N° d'ordre NNT : 2020LYSE2016

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 476 Neurosciences et Cognition

Discipline : Psychologie cognitive

Soutenue publiquement le 23 juin 2020, par :

Jennifer KRZONOWSKI

Perception et production de monophtongues de l'anglais par des apprenants francophones : effets d'entraînements en perception et en production.

Devant le jury composé de :

Frédéric ISEL, Professeur des universités, Université Paris Nanterre, Président

Sophie DUFOUR, Directrice de Recherches, C.N.R.S. Marseille, Rapporteure

Sophie HERMENT, Professeure des universités, Aix-Marseille Université, Rapporteure

Emmanuel FERRAGNE, Maître de conférences, Université Paris Diderot, Examinateur

François PELLEGRINO, Directeur de Recherches, C.N.R.S., Directeur de thèse

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « Paternité – pas d'utilisation commerciale – pas de modification » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

Université Lumière Lyon 2
École doctorale Neurosciences et Cognition (NSCo)
Laboratoire Dynamique Du Langage – UMR 5596

THÈSE

Réalisée en vue de l'obtention du grade de

DOCTEUR EN PSYCHOLOGIE DE L'UNIVERSITÉ LYON 2

Discipline : Sciences cognitives Mention Psychologie cognitive

Perception et production de monophtongues de l'anglais par des apprenants francophones : effets d'entraînements en perception et en production

Par **Jennifer Krzonowski**

Sous la direction d'**Emmanuel Ferragne** et **François Pellegrino**

Présentée et soutenue publiquement à Lyon le 31 mars 2020

Devant un jury composé de :

François Pellegrino (DR)	DDL – Univ. Lyon 2 – CNRS	Directeur
Emmanuel Ferragne (MCF)	CLILLAC-ARP – Univ. de Paris	Directeur
Sophie Dufour (DR)	LPL – Aix-Marseille Univ. – CNRS	Rapporteure
Sophie Herment (PR)	LPL – Aix-Marseille Univ. – CNRS	Rapporteure
Frédéric Isel (PR)	MoDyCo – Univ. Paris Nanterre - CNRS	Président

Résumé

Dans cette thèse nous testons l'effet de différents types d'entraînements sur l'acquisition des monophtongues /æ, ʌ, ɑː, ɪ, et iː/ de l'anglais par des apprenants francophones adultes. Des mesures comportementales sont réalisées avant et après entraînements afin d'évaluer les changements sur la perception et la production de ces voyelles. De plus, une étude en potentiels évoqués a été menée afin de mettre en évidence des corrélats neuronaux consécutifs à l'entraînement du contraste /æ-ʌ/.

Deux types d'entraînements ont donc été mis en place : 1) le premier, du type *High Variability Phonetic Training* (*HVPT*) comportait des tâches classiques d'identification et de discrimination avec *feedback* sur la justesse des réponses. Cet entraînement était proposé à 16 participants constituant le groupe PE. 2) le second reposait sur une tâche d'imitation de mots dans lequel un *feedback* était donné sur la justesse des voyelles produites à partir des paramètres acoustiques mesurés en quasi temps réel. Cet entraînement était proposé à 16 participants constituant le groupe PR. Un troisième groupe ne recevant pas d'entraînement constituait le groupe contrôle (groupe C).

En perception, les résultats montrent que l'entraînement *HVPT* améliore significativement les performances des apprenants en identification et en discrimination comparativement au groupe C, alors que l'amélioration du groupe PR est limitée à l'identification.

Les performances en production ont été évaluées par le biais de deux méthodes : 1) une méthode « objective » basée sur l'analyse des paramètres acoustiques des voyelles. 2) une méthode « subjective » basée sur une tâche d'identification des voyelles produites réalisée par des juges natifs. L'entraînement de type *HVPT* semble être le plus enclin à améliorer les performances en production des apprenants. Cependant, il a été observé que l'amélioration en production était restreinte à l'intelligibilité telle qu'évaluée par des locuteurs natifs et que l'accent étranger mesuré par le biais de mesures acoustiques n'était pas significativement réduit dans les groupes expérimentaux comparés au groupe contrôle.

L'étude en électroencéphalographie n'a pas permis d'observer les corrélats attendus mais nous discutons les améliorations potentielles à apporter à notre protocole.

Mots-clés : acquisition de la phonologie L2, voyelle, production, perception, entraînement perception, entraînement en production, potentiels évoqués.

Abstract

This work examines the effects of different kinds of training for the acquisition of the British English monophthongs /æ, ʌ, ɑː, ɪ and iː/ by French late learners. Behavioral measures were computed before and after training in order to evaluate changes occurring on perception and production abilities. Furthermore, an Event Related Potentials study (MMN) was conducted in order to observe the neural correlates of the acquisition of the trained /æ-ʌ/ contrast.

A first High Variability Phonetic Training including identification and discrimination tasks with feedback was designed. It was proposed to 16 French learners constituting the so-called PE-Group. 16 other participants constituting the PR-Group received the second training program composed of a word imitation task with visual feedback. The feedback was based on the analysis of the acoustic parameters of the vowel produced by the learners, allowing them to evaluate their own production. A third group, who did not receive any training program was also included (C-Group).

The PE-group was shown to perform better after training in identification and discrimination compared to the C-Group, and the PR-Group improved in identification only. Production performances were evaluated through two different methods: 1) one based on the acoustic parameters of the vowels; 2) the other based on the classification of the produced vowels by native English speakers. It was shown that the *HVPT* was better to improve production performances. However, it was observed that production improvement was restricted to intelligibility, as evaluated by native speakers, and that foreign accent reflected by acoustic measures was not significantly reduced in experimental groups (PE and PR) compared to the C-Group.

The expected Event Related Potentials were not observed, but we discuss potential improvements to our protocol.

Keywords: L2 phonology acquisition, vowel, production, perception, perceptual training, production training, Event Related Potentials.

Remerciements

Je souhaite en tout premier lieu remercier mes deux directeurs de thèse, François Pellegrino et Emmanuel Ferragne, de m'avoir accompagnée jusqu'ici.

François, un immense merci d'avoir été là pile au bon moment, quand j'avais baissé les bras. Merci pour toute l'aide que tu m'as apportée pour me permettre de terminer ce travail, pour tes conseils toujours si justes et pour ton aide avec ces maudits ERP.

Manu, tu es celui sans qui je ne serais ni où je suis, ni qui je suis, tant sur le plan professionnel que personnel. Depuis ce jour, il y a bientôt 10 ans, où tu nous as présenté ton projet sur l'*Hypothèse de la Phonémicité Gradiente* à Marie et moi, tu m'as appris tout ce que tu pouvais avec confiance, patience et bienveillance. J'ai eu une chance immense d'avoir un directeur comme toi. Merci pour toutes ces années où on a bien travaillé ensemble dans la bonne humeur.

Je remercie les membres du jury, Sophie Dufour et Sophie Herment d'avoir accepté d'être les rapporteures de mon travail et Frédéric Isel d'avoir accepté le rôle de président du jury. Je remercie également Barbara Tillmann et Noël Nguyen d'avoir accepté de faire partie de mon comité de suivi de thèse. Merci à tous les deux pour vos précieux conseils.

Je remercie tous les locuteurs anglais qui ont participé à cette étude. Au passage, je remercie Maud et Marie qui, depuis Paris et Reading, m'ont été d'une grande aide pour recruter des participants quand à Lyon c'était compliqué.

Je remercie également tous les étudiants qui ont accepté de participer à cette étude et qui se sont donnés du mal à réaliser toutes ces tâches.

Merci au LabEx ASLAN d'avoir financé cette étude.

Un immense merci à mes quatre relecteurs de coquilles (et plus) Nath, Séverine, Jeff et Antoine.

Nath, merci pour ta présence, ton soutien inconditionnel et pour tous ces bons moments dans votre bureau ou dehors.

Je souhaite remercier tous mes collègues pour leur bienveillance qui m'a permis de terminer enfin cette thèse. Merci à Sophie Kern et Antoine Guillaume, les deux directeurs du labo qui se sont succédés le temps de la réalisation de ma thèse et qui ont fait en sorte de me mettre tous

les moyens à ma disposition pour que je puisse mener ensemble mon travail d'ingénieure et de doctorante. Un merci tout particulier à Mélanie, Françoise et Alice pour vos encouragements et vos petites phrases qui, sans que vous le sachiez peut-être, ont trotté dans ma tête et m'ont permis de terminer. Véro, tu fais également partie des personnes qui m'ont menée jusqu'ici, merci de m'avoir fait confiance et de m'avoir toujours soutenue.

Egidio, merci pour toutes les aventures dans lesquelles tu m'as embarquée ces dernières années.

Seb, merci de toute l'aide que tu m'as apportée au début de ma thèse, quand il était question de pseudo-randomisation et de twinours à l'époque du « bureau de la bonne humeur et de la non-approximation ». Et merci pour ces années de rire au quotidien.

Merci Christian pour ton aide dans tous les petits tracas informatiques.

Un grand merci à Arnaud, Linda et Rabia qui m'ont permis de partir en mission, défrayer mes participants et tant d'autres choses.

Je souhaite également remercier tous mes collègues de la MSH : Valérie, Christian, Jean-Luc, Jean-Pierre, Emma. On ne prend pas le temps de vous le dire mais si vous n'étiez pas là, les choses seraient beaucoup moins faciles pour nous dans les labos.

Merci à mes amies du labo d'avoir été là pour me montrer l'exemple, pour les pauses soleil, les pauses déjeuner, les mercredis poulets et pour m'avoir encouragée jusqu'au bout : MP, Marie, Émilie, Agathe, Priyanka, Aurore et Séverine. Et merci Leslie, d'avoir été mon modèle de courage.

Je remercie tous mes amis de la fac et d'ailleurs, les Psychopotes, Marion et Rémy, Kim et Louis, Chloé. Merci pour votre soutien pendant toutes ces années et surtout sur la fin.

Jeff, mon meilleur ami, meilleur soutien en toutes circonstances, merci d'avoir été là jusqu'au bout et d'être là tout court.

Je remercie ma famille de toujours me faire confiance quoi que j'entreprenne. Merci d'être là tous les quatre.

Enfin, Antoine, merci pour tout ce que tu as fait ces dernières semaines pour me permettre de ne penser qu'à ma thèse. Tu as été parfait, comme toujours. Je ne peux pas rêver meilleur homme de ma vie d'adulte.

Table des matières

Résumé ..... **i**

Abstract ..... **ii**

Remerciements ..... **iii**

Table des matières ..... **v**

Liste des figures ..... **viii**

Liste des tableaux ..... **xi**

Préambule **1**

Partie 1. Introduction générale **5**

1 Acquisition de la phonologie d'une langue seconde **5**

1.1 Modèles théoriques de l'acquisition de la phonologie 6

1.1.1 Modèles de perception 6

1.1.2 Modèles liés à la production 17

1.2 Acquisition L2 et potentiels évoqués 21

1.2.1 Les potentiels évoqués liés à des traitements phonologiques 23

1.2.2 Potentiels évoqués et perception des sons de L2 27

2 Les entraînements pour l'acquisition de la L2 ..... **31**

2.1 Les entraînements perceptifs 31

2.1.1 Les débuts 31

2.1.2 Les entraînements perceptifs pour l'acquisition des voyelles 34

2.2 D'autres types d'entraînements pour l'acquisition de contrastes vocaliques de L2 38

2.3 Entraînements perceptifs et potentiels évoqués 41

3 Les systèmes vocaliques du français et de l'anglais britannique **48**

3.1 Système vocalique du français 48

3.2 Système vocalique de l'anglais britannique 50

3.3 Difficultés des apprenants francophones pour les voyelles de l'anglais 51

4 Objectifs de l'étude 52

Partie 2. Contribution expérimentale 59

1 Matériel expérimental 59
2 Méthodologie expérimentale des entraînements 66

2.1 Participants 66

2.2 Déroulement des entraînements 67

2.2.1 Entraînement en perception (Groupe PE) 69

2.2.2 Entraînement en production (Groupe PR) 72

3 Évaluation en perception 75

3.1 Hypothèses 75

3.2 Méthodologie 76

3.2.1 Description des tâches utilisées 76

3.2.2 Traitement des données 78

3.3 Résultats 79

3.3.1 Comparaison avec les locuteurs natifs 80

3.3.1 Évaluation de l'effet des entraînements 88

3.4 Discussion des résultats en perception 90

4 Évaluation en production 96

4.1 Objectifs et hypothèses 96

4.2 Méthodologie 98

4.2.1 Description des tâches utilisées 98

4.2.2 Évaluation *objective* des paramètres acoustiques 99

4.2.3 Évaluation *subjective* des productions 100

4.3 Résultats 101

4.3.1 Production des voyelles par les apprenants francophones avant entraînement : comparaison avec les productions natives 101

4.3.2 Évaluation de l'effet des entraînements sur la production 114
4.4 Discussion des résultats en production 135
4.4.1 Évaluation avant entraînement 135
4.4.2 Évaluation des entraînements 138
5 Évaluation en EEG 140
5.1 Objectifs et hypothèses 140
5.2 Méthodologie 141
5.2.1 Matériel expérimental 141
5.2.2 Protocole expérimental 143
5.2.3 Analyses des données 144
5.3 Résultats 146
5.4 Discussion des résultats en EEG 150
Partie 3. Conclusion générale 153
1 Résumé des résultats 153
2 Perspectives 157
Référence bibliographiques 165

Liste des figures

Figure 1 : Un nouveau son ayant deux étiquettes possibles : /ɛ/ ou /ɪ/ (tiré de Pierrehumbert, 2001) 8
Figure 2 : Représentation de l'effet d'aimant perceptif (image reproduite de Kuhl et Iverson, 1995) 11
Figure 3 : Représentation schématique du PAM (reproduit de Best et al., 2016)..... 14
Figure 4 : Représentation schématique des prédictions du modèle L2LP d'Escudero (selon Burgos, 2018) 17
Figure 5 : Schématisation du modèle DIVA (reproduit de Guenther, 2006).....19
Figure 6 : Les effets de la déviance sur la MMN pour les électrodes Cz et Pz (image tirée de Näätänen et Picton, 1987) 24
Figure 7 : Réponses MMN observées pour les stimuli langagiers et les tons chez des enfants finlandais et russes (image reproduite de Nenonen et al., 2003) 28
Figure 8 : Grandes moyennes des différences entre déviants et standards dans les quatre blocs de voyelles pour les locuteurs finnois monolingues en ligne fine, les étudiants finnois en ligne épaisse et les locuteurs anglais monolingues en pointillés. (image reproduite de Peltola et al., 2013).....30
Figure 9 : Exemple de feedback visuel proposé aux apprenants dans l'étude de Kartushina et collègues (2015) à gauche (a). Le carré blanc correspond à la voyelle produite par l'apprenant et le gris à la cible native à atteindre. La représentation de droite (b) montre les unités et l'étendue utilisée qui correspond à l'espace des 4 voyelles de leur étude. (image reproduite de Kartushina et al., 2015) 41
Figure 10 : MMN en Fz chez les locuteurs de langue maternelle anglaise et finnois avant entraînement pour les stimuli beat et bit avec des durées normales (donc différentes) ou égales et les changements de fréquence et durée dans la condition contrôle avec les stimuli non langagiers. (image reproduite de Ylinen et al., 2010).....46
Figure 11 : MMN sur l'électrode Fz en prétest et posttest pour les stimuli beat et bit avec des durées normales ou égales et pour les changements de fréquence et durée dans la condition contrôle avec les stimuli non langagiers. (Image reproduite de Ylinen et al,. 2010) 47
Figure 12 : Trapèze vocalique des voyelles orales du français (reproduit de Capliez, 2011) 50
Figure 13 : Trapèze vocalique des monophtongues de l'anglais (reproduit de Capliez, 2011) 51
Figure 14 : Voyelles du français (d'après Georgeton et al., 2012) et monophtongues de l'anglais (d'après Deterding, 1997) pour les femmes 55
Figure 15 - Interface du logiciel ROCme! utilisé pour l'enregistrement des locuteurs natifs 62
Figure 16 : Estimation formantique réalisée sous Praat.....63
Figure 17 : Représentation dans l'espace F1/F2 des voyelles des stimuli utilisés dans les tests.....64

Figure 18 : Représentation dans l'espace F1/F2 des voyelles des stimuli utilisés dans les entraînements.....65
Figure 19 : Durées moyennes des voyelles utilisées dans les tests et les entraînements 66
Figure 20 : Interfaces de la tâche d'identification à deux choix forcés de l'entraînement en perception 71
Figure 21 : Écran présenté dans la tâche de répétition pour écouter le mot à répéter 73
Figure 22 : Feedback visuel pour la tâche de répétition de mots dans l'entraînement en production.....74
Figure 23 : Écran présenté lors de la tâche d'identification à 5 choix forcés 77
Figure 24 : Écran présenté dans la tâche de discrimination *oddity* 78
Figure 25 : Scores en identification pour le groupe natif pour chacune des voyelles testées. 80
Figure 26 : Scores des 3 groupes C, PE et PR en identification en prétest et posttest pour chacune des voyelles testées.. 82
Figure 27 : Scores en discrimination du groupe N 85
Figure 28 : Scores dans la tâche de discrimination pour chacun des groupes selon le contraste. 86
Figure 29 : Amélioration entre prétest et posttest dans la tâche d'identification et de discrimination. 89
Figure 30 : Écran présenté lors de la tâche de production en anglais 99
Figure 31 : Écran d'instructions présenté aux locuteurs natifs ayant évalué les productions des groupes expérimentaux (N, C PE et PR) 100
Figure 32 : Voyelles de l'anglais et du français produites par des natifs 103
Figure 33 : Représentation dans l'espace F1/F2 des voyelles produites dans les mots bée et baie 104
Figure 34 : Voyelles de l'anglais et du français produites par des apprenants français 107
Figure 35 : Voyelles de l'anglais produites par des locuteurs natifs et par les apprenants français 109
Figure 36 : Comparaison des durées moyennes des voyelles anglaises et françaises produites par des locuteurs L1 et comparaison des durées des voyelles anglaises produites par des locuteurs de L1 et de L2. 111
Figure 37 : Matrice de confusion des voyelles des natifs selon le modèle de classification lda.m1 et des apprenants 113
Figure 38 : Durées des voyelles produites par chacun des groupes, en prétest et en posttest. 117
Figure 39 : Représentation des voyelles produites par les 16 locuteurs natifs (Groupe N) dans l'espace des deux premiers formants. 120
Figure 40 : Ellipses de dispersion des voyelles produites en prétest et en posttest pour chacun des trois groupes d'apprenants. 121
Figure 41 : Voyelles produites par la participante PE05 dans l'espace des deux premiers formants en prétest et en posttest. 122
Figure 42 : Proportions de classification correcte à l'analyse discriminante linéaire des voyelles produites selon les groupes en prétest et en posttest 124

Figure 43 : Matrice de confusion issue de la classification des voyelles du Groupe N selon le modèle lda.m1 appris sur les données des locuteurs ayant enregistré le matériel expérimental 126
Figure 44 : Matrices de confusion de la classification des voyelles produites par les apprenants en prétest et en posttest 127
Figure 45 : Accords inter-juges (Kappa de Fleiss) selon les voyelles dans les groupes de juges 129
Figure 46 : Proportions d'identification correcte des voyelles produites par les apprenants en prétest, en posttest et par les locuteurs natifs dans la tâche de jugement par des locuteurs natifs.....130
Figure 47 : Matrice de confusion issue de la tâche d'identification réalisée par les juges natifs des voyelles produites par les locuteurs du groupe N 131
Figure 48 : Matrices de confusion issues de la tâche d'identification réalisée par les juges natifs sur les données apprenants en prétest et en posttest 132
Figure 49 : Distribution des scores individuels moyens de classification correcte à l'analyse discriminante linéaire selon le groupe et le test (prétest vs. posttest) 133
Figure 50 : Distribution des scores individuels moyens d'identification correcte par les juges natifs selon le groupe et le test (prétest vs. posttest).....134
Figure 51 : Représentation des voyelles utilisées dans la tâche en EEG.....142
Figure 52 : iMMN (analyse initiale 100 standards / 100 déviants par condition).....146
Figure 53 : Onde iMMN pour le contraste /r/-/l/ sur l'électrode Fz dans l'étude de Zevin et al. (2010) pour les apprenants japonais.....147
Figure 54 : MMN classique (analyse 100 standards / 100 déviants) 148
Figure 55 : iMMN (analyse 300 standards / 100 déviants). 148
Figure 56 : iMMN (signaux recalés) 150

Liste des tableaux

Tableau I : Classification des voyelles orales du français (d'après Gess, Lyche et Meisenburg, 2012, p. 4).....49
Tableau II : Classification des voyelles orales de l'anglais.....51
Tableau III : Déroulement du protocole expérimental suivi dans les groupes. 56
Tableau IV : Paires minimales utilisées dans les phases test et les phases d'entraînement 60
Tableau V : Origine géographique des locuteurs et locutrices ayant enregistrés les stimuli expérimentaux 61
Tableau VI : Âge des locuteurs anglais ayant enregistré le matériel expérimental.....61
Tableau VII : Valeurs moyennes des trois premiers formants (en Hz) et de la durée des voyelles (en ms) des stimuli utilisés dans les phases tests.....63
Tableau VIII : Valeurs moyennes des trois premiers formants (en Hertz) et de la durée des voyelles (en ms) des stimuli utilisés dans les entraînements.....64
Tableau IX : Âges des participants selon les groupes 67
Tableau X : Déroulement des entraînements selon les groupes 68
Tableau XI : Listes de stimuli utilisés dans les séances 4 et 5 69
Tableau XII : Valeurs moyennes des trois premiers formants (en Hz) et de la durée des voyelles (en ms) utilisées pour représenter les valeurs cibles en fonction du sexe du participant dans l'entraînement en production.....75
Tableau XIII : Table d'analyse du modèle linéaire mixte pour les scores *d'* en identification en prétest.....83
Tableau XIV : Table d'analyse du modèle linéaire mixte pour les scores *d'* en identification en posttest.....83
Tableau XV : Table d'analyse du modèle linéaire mixte pour les scores *d'* en discrimination en prétest.....87
Tableau XVI : Table d'analyse de modèle linéaire mixte pour les scores *d'* en discrimination en posttest 87
Tableau XVII : Table d'analyse du modèle linéaire mixte des scores *d'* en identification des groupes d'apprenants en prétest et posttest 88
Tableau XVIII : Table d'analyse de la différence des scores *d'* entre prétest et posttest en identification 89
Tableau XIX : Table d'analyses du modèle linéaire mixte des scores *d'* en discrimination des groupes d'apprenants en prétest et posttest 90
Tableau XX : Table d'analyse de la différence des scores *d'* entre prétest et posttest en discrimination.....90
Tableau XXI : Productions évaluées par chacun des six juges natifs 101
Tableau XXII : Répartition des voyelles dans le premier sous-ensemble en fonction de la langue maternelle des locuteurs 102

Tableau XXIII : Distances euclidiennes moyennes en Hertz entre les voyelles d'une même région acoustique dans l'espace F1/F2 produites par les locuteurs français pour les deux langues, pour les femmes et les hommes 108
Tableau XXIV : Distances acoustiques moyennes entre les voyelles d'une même région acoustique dans les voyelles produites par les locuteurs natifs et les locuteurs français.....110
Tableau XXV : Pourcentages de classification correcte des voyelles produites par les locuteurs natifs ayant enregistré les stimuli et ceux du groupe N selon les différents modèles d'analyse discriminante linéaire testés.....112
Tableau XXVI : Taux de classification correcte des voyelles [i] natives selon les différents modèles testés 114
Tableau XXVII : Composition du corpus utilisé pour évaluer les performances en production en prétest et en posttest 116
Tableau XXVIII : Table d'analyse des durées vocaliques en fonction du Groupe et de la Voyelle en prétest 118
Tableau XXIX : Table d'analyse des durées vocaliques en fonction du Groupe et de la Voyelle en posttest.....118
Tableau XXX : Table d'analyse des durées vocaliques en fonction du Test (prétest vs. posttest), du Groupe et de la Voyelle. 119
Tableau XXXI : Différents modèles de classification utilisés pour classifier les voyelles produites par les locuteurs natifs ayant enregistré le matériel expérimental.....124
Tableau XXXII : Caractéristiques temporelles (en ms) et spectrales (en Hz) des stimuli utilisés dans la tâche en EEG. 142
Tableau XXXIII – Organisation de l'enregistrement EEG en 4 blocs.....143
Tableau XXXIV : Tableau récapitulatif des effets des entraînements dans les différentes tâches. 154

Préambule

Parmi les compétences linguistiques qui font l'objet d'un enseignement spécifique dans l'acquisition de l'anglais langue seconde (L2), la prononciation est restée longtemps reléguée au second plan. En effet, alors que les aspects syntaxiques, morphologiques et lexicaux sont acquis par le biais d'un enseignement explicite (à travers les cours de grammaire et les listes de vocabulaires), on note que 1) la composante phonético-phonologique est supposée être acquise implicitement via l'écoute répétée des sons de la L2 et, 2) il est de toute façon superflu d'y consacrer du temps puisqu'il est entendu qu'un accent étranger ne nuit pas à la compréhension.

À ces deux idées reçues, il convient d'objecter que :

1. dans le cas d'apprenants tardifs ayant dépassé l'âge limite d'acquisition de la phonologie selon l'hypothèse de la période critique (Lenneberg, 1967), la simple écoute répétée, dans le cadre de la salle de classe, de contrastes phonologiques spécifiques à la L2 ne saurait suffire puisque, par définition, ces apprenants présentent une surdité relative aux contrastes absents de leur langue maternelle (L1). De nombreuses études comportementales et neurophysiologiques ont en effet montré que, durant leur première année, les enfants perdent la capacité à discriminer les phonèmes qui ne sont pas utilisés dans leur langue maternelle (Dehaene-Lambertz & Dehaene, 1994 ; Kuhl et al., 1992).
2. de nombreuses études, notamment dans le domaine de la sociophonétique, soulignent que la stigmatisation induite par un accent étranger présente des enjeux sociétaux de tout premier ordre. Il est avéré, par exemple, qu'un accent étranger peut être plus stigmatisant qu'un type ethnique (Kinzler et al., 2007; Kinzler et al., 2009), et qu'un témoignage en justice produit avec un accent étranger discrédite le témoin (Frumkin, 2007). Par ailleurs, un accent étranger entraîne un jugement défavorable de la part de l'auditeur même lorsque le locuteur est parfaitement intelligible (Munro, Derwing et Sato, 2006) et ce jugement négatif est valable quel que soit l'accent (Lindemann, 2003).

La mise en place de techniques solides pour l'enseignement de la phonologie de l'anglais langue seconde s'impose donc comme une nécessité absolue, surtout quand on constate que 1) la plupart de nos contemporains sont amenés à apprendre l'anglais et, 2) l'enseignement le plus

développé en la matière (celui qui est dispensé dans les départements d'anglais) se cantonne le plus souvent à des techniques de phonétique corrective rudimentaires. Dans le cadre de l'enseignement d'une langue seconde, la maitrise de la prononciation implique pour un apprenant d'acquérir un système phonético-phonologique complet aussi bien en perception qu'en production. Ce système comprend des éléments suprasegmentaux et segmentaux nécessaires à l'apprentissage morpho-lexical et discursif (Detey et al., 2018). Ainsi Detey (2016) propose de distinguer au moins quatre approches dans l'enseignement d'une L2 : 1) l'enseignement de la prononciation comportant les activités d'apprentissage du système phonético-phonologique par l'intermédiaire du reste du matériel à apprendre ; 2) la correction phonologique qui consiste par exemple à traiter des confusions phonémiques ; 3) la correction phonétique qui consiste à traiter les productions inappropriées du point de vue phonétique, entrainant par exemple un accent étranger ; enfin 4) le développement des habiletés socio-phonétiques et phonostylistiques permettant à l'apprenant de produire une parole appropriée au contexte, par exemple social, dans lequel il évolue. Dans cette thèse, nous nous situerons au niveau des approches 2 et 3, puisque nous nous intéresserons à l'acquisition de contrastes vocaliques de l'anglais connus pour poser des problèmes récurrents aux apprenants francophones. Nous souhaitons mettre en place des méthodes d'entraînements automatisées permettant de corriger les erreurs phonémiques et phonétiques des apprenants tant au niveau perceptif qu'en production.

Depuis la fin des années 80, de nombreuses recherches ont tenté de mettre au point des méthodes permettant d'améliorer les compétences de locuteurs de L2 du point de vue de la phonologie, et parmi ces recherches une technique a rencontré un grand succès, le *High Variability Phonetic Training* (*HVPT*) comme le montre Thomson (2018) dans une revue de question. Barriuso et Hayes-Harb (2018) considèrent cette méthode comme un excellent candidat à l'application des recherches sur la prononciation en L2 dans le champ de l'apprentissage et de l'enseignement des langues secondes en situation écologique. Cette méthode, qui repose sur des tâches de perception, a été appliquée à des contrastes consonantiques (e.g., Bradlow et al., 1997, 1999 ; Lively et al., 1993, 1994 ; Logan et al., 1991), comme à des contrastes vocaliques (e.g., Akahane-Yamada et al., 1998 ; Aliaga-García & Mora, 2009 ; Carlet & Cebrian, 2014 ; Cebrian & Carlet, 2015 ; Inceoglu, 2016), et ce pour différentes paires Langue maternelle (L1) - Langue seconde (L2). Le *HVPT* présente pour l'enseignement l'avantage d'être peu couteux en ressources matérielles et humaines puisqu'il ne nécessite qu'un simple ordinateur. Cependant, dans le contexte de l'acquisition de la

phonologie de l'anglais par des apprenants francophones, les preuves de l'efficacité de cette technique sont encore peu nombreuses, notamment pour les monophtongues. La seule étude qui, à notre connaissance, a porté sur l'acquisition des voyelles orales de l'anglais par des apprenants francophones a été réalisée par Iverson, Pinet et Evans (2012) et a montré des résultats mitigés, notamment quant à la production des voyelles. Nous avons donc souhaité mettre au point un nouveau type d'entraînement dont l'objectif est de mettre l'accent sur la production des voyelles tout en ne négligeant pas leur perception. Tout comme le *HVPT*, nous avons fait en sorte que cet entraînement soit automatisé. Ainsi, un premier objectif de cette thèse était de mettre au point ce nouvel entraînement et de le comparer au *HVPT* quant à son efficacité pour améliorer les compétences en perception, mais aussi en production de voyelles de l'anglais, par des apprenants francophones.

De plus, nous avons souhaité observer dans quelle mesure l'usage de la nouvelle technique d'entraînement phonétique a un impact sur le traitement des sons de parole entrainés à un niveau pré-attentionnel. Pour cela, nous avons souhaité observer l'évolution de potentiels évoqués liés au traitement de voyelles faisant l'objet de l'entraînement. Ainsi, en plus de la méthodologie comportementale habituelle nécessaire à l'évaluation de l'efficacité d'un entraînement sur les compétences, nous avons mis en place, une étude en électroencéphalographie avec la méthode des potentiels évoqués.

Le premier chapitre de thèse présente tout d'abord une revue de la littérature sur les principaux modèles d'acquisition d'une langue seconde et décrit les indices neuronaux susceptibles de témoigner de la perception des sons d'une L2. Nous ferons ensuite une synthèse des études dont l'objectif est de mettre au point des techniques d'entraînements pour l'acquisition de la phonologie de l'anglais L2. Enfin, nous détaillerons les difficultés rencontrées par les apprenants francophones pour les voyelles orales de l'anglais avant de décrire les objectifs de notre étude.

Cette recherche a nécessité la construction d'un matériel expérimental spécifique ainsi que la mise en place d'une méthodologie particulière afin de mettre au point et d'évaluer des entraînements à destination d'apprenants francophones pour plusieurs contrastes vocaliques de l'anglais. Le second chapitre de cette thèse présentera ces différents éléments méthodologiques ainsi que les résultats issus des expériences réalisées. Ces dernières portent sur les effets des entraînements proposés sur la perception et la production des voyelles étudiées, mais également sur les corrélats électroencéphalographiques de la perception de voyelles ayant fait

l'objet d'un entraînement. Enfin, nous ferons une conclusion générale de l'étude ouvrant sur des perspectives de recherche.

Avant d'aller plus loin, nous souhaitons apporter quelques précisions terminologiques nécessaires au lecteur. Tout d'abord, alors qu'en français le terme d'*opposition phonémique*, que l'on identifie par le biais de paires minimales, est traditionnellement utilisé, nous avons ici un usage un peu plus « international » des termes, qui fait de *contraste phonémique* le strict synonyme d'*opposition phonémique*. De plus, bien que nous ayons conscience de la différence entre le *phonème* ou catégorie phonémique, qui est une représentation abstraite, d'une part, et la *réalisation d'un phonème*, d'autre part, nous utilisons parfois ici le terme de phonème là où d'autres utiliseraient le terme *son* ou *réalisation*. Prenant une position proche de celle des modèles à exemplaires, notre définition du phonème est ainsi assez concrète, ce qui amène à parler par exemple de distance acoustique entre phonèmes, en un raccourci visant à alléger la lecture. Enfin, nous tenons à préciser que, conformément à l'usage international, le point sera utilisé comme séparateur de décimales.

Partie 1. Introduction générale

1 Acquisition de la phonologie d'une langue seconde

Percevoir et produire les sons d'une langue qui est autre que la langue maternelle n'est pas chose facile et, depuis déjà plusieurs décennies, le fameux « accent étranger » fait l'objet de très nombreuses recherches destinées à en comprendre les origines. Ces études ont mis en évidence de nombreux facteurs influant sur l'acquisition de la phonologie d'une langue seconde (L2). Dans une revue sur les facteurs affectant cet accent étranger, Piske et collaborateurs (2001) mettent en avant l'influence du temps de résidence dans une communauté parlant la L2 en question. Bien que le genre des apprenants ne semble jouer un rôle que dans les premiers temps de l'apprentissage selon Flege (1988), il aurait aussi une influence : les femmes reçoivent en effet de meilleurs scores en L2 que les hommes selon certaines études (Asher & García, 1969 ; Thompson, 1991). Le nombre d'années d'instruction formelle intervient aussi sur le niveau de compétence (e.g., Flege & Fletcher, 1992), de même que la motivation des apprenants (Elliott, 1995 ; Purcell & Suter, 1980 ; Suter, 1976) et leur habileté naturelle dans une composante cognitive inhérente à l'apprentissage : la capacité à imiter (Suter, 1976). Pour finir, la fréquence d'utilisation de la L2 est un aspect quantitatif qui retentit aussi sur les compétences dans cette langue (Purcell & Suter, 1980 ; Suter, 1976). L'influence de tous ces facteurs sur l'accent étranger est attestée par des preuves expérimentales mais, selon l'hypothèse de la période critique, le facteur le plus déterminant serait l'âge d'acquisition de la L2 (Lenneberg, 1967). C'est Penfield et Robert (Lenneberg, 1960 pour une revue) qui ont introduit en 1959 l'idée selon laquelle il existerait une période critique délimitée avant 9 ans pour l'apprentissage d'une langue. Lenneberg (1967) a précisé plus tard que la capacité à apprendre une langue de manière naturelle, par simple exposition, disparait à la puberté et qu'il n'est plus alors possible d'acquérir une L2 avec un niveau de locuteur natif, à moins de fournir des efforts considérables. Il a également indiqué qu'il devenait extrêmement difficile d'effacer un accent étranger une fois la puberté passée. Depuis Lenneberg (1967), Singleton (2005) relève qu'au moins dix versions différentes de cette hypothèse ont été proposées avec un âge limite qui varie. De plus, un ensemble d'expériences concourt à montrer que cette hypothèse est trop stricte puisque des cas d'apprenants de L2

tardifs dont l'accent ne pouvait pas être distingué de celui de locuteurs natifs sont documentés (e.g., Bongaerts et al., 1995). Fort de ces observations, un nouveau champ de recherche s'est développé en vue de comprendre les mécanismes qui sous-tendent l'acquisition des sons de L2 à l'âge adulte et d'expliquer comment se créent de nouvelles catégories phonémiques. Nombre de ces recherches ont porté sur la perception et la production inter-langues dans le but de mettre en évidence les influences du système phonologique de la langue maternelle (L1) sur la manière dont des sons non-natifs sont perçus et produits. Des modèles théoriques ont alors été développés afin de prédire la perception de la L2 à partir des similarités/ressemblances avec la L1.

Nous présenterons ici différents modèles développés pour expliquer l'acquisition de la phonologie de la L1 et de la L2 en perception et en production. Les principaux modèles cités dans les études sur l'acquisition de la phonologie de la L2 portent sur la perception. Leur hypothèse sous-jacente est la primauté de la perception sur la production : les difficultés de production des sons en L2 trouveraient leur origine dans des anomalies en perception des sons de la L2 (Best & Tyler, 2007 ; Flege, 1995). En plus de ces modèles, nous avons choisi de présenter des modèles d'acquisition de la production de sons en L1 qui nous semblaient pertinents compte-tenu du fait que l'un des objectifs de cette thèse est de proposer un entraînement particulièrement orienté sur la production.

1.1 Modèles théoriques de l'acquisition de la phonologie

1.1.1 Modèles de perception

1.1.1.1 La théorie des modèles à exemplaires

La théorie des exemplaires est initialement une théorie de la perception en psychologie (Goldinger, 1998). Cette théorie propose que les êtres humains catégorisent et mémorisent leurs expériences comme des nuages d'exemplaires, c'est-à-dire de grands ensembles de souvenirs d'expériences individuelles, personnelles. Cette théorie a été appliquée à la perception des sons de parole, notamment par Johnson (1997). En phonologie, la théorie des exemplaires propose que nous ne formons pas des abstractions de propriétés formelles des phonèmes et des mots, mais que nous stockons leurs occurrences individuelles telles qu'elles sont perçues. Ces éléments épisodiques donnent lieu à la création d'un lexique mental

contenant non seulement des informations très précises sur les propriétés linguistiques (contenant toutefois le strict minimum nécessaire à ce que le message soit compris) mais aussi extra-linguistiques des sons. En effet, en plus de données sur les caractéristiques acoustiques des sons perçus, des données sur le locuteur, le style de parole ou encore le dialecte sont stockées. La théorie des exemplaires propose au final que les sons de parole sont stockés en mémoire comme des nuages d'exemplaires labélisés. Elle constitue un bon modèle pour traiter de la catégorisation des voyelles (Ettlinger et Johnson, 2009).

Pierrehumbert (2001) résume la théorie des exemplaires avec le cas du phonème. Une telle catégorie serait représentée par un nuage de nombreux exemplaires, d'autant plus proches les uns des autres qu'ils sont très similaires. Les exemplaires stockés rendent compte de la variation liée aux corrélats physiques des catégories. Par exemple, un exemplaire de la voyelle /ɛ/ peut être produit avec une variété de valeurs de formants, de durée, de fréquence fondamentale, etc. L'espace des paramètres acoustiques est cartographié selon les étiquettes du système catégoriel. La grande quantité des exemplaires permet notamment d'expliquer les effets de fréquence.

Bien que les capacités d'un individu à mémoriser des exemplaires soient extrêmement importantes (K. Johnson, 1997), la quantité de parole qu'il doit traiter dans une vie entière est tellement importante que le modèle de Pierrehumbert ne prétend pas que chaque exemplaire rencontré reste stocké à vie. Ce modèle propose deux alternatives : une dégradation de la mémorisation des épisodes les plus anciens en faveur des souvenirs récents et un effet de la granularité des paramètres acoustiques (e.g., durée ou F0) menant à un pouvoir de résolution permettant un codage ne différenciant pas inutilement les trop petites différences.

Ainsi, dans ce modèle, chaque exemplaire stocké a une influence persistante relative. Les éléments les plus récents et les plus fréquents sont les plus influents. Dans la Figure 1, chaque ligne verticale représente un exemplaire. Les lignes continues sont des exemplaires du phonème /ɪ/, tandis que les lignes pointillées sont des exemplaires de /ɛ/. La hauteur des barres est proportionnelle à la force de l'exemplaire. La position des barres sur l'axe des abscisses est fonction de la valeur du second formant de l'exemplaire. Cette représentation selon un seul paramètre acoustique est volontairement simplifiée. Un exemplaire est en réalité associé à de nombreux paramètres acoustiques ; l'espace de ces paramètres est pluridimensionnel.

Lorsqu'un nouvel élément (représenté par l'astérisque sur la Figure 1 : Un nouveau son ayant deux étiquettes possibles : /ɛ/ ou /ɪ/ (tiré de Pierrehumbert, 2001) est rencontré, il doit être catégorisé et trois opérations s'exécutent : le nouvel élément est d'abord soumis à un encodage perceptif qui lui attribue une place dans l'espace paramétrique, ensuite certains exemplaires sont identifiés comme les plus proches en terme de distance dans l'espace paramétrique, puis les exemplaires plus proches sont sélectionnés et, selon leurs étiquettes, les étiquettes les plus probables pour le nouvel élément sont calculées. Dans l'exemple présenté dans la Figure 1, la valeur de F2 de /ɪ/ est en général plus élevée que celle de /ɛ/, mais les exemplaires des deux phonèmes présentent une zone de chevauchement. La voyelle à catégoriser présente donc une position ambiguë entre /ɪ/ et /ɛ/. C'est pourtant le phonème /ɪ/ qui est identifié car il possède plus d'exemplaires actifs dans le voisinage du nouvel élément.

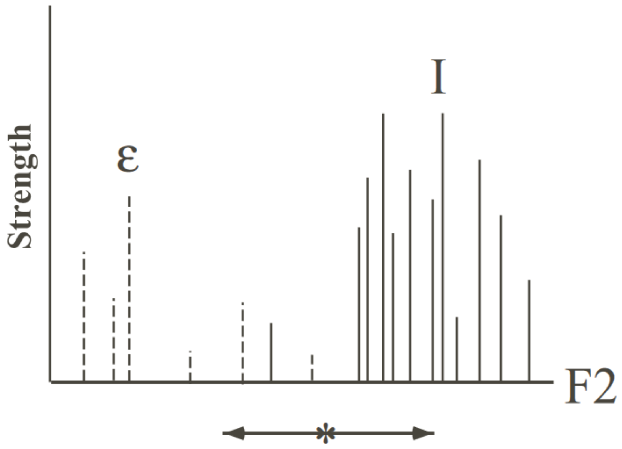

Figure 1 : Un nouveau son ayant deux étiquettes possibles : /ɛ/ ou /ɪ/ (tiré de Pierrehumbert, 2001)

La fréquence, quant à elle, n'est pas encodée de manière explicite dans le modèle, mais elle est intrinsèquement liée aux représentations des catégories cognitives. Les étiquettes les plus fréquentes ont des exemplaires activés de telle sorte qu'ils ont plus d'influence dans les cas ambigus.

Selon Pierrehumbert, les avantages de cette approche sont les suivants : d'abord, l'acquisition des connaissances phonétiques et phonologiques consiste en l'acquisition d'un grand nombre d'exemplaires, ce qui permet de rendre compte des différences extrêmement fines qui existent entre les langues. Ensuite, un élément se trouvant au centre d'un nuage est reconnu dans la catégorie, même si il n'y a pas d'exemplaire de cette catégorie qui se trouve exactement au centre du nuage. Ainsi, il n'y a pas besoin de faire référence à un prototype. Pour une catégorie phonologique, les exemplaires extrêmes sont parfois mieux jugés que ceux qui se situent au centre du nuage. Ainsi les meilleurs exemplaires des voyelles /i/ et /u/, selon une tâche de

jugement par des auditeurs, sont ceux présentant des valeurs de formants plus extrêmes que ceux associées à des productions plus typiques. Étant donné que l'attribution d'une catégorie est contrastive avec les autres, la distance d'un candidat à tous les exemplaires des catégories adverses augmente sa qualité subjective. Ce qui compte n'est pas que ce soit un bon /i/ mais plutôt que cet exemplaire serait un très mauvais /u/. Enfin, les modèles à exemplaires permettent de modéliser les effets de fréquence puisque la fréquence est considérée comme le mécanisme par lequel les souvenirs des catégories sont stockés et les nouveaux exemplaires classifiés.

Pierrehumbert a ainsi développé quatre modèles de la production phonologique. Le premier modèle rend compte de la capacité à produire différentes réalisations des catégories (variations dans l'occurrence produite et variations inter-locuteurs) et de les reconnaitre malgré ces variations. Ce modèle rend ainsi compte de la robustesse des catégories entre les locuteurs malgré la variation de réalisation des catégories. Le modèle ne présente cependant pas de biais systématique alors qu'un tel biais est nécessaire pour rendre compte de la tendance à sous-articuler par économie (Lindblom, 1990).

Un deuxième modèle est alors dérivé du premier en introduisant un biais. Il rend compte du changement phonétique par lénition en la reliant correctement à la fréquence. Par exemple, en anglais, dans *every*, *evening*, qui sont fréquents, le second ‹e› (schwa) est difficile à entendre ([ˈɛvri], [ˈiːvnɪŋ]), alors que dans *mammary*, *artillery*, qui sont rares, le second *a* de *mammary* et le ‹e› de *artillery* sont réalisés explicitement ([ˈmæməri], [ɑːˈtɪlәri]). Le modèle, ainsi modifié, permet les prédictions suivantes : chaque mot a une production variable, l'effet de la fréquence sur le taux de lénition est graduel ; l'effet de la fréquence sur le taux de lénition doit être observé dans le discours individuel (une communauté d'individus n'est pas une moyenne entre les générations qui la compose), l'observation synchronique de la prononciation de mots de différentes fréquences et l'observation diachronique que la prononciation d'un mot par une personne change au cours du temps. Ce modèle prédit également que la distribution doit avoir un biais, faible ou important selon le taux de changement. Un locuteur qui change d'environnement linguistique modifie ses schémas de prononciation seulement graduellement. Les changements historiques prennent des décennies ou des siècles. Ils affectent moins les personnes âgées que les jeunes et ainsi la variation est liée à la coexistence des générations. Pour cela, le modèle suggère deux facteurs possibles : premièrement, les personnes âgées peuvent avoir plus d'exemplaires que les jeunes pour un *pattern* donné, ainsi les exemplaires

anciens dominent (cette explication n'est valide que si le modèle d'oubli a une lente altération). Deuxièmement les personnes âgées apportent moins de nouveaux exemplaires que les jeunes du fait de la lenteur et du manque de robustesse du processus de formation de nouveaux souvenirs avec l'âge. Les statistiques des productions sont donc dominées par les exemplaires appris au plus jeune âge.

Dans un modèle avec du bruit en production, la variance pour une catégorie donnée augmente avec son utilisation ; quand il y a un biais systématique de production, l'amplitude que le biais donne à la distribution aggrave son étalement. Cependant, la pratique montre l'inverse, par exemple, un enfant qui apprend le violon produit une plus grande variabilité de notes au début de son apprentissage et devient de plus en plus précis avec les années de pratique. De la même manière, la variabilité phonétique associée à une certaine catégorie phonologique diminue graduellement jusqu'à la fin de l'enfance (Lee, Potamianos et Narayanan, 1999). Le modèle précédent ne permettait pas de rendre compte de cet effet. Un troisième modèle corrige cela en modifiant la manière dont les productions présentent une certaine variation aléatoire : cette dernière ne consiste plus à choisir un exemplaire au hasard mais à choisir un point au hasard dans l'espace paramétrique. L'intégration est un résultat de ce changement car la nouvelle approche tend à se déplacer plus près de la moyenne de la distribution.

Enfin, une dernière extension du modèle parvient à rendre compte de la neutralisation d'une opposition phonologique : une catégorie marquée, quand elle est assez proche d'une catégorie non marquée, finit par être assimilée par celle-ci.

Pierrehumbert conclut que la dynamique exemplariste, qui suppose que la mémoire soit constituée d'étiquettes d'exemplaires, reproduit plusieurs faits observés en phonologie :

- la possibilité d'apprendre de nouveaux *patterns* phonétiques d'une langue
- la lente évolution du langage d'un adulte
- le fait que la lénition est plus marquée pour les mots fréquents
- le phénomène de neutralisation, pour lequel cette théorie propose un modèle précis.

Cette approche présente un intérêt fondamental. Elle va au-delà de la théorie des prototypes puisqu'elle intègre de manière inhérente et centrale la variabilité des productions, la variabilité inter-locuteurs et l'évolution diachronique.

1.1.1.2 Le Native Language Model de Kuhl

Selon le *Native Language Magnet model* (NLM), le langage et les capacités perceptives sont innés et modelés dès les premiers mois par l'exposition de l'enfant à sa langue maternelle (L1). Cette théorie repose sur des résultats montrant que la perception des adultes et des enfants est altérée ou contrainte par les expositions passées au langage et non par un défaut de plasticité cérébrale. En effet, selon ce modèle, durant la première année de vie, les bébés ont des capacités perceptives non spécifiques au langage et sont donc capables de discriminer tous les sons des langues du monde. Durant cette première année de vie, sous l'effet de l'exposition à leur langue maternelle, ces capacités de discrimination se focalisent sur les sons de la L1 et l'enfant devient ainsi un expert pour percevoir dans sa langue (Kuhl, 1993). Ce modèle repose sur la théorie de l'effet d'aimant perceptif (Kuhl, 1991) qui stipule que tous les exemplaires d'une même catégorie phonologique ne sont pas perçus de manière équivalente. Cette théorie accorde un rôle particulier aux prototypes qui exerceraient un effet d'attraction perceptive sur les sons qui les entourent dans l'espace acoustique, contractant ainsi l'espace perceptif autour de lui comme le montre la Figure 2. L'auditeur aurait tendance à assimiler perceptivement au prototype les sons qui lui sont proches, et il discriminerait donc plus difficilement des sons proches du prototype que des sons qui en sont éloignés. En d'autres termes, des différences entre des exemplaires proches des prototypes seraient plus difficiles à percevoir que des différences acoustiquement équivalentes entre des exemplaires éloignés des prototypes.

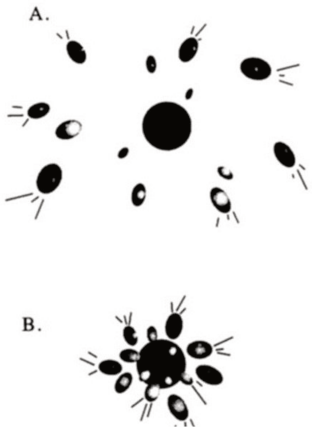

Figure 2 : Représentation de l'effet d'aimant perceptif. En A. l'espace acoustique, en B. l'espace perceptif. Image reproduite de Kuhl et Iverson (1995)

En ce qui concerne plus particulièrement l'acquisition des sons de la L2, ce modèle suppose que, lorsqu'ils sont perceptivement proches de sons de la L1, ils sont d'abord perçus comme

de mauvais exemplaires de catégories de la L1 et apparentés à des catégories de cette L1. Le modèle décrit différents niveaux de difficulté d'intégration L1-L2. En effet, plus un son de L2 est similaire à un son existant dans le système phonologique de la L1, plus le traitement de ce son risque d'être influencé par l'effet d'aimant perceptif, ce qui rendrait d'autant plus difficile la discrimination avec les sons de la L1 et un traitement adapté de ce son. Au contraire, plus un son de L2 diffère des sons existants en L1, plus la discrimination devient facile puisque ce son n'est pas soumis à l'effet d'aimant perceptif. Au sujet de la production de sons de L2, ce modèle considère que la perception précède et altère la production (Iverson & Kuhl, 1995; Kuhl & Iverson, 1995), ce qui expliquerait en partie la persistance de l'accent étranger.

1.1.1.3 Le Perceptual Assimilation Model (PAM) de Best

Le modèle d'assimilation perceptive (Best, 1995) est une extension de l'approche réaliste directe de la perception (Fowler, 1986 ; Gibson & Gibson, 1955), car le principe au centre du modèle dit que la perception est directement créée par l'environnement et qu'elle n'est pas filtrée par des constructions mentales. Dans le cadre du PAM, les représentations mentales des catégories phonétiques ne sont pas nécessaires pour mener à l'apprentissage perceptif. Ce modèle s'oppose aux modèles représentationistes et s'appuie fermement sur le fait que la perception n'est pas efficace à 100%. En effet, durant l'enfance, la perception des sons est guidée par l'inventaire phonologique de la L1 et, par conséquent, identifier des sons familiers devient naturel et écologique. Au début, les sons seraient organisés en termes de groupes de gestes de parole et non pas sous forme de gestes individuels présentés séquentiellement. Cependant, une fois acquise, cette manière de percevoir les sons comme des groupes de gestes, rend difficile la perception des sons de parole qui se singularise sur le plan des gestes et associations de gestes, comme les sons de langues étrangères. Selon Best, les segments non-natifs présentent vraiment des éléments gestuels qui ne correspondent précisément à aucune constellation gestuelle présente en L1.

Best et Tyler (2007) ont étendu leur modèle à la perception des sons de L2 (PAM-L2) et décrivent différents profils d'assimilation perceptive des phonèmes de L2 dans l'inventaire phonémique de la L1. Le PAM-L2 postule que la L1 n'entrave pas nécessairement la perception des sons de la L2. Ce modèle fait même des prédictions sur le degré de réussite de la perception de certains sons de L2 par un apprenant en fonction des degrés de similarité et/ou de dissimilarité avec les sons de la L1. Le PAM-L2 prédit que la catégorisation émerge quand

un son de L2 est perçu comme un son existant en L1, et il prédit une non-catégorisation en cas d'échec d'attribution à une catégorie. En d'autres termes, quand des apprenants de L2 incorporent un nouveau son de L2 à leur système phonologique, ils assimilent ce nouveau son selon les traits acoustiques présents dans leur L1, en percevant cet élément nouveau comme un son préexistant dans leur système phonologique au lieu de créer une nouvelle catégorie. Enfin, une dernière option est envisagée par le PAM-L2 et implique les sons non perçus comme du signal de parole. Ces sons ne seraient alors pas assimilés à des sons des catégories préexistantes en L1. Le PAM-L2 décrit ainsi cinq schémas d'assimilation des sons de L2 non familiers dans l'espace phonologique de la L1 :

1. Assimilation en deux catégories : l'apprenant perçoit deux sons de la L2 en se référant à deux catégories de la L1 (TC : Two-Categories). La discrimination de ces deux sons est prédite comme excellente.
2. Assimilation selon la qualité dans la catégorie (CG : Categories-Goodness) : l'apprenant associe les deux sons de la L2 à une même catégorie de la L1, mais l'un des deux sons est perçu comme un meilleur exemplaire de cette catégorie. Il est prédit que la discrimination de ces deux sons sera de modérée à bonne.
3. Assimilation en une seule catégorie (SC : Single-Category) : les deux sons de la L2 sont perçus comme un même phonème de la L1, mais aucun des deux n'est un très bon exemplaire. Dans ce cas, la discrimination devrait être mauvaise.
4. Assimilation en non-catégorie/catégorie (UC : Uncategorized-Categorized) : l'un des deux sons est assimilé à une catégorie native et l'autre n'est pas catégorisé. La discrimination est censée être de bonne à très bonne.
5. Assimilation non-catégorie/non-catégorie (UU : Uncategorized-Uncategorized) : aucun des deux sons n'est catégorisé. La difficulté de discrimination de ces deux sons dépendra de la façon dont ils sont perceptivement proches entre eux et proches des sons natifs. La discrimination peut varier de mauvaise à excellente.

La Figure 3 schématise ces prédictions :

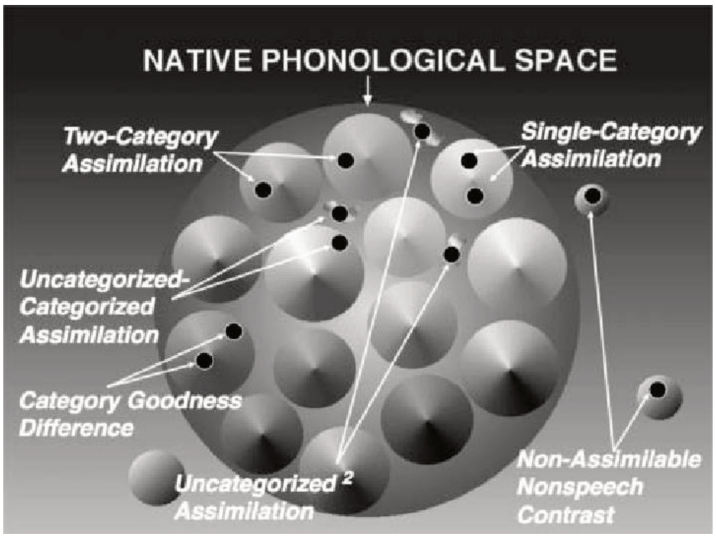

Figure 3 : Représentation schématique du PAM illustrant l'espace phonologique d'un adulte dans lequel les cônes représentent les catégories consonantiques qui ont été délimitées et façonnées par l'expérience avec la perception et la production de la langue maternelle et les principaux patterns d'assimilation perceptive des contrastes consonantiques. Les paires de points noirs représentent des contrastes consonantiques non natifs et les flèches indiquent les patterns d'assimilation prédits par le modèle. (Reproduit de Best et al., 2016)

Selon Best & Tyler (2007), l'expérience d'un locuteur de L1 peut systématiquement affecter la perception d'un contraste de L2. Cependant, ils font l'hypothèse qu'un apprentissage perceptif continu à l'âge adulte sous l'effet de l'exposition à la L2 provoque des réajustements perceptifs. Dans ce modèle, les catégories phonétiques peuvent évoluer pour incorporer les propriétés supplémentaires des sons de L2. Les auteurs ne se positionnent pas clairement sur la question de la précédence de la perception sur la production. Cependant, ils n'écartent pas le fait qu'une perception correcte d'un son de L2 peut faciliter sa production.

1.1.1.4 Le Speech Learning Model de Flege (SLM)

Ce modèle a été développé pour rendre compte des difficultés que présentent les apprenants pour produire des sons de L2 sans accent étranger. Il essaie d'expliquer le phénomène selon lequel plus une L2 est apprise tôt, meilleure seront les capacités en perception et en production de l'apprenant et plus il se rapprochera d'un locuteur natif. Le SLM présente quatre postulats. Premièrement, les mécanismes utilisés pour l'acquisition de la L1 restent intacts tout au long de la vie et peuvent être utilisés pour l'acquisition de la L2. Deuxièmement, les sons de parole ou catégories phonétiques sont stockés en mémoire à long terme et elles sont, troisièmement, établies durant l'enfance qu'ils sont susceptibles d'évoluer au cours de la vie. Le modèle suppose que les catégories phonétiques de la L1 établies dans les premières années de vie sont

modifiées par la suite pour intégrer les propriétés des sons de L1 et L2 identifiés comme des réalisations de chacune des catégories (Flege, 1995, 2003). Il est donc explicitement stipulé que les catégories de L1 et L2 coexistent dans un même espace phonologique et donc s'influencent mutuellement. Quatrièmement, les locuteurs qui parlent plusieurs langues tentent de maintenir des contrastes entre les catégories phonétiques de L1 et L2.

Ce modèle prédit différents niveaux de réussite dans la perception de sons de la L2 en fonction des distances acoustiques entre les catégories phonémiques de L1 et L2. Selon la terminologie employée dans ce modèle, un son de L2 peut être perçu comme « identique », « similaire » ou « nouveau ». Plus un son de L2 est éloigné acoustiquement et perceptivement d'un phonème de L1, plus il a de chances d'être perçu comme un nouveau son et donc de générer la formation d'une nouvelle catégorie, à condition que l'exposition à cette nouvelle catégorie soit suffisante. En effet, Flege et Liu (2001) insistent sur le rôle de l'input dans la formation de nouvelles catégories de L2 et sur le fait que celui-ci doit intervenir en quantité et qualité suffisantes. Les sons considérés dans le modèle comme « identiques », c'est-à-dire non discriminables d'un équivalent en L1, ne sont pas censé poser de problème aux apprenants et ne nécessiteront pas la formation de nouvelles catégories. À l'inverse, plus un son de L2 est perçu comme « similaire » à un son de L1, plus il sera acquis difficilement. L'accent étranger résulterait ainsi de la difficulté à établir des catégories distinctes pour les sons « similaires » en L1 et L2. Le fait des détecter des différences entre les sons de L1 et L2 apparait ainsi comme primordial pour établir des catégories cibles efficaces pour les sons de L2.

Le SLM attribue les erreurs en production principalement à une perception incorrecte. De plus, même si une nouvelle catégorie est générée, cela n'implique pas nécessairement que la production du son en question soit réalisée avec la justesse d'un locuteur natif. Mais le modèle permet que l'exposition à un input suffisant et de bonne qualité permette une évolution favorable.

Le modèle décrit aussi deux mécanismes d'interaction L1/L2 (Flege et al., 2003). Le premier est appelé « assimilation de catégorie phonémique » et prévoit qu'une catégorie de L2 soit assimilée à une catégorie de L1 tant que les occurrences de la catégorie de L2 sont identifiées par la catégorie de L1. Le second mécanisme est appelé « dissimilation de catégorie phonémique » et prévoit que les apprenants de L2 qui parviennent à créer une nouvelle catégorie le long d'une dimension acoustique vont placer la frontière entre la catégorie de L2 et celle de L1 la plus proche aussi loin que possible de cette dernière.

1.1.1.5 Le Second Language Linguistic Perception model d'Escudero (L2LP)

L'hypothèse de la perception linguistique optimale (Escudero et Boersma, 2003) se trouve à la base des trajectoires d'apprentissage en L2 prédites selon le L2LP (Escudero Neyra, 2005 ; Escudero, 2009). Cette hypothèse dit que les apprenants perçoivent d'abord les sons de la L2 de façon équivalente à la production de sons similaires dans leur L1. Ainsi l'auditeur aura tendance à utiliser les indices acoustiques pertinents dans sa langue pour traiter ceux de la L2. Ce modèle propose le résultat de l'acquisition de la L1 comme le point de départ de l'acquisition de la L2, en prédisant que les différences et similarités entre les phonèmes des deux langues vont influencer la trajectoire du développement. À partir de là, trois scenarios sont envisagés. Comme le PAM, ce modèle fait des prédictions sur la perception des contrastes (deux phonèmes) plutôt que sur les phonèmes isolés comme dans le SLM. Les trois scénarios sont schématisés dans la Figure 4 reproduite de Burgos (2018).

1. Quand la majorité des productions d'un contraste de L2 sont acoustiquement proches d'un seul son typique de L1, le L2LP parle d'un *nouveau scenario* (équivalent du *Single Category* du PAM). Dans ce cas, les apprenants doivent soit créer une nouvelle catégorie en L2, soit diviser une catégorie de L1 en deux. Comme le PAM, ce modèle prédit que l'acquisition du contraste sera difficile.
2. Inversement, quand la majorité des productions d'un contraste de L2 sont acoustiquement proches des productions de deux catégories de L1 différentes, les apprenants sont confrontés à un *scenario similaire* (PAM : *two-categories*). Selon le L2LP, les catégories de L1 existantes sont alors répliquées et ajustées pour que leurs frontières viennent s'adapter aux contrastes de L2. Ce scenario est prédit comme moins problématique que la création de nouvelles catégories, ce qui est contraire au modèle de Flege qui prédit que l'acquisition de nouveaux sons est plus facile que celle de sons anciens ou similaires.

3. Enfin le scenario du *sous-set* (PAM : categorized/uncategorized-uncategorized) considère les cas où un son de L2 est perçu comme plus d'une catégorie de L1 (*multiple category assimilation*). Ce scenario est censé poser moins de problème que celui appelé *nouveau scenario* puisqu'il ne nécessite pas la création d'une nouvelle catégorie.

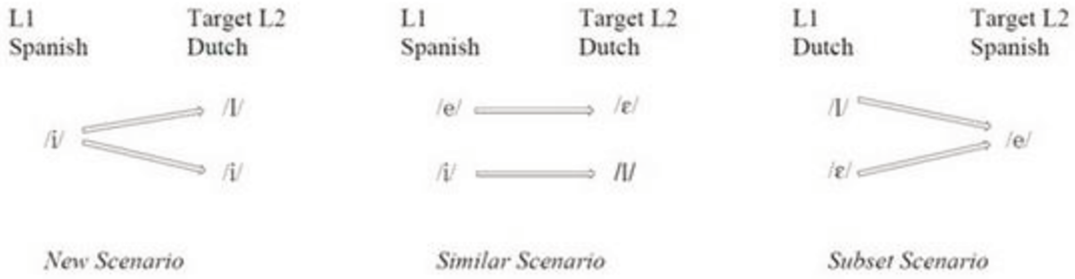

Figure 4 : Nouveau scenario, scenario similaire et scenario du sous-set d'après le modèle L2LP d'Escudero. La schématisation utilise la paire L1-Espagnol/L2-Néerlandais pour les deux premiers scenarios mais la paire L1-Neerlandais/L2-Espagnol pour le scenario du sous-set car ce scenario n'apparait pas pour la première paire.

1.1.2 Modèles liés à la production

1.1.2.1 La théorie motrice de la perception de la parole du laboratoire Haskins

Ce premier modèle est certes un modèle de perception, mais il accorde une place toute particulière à la production dans l'acte de perception et ainsi dans l'acquisition de nouveaux sons de parole. La première affirmation de la théorie motrice de la perception de la parole développée dans Liberman et Mattingly (1985) est que les objets de la perception de la parole sont les gestes phonétiques attendus du locuteur. Ces gestes sont représentés dans le cerveau comme des commandes invariantes qui génèrent des mouvements des articulateurs par le biais de certaines configurations linguistiques significatives. Ces commandes motrices sont la réalité physique sous-tendant les notions de phonétique traditionnelle, comme par exemple l'antériorisation de la langue, l'arrondissement des lèvres ou l'abaissement de la mâchoire, qui permettent de catégoriser les phonèmes. Elles sont les événements élémentaires de la perception et de la production de la parole. Les segments phonétiques sont simplement des groupes (une constellation) d'un ou plusieurs événements gestuels élémentaires. Percevoir un son de parole, consiste à percevoir un *pattern* spécifique de gestes attendus. Ce modèle parle de « geste attendu » car, notamment à cause de la coarticulation, mais pas uniquement, les gestes ne sont pas directement représentés dans le signal acoustique ou dans les mouvements articulatoires observés. Selon les auteurs, il n'est pas simple de définir des gestes spécifiques,

ni de les relier à une conséquence observable. Cependant des gestes invariants doivent exister pour qu'une théorie de la production de la parole puisse être établie.

Le second postulat de la théorie découle du premier : si la perception et la production de la parole partagent les mêmes invariants, elles doivent être intimement liées. Les auteurs supposent que ce lien n'est pas une association apprise qui résulterait du fait que ce que les gens entendent quand ils écoutent de la parole est similaire à ce qu'ils font quand ils parlent. L'originalité du modèle est de proposer que les bases de ce type de lien est, par nature, présent de manière innée, et ne requière qu'un développement épigénétique pour se mettre en jeu. D'après ce postulat, la perception des gestes intervient dans un mode de traitement spécialisé, différent en de nombreux aspects du mode de traitement auditif général, qui est aussi responsable de la production des structures phonétiques et d'une grande partie de la spécialisation du langage. La fonction adaptative du côté perceptif de ce mode est de faire la conversion automatique du signal acoustique en gestes pour permettre au locuteur de percevoir les structures phonétiques sans médiation par les apparences auditives que les sons pourraient avoir sur la base d'informations purement psychoacoustiques. Enfin, la troisième affirmation est que le système moteur est recruté lors de la perception de la parole, ce qui est cohérent avec des données empiriques issues d'études menées en imagerie cérébrales mettant en évidence des activations du cortex moteur lors de tâches de perception de la parole (e.g., Callan, Callan, Gamez, Sato, & Kawato, 2010 ; Pulvermüller et al., 2006 ; Wilson, Saygin, Sereno, & Iacoboni, 2004).

1.1.2.2 Le modèle DIVA de Guenther

Le modèle DIVA (*Directions Into Velocities of Articulators*) développé par Guenther (1994) est un modèle de production de la parole qui propose une approche computationnelle et neuroanatomique du réseau de régions cérébrales impliquées dans l'acquisition et la production de la parole (Tourville et Guenther, 2011). DIVA est un réseau de neurones adaptatif qui décrit les interactions sensorimotrices du contrôle articulatoire dans la production de parole. Le modèle est schématisé dans la Figure 5. Chaque bloc correspond à un ensemble de neurones qui constituent une représentation neuronale. Dans la description du modèle, le terme de « carte » est utilisé pour désigner de tels ensembles de cellules représentés par les boîtes. Les flèches schématisent les transformations d'une représentation neuronale à une autre. Ces transformations sont effectuées par le filtrage des activations d'une carte par des synapses

projetées sur une autre carte. Les poids synaptiques sont appris au cours d'une phase de babillage censée représenter globalement celle généralement vécue par un nourrisson au cours de son développement normal. Pendant cette phase, les mouvements aléatoires des articulateurs de la parole fournissent des signaux de retour tactiles, proprioceptifs et auditifs qui sont utilisés pour apprendre les correspondances entre les différentes représentations neuronales. Après avoir babillé, le modèle peut rapidement apprendre à produire de nouveaux sons à partir d'échantillons audio qui lui sont donnés en entrée.

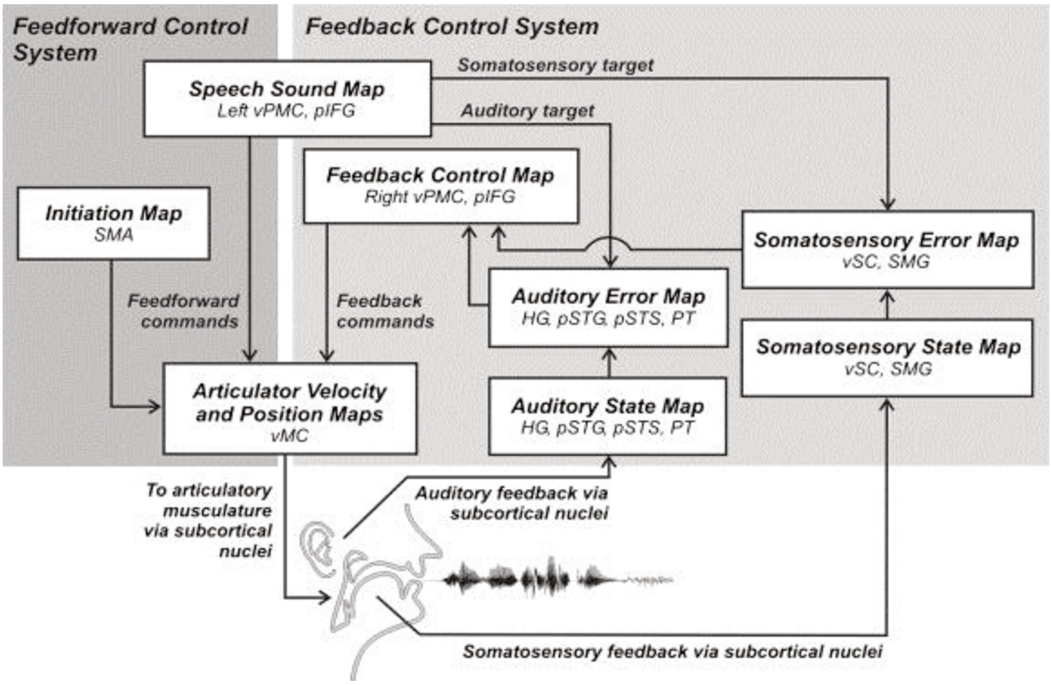

Figure 5 : Schématisation du modèle DIVA (reproduit de Guenther, 2006)

Dans ce modèle, la production d'un son de parole débute par l'activation des neurones associés à ce son dans la *carte du son* du modèle (*Speech Sound Map*). Un son de parole peut être un phonème, une syllabe ou même une séquence de courtes syllabes. L'activation des neurones de la *carte du son de parole* produit une commande motrice qui arrive dans le cortex moteur primaire via deux sous-systèmes. *Le système de contrôle par anticipation* (*Feedfoward Control System*) projette directement à partir de la carte du son de parole jusqu'aux unités du contrôle articulatoire dans le cervelet et dans le cortex moteur primaire. Le système de rétrocontrôle (*Feedback Control System*), lui-même composé d'un sous-système de rétro-contrôle auditif et d'un sous-système de rétro-contrôle somatosensoriel, implique des projections plus lentes et indirectes qui traversent les aires sensorielles du cerveau.

Le modèle DIVA décrit comment les enfants apprennent à produire des sons grâce au babillage et à l'imitation. Selon ce modèle, une combinaison d'informations motrices, auditives et somatosensorielles générées durant le babillage aléatoire et répété est utilisée pour façonner les projections synaptiques entre les cartes des erreurs sensorielles et le cortex moteur via une carte de rétro-contrôle dans le cortex prémoteur ventral droit. Plus tard, dans l'apprentissage par imitation, les cartes d'erreurs stockent les écarts entre les états sensoriels attendus et produits. Les transformations sensori-motrices apprises durant le babillage permettent aux erreurs sensorielles détectées d'être intégrées dans des commandes motrices correctives pendant le stade de l'imitation.

L'étape d'imitation décrit comment se déroule l'apprentissage spécifique à une syllabe lorsqu'un nouveau son de la langue maternelle à apprendre est présenté à un nourrisson. La détection d'un nouveau son provoque une action des neurones d'une nouvelle carte de son qui n'a encore jamais été activée. Le premier modèle apprend une cible auditive pour le nouveau son, représenté comme un signal acoustique variant dans le temps. Cette cible auditive est encodée dans les projections synaptiques depuis la carte du son de parole vers la carte des erreurs auditives. La cible encode la variabilité autorisée dans le signal acoustique tout au long de la durée de la syllabe. L'utilisation de régions-cibles, plutôt que de points, est un aspect important du modèle qui permet une explication unifiée pour un large spectre de phénomènes de production de parole. Les neurones de la carte du son de parole qui représentent le nouveau son peuvent également être utilisés pour produire le son ; ils sont activés lorsque le son est perçu et quand ce même son est produit.

Dans l'étape suivante du processus d'apprentissage par imitation, le modèle tente de produire un son en activant les neurones de la carte du son de parole appropriée. Cela conduit à la lecture d'une commande d'anticipation. Lors de la première tentative pour produire un nouveau son, aucune commande d'anticipation ajustée pour ce son n'existe. Le modèle prédit alors que la lecture de cette commande d'anticipation produira une erreur auditive et que le système devra utiliser le sous-système de rétrocontrôle pour aider à former la prochaine tentative de production du son, en transformant les erreurs auditives en commandes motrices correctives via la carte de rétro-contrôle dans le cortex prémoteur ventral droit. L'activité de la cellule de l'erreur auditive représente l'erreur dans l'espace des fréquences de formants ; cette erreur doit être transformée en une commande corrective de la représentation motrice ou articulatoire. La

transformation de l'erreur auditive en une commande motrice corrective est apprise pendant le cycle de babillage du modèle.

À chaque tentative de produire le son, les signaux de commande du sous-système de commande par anticipation sont mis à jour et intègrent les commandes redéfinies qui ont été générées par le sous-système de rétro-contrôle auditif pour cette tentative. Cela conduit à une commande par anticipation plus efficace pour la prochaine tentative. Ensuite, la commande par anticipation suffit à elle seule pour produire le son dans des circonstances normales. C'est-à-dire que la commande par anticipation est suffisamment précise pour générer peu ou pas d'erreur auditive lors de la production du son : elle ne provoque donc pas l'activation du sous-système de rétro-contrôle auditif. À ce niveau, le nouveau son peut être produit de manière suffisamment fluide. Si la production du son est répétée, une région cible somatosensorielle pour ce son est apprise, comme pour la région cible auditive. Cette cible représente les sensations tactiles et proprioceptives attendues associées au son et est utilisée dans le sous-système de rétro-contrôle somatosensoriel pour détecter les erreurs somatosensorielles.

1.2 Acquisition L2 et potentiels évoqués

De nombreuses recherches en psycholinguistique ont cherché à comprendre comment le cerveau humain traite et intègre les différents éléments linguistiques qu'il reçoit pour en comprendre le sens. Si les locuteurs natifs ou des apprenants font face à des difficultés en compréhension, il est nécessaire d'étudier le décours temporel très précis des traitements réalisés pour comprendre ce qui se passe exactement et à quel moment le problème survient. La plupart des traitements linguistiques intervenant dans la seconde après le début du stimulus linguistique, les méthodes comportementales ne sont pas suffisamment précises du point de vue temporel. Elles sont limitées à un seul point de données par item et ne peuvent être utilisées que chez certaines populations en mesure de coopérer. De plus, les comportements observables par ces méthodes peuvent être influencés par des processus méta-linguistiques ou liés à la tâche à réaliser (e.g., prise de décision, action motrice). Pour contourner ces difficultés, la méthode des potentiels évoqués s'avère de premier choix. En effet, l'étude électrophysiologique des potentiels évoqués (*Event Related Potentials,* ERP) fournit des ondes cérébrales reflétant l'activité neuronale sans délais, avec une précision temporelle de l'ordre de la milliseconde, tout au long du stimulus et pour tout type de populations (e.g., enfants, patients, etc.). Elle ne nécessite pas forcément la réalisation d'une tâche particulière par les participants et permet de

distinguer différents processus cognitifs. Techniquement, les potentiels évoqués sont obtenus en enregistrant un signal EEG (électroencéphalographique) par l'intermédiaire d'électrodes placées à la surface du crâne, ce qui a l'avantage d'être une méthode relativement simple à mettre en place et non-invasive. Le signal est ensuite synchronisé à l'apparition des stimuli afin d'extraire les ondes consécutives aux stimuli. La méthode des potentiels évoqués en EEG permet de recueillir l'activité électrique spécifique du cerveau liée à un stimulus interne (e.g., prise de décision, préparation motrice) ou externe, transmis par un canal sensoriel (e.g., stimulation auditive ou visuelle) (Luck, 2014). Cette activité électrique étant extrêmement faible (de l'ordre du microvolt), il est nécessaire de répéter l'enregistrement un très grand nombre de fois afin de moyenner toutes ces mesures pour augmenter le rapport signal sur bruit. Les potentiels évoqués sont (le plus souvent) représentés par le biais de trois paramètres : l'amplitude de l'onde, sa latence (i.e., temps d'occurrence le plus précoce après le stimulus) et sa polarité. Les ondes sont généralement nommées en fonction de leur polarité, positive (P) ou négative (N), et de leur latence (Fonteneau, Frauenfelder et Rizzi, 1998). La plupart des traitements psycholinguistiques interviennent dans les 1000 ms après le début d'un mot et génèrent des ondes positives et négatives avec des latences spécifiques. Par exemple, la discrimination automatiques de deux phonèmes génère une onde négative de distribution frontale aux alentours de 200 ms appelée la *MisMatch Negativity* (MMN), alors que des difficultés de traitement syntaxique sont souvent reflétées par une négativité antérieure gauche (LAN) entre 300 et 500 ms suivie d'ondes positives postérieures autour de 600 ms (P600) (Kaan, 2007 ; Luck, 2014). Le profil des potentiels évoqués diffère selon leur distribution sur le scalp, leur amplitude et leur durée. Il faut toutefois noter que la distribution de l'activation sur le scalp n'indique pas la localisation de la source d'activité, mais est plutôt le reflet de l'orientation des neurones impliqués. Cette absence de relation directe entre la distribution de surface d'un composant ERP et l'emplacement de ses générateurs neuronaux dans le cerveau est une limite des ERP par rapport à l'imagerie cérébrale fonctionnelle (IRMf), mais leur résolution temporelle est de l'ordre de la milliseconde quand celle de l'IRMf est de l'ordre de la seconde. Elle permet donc d'appréhender la dynamique temporelle des processus langagiers en temps réel.

Dans l'étude du langage, les principales composantes étudiées en potentiels évoqués sont la N200, générée lors du traitement d'informations phonologiques, le complexe N1-P2, la N350, pour les traitements lexicaux, la N400, révélant des processus de traitement sémantique, la P600 et l'ELAN (*Early Left Anterior Negativity*) générées lors du traitement d'informations

syntaxiques, la MMN (Fonteneau et al., 1998 ; Pulvermüller & Shtyrov, 2006) que nous détaillerons plus loin et la P300 qui est associée à des processus décisionnels (Isel, 2017). Parmi ces potentiels évoqués, la MMN est l'onde de prédilection pour étudier la perception des phonèmes.

1.2.1 Les potentiels évoqués liés à des traitements phonologiques

1.2.1.1 Le complexe N1-P2

Le complexe N1-P2 est un potentiel évoqué obligatoire indice du traitement des infirmations sensorielles (Isel, 2017) qui peut refléter les représentations auditives centrales de la parole et qui est indépendant de l'attention (Martin et al., 1997 ; Sharma & Dorman, 1999, 2000). La N1 atteint son amplitude maximale dans des régions fronto-centrales. Selon la durée du stimulus, la latence du pic de cette onde se situe autour de 100 ms et il est suivi par une positivité − la P2 − présentant un pic autour de 175 ms. Ce complexe est censé refléter l'activation synchrone de neurones en réponse à une stimulation auditive. Plusieurs sources contribuent à l'activité enregistrée à la surface du scalp dans la période temporelle de ce complexe. Cependant, des études ont montré que la source probable de cette activité se situerait à proximité du cortex auditif primaire (Näätänen et Picton, 1987).

1.2.1.2 La *Mismatch Negativity*

La « *Mismatch Negativity* » (MMN) ou « négativité de discordance » est une composante électroencéphalographique fronto-centrale négative reflétant la détection de tout changement discriminable dans une séquence de stimuli auditifs répétitifs (Pulvermüller et Shtyrov, 2006). Elle est le plus souvent étudiée avec le paradigme *oddball*, qui consiste en la présentation de stimuli déviants rares parmi des stimuli standards répétés/fréquents. La MMN est mise en évidence en calculant la différence entre la réponse évoquée par le stimulus déviant et celle évoquée par le stimulus standard. C'est une réponse automatique et indépendante de l'attention (i.e., elle est générée que le participant soit attentif ou non aux stimuli). Les études montrent qu'elle est générée essentiellement dans le cortex temporal supérieur (Pulvermüller et Shtyrov, 2006). Le son répétitif (standard) forme une trace mnésique dans le système auditif, et si un nouveau son (déviant) ne correspond pas à cette trace, alors la MMN apparaît environ 50 ms

après le début du stimulus déviant et présente généralement un pic à 100-250 ms (Näätänen et Picton, 1987). La MMN est le signe d'un processus perceptif de comparaison automatique du stimulus rare/déviant et d'une représentation neuronale en mémoire formée par le standard. La localisation et la latence de la MMN varient en fonction des caractéristiques des stimuli auditifs (Pulvermüller et Shtyrov, 2006). L'amplitude de la MMN diminue avec l'augmentation de la probabilité d'apparition des stimuli déviants et avec la diminution des différences entre stimuli standards et déviants, comme le montre la Figure 6 (Näätänen et Picton, 1987). Cette composante est sensible à des changements de différentes natures dans les caractéristiques des stimuli auditifs : acoustiques, phonologiques, lexicaux, sémantiques ou syntaxiques (voir (Kujala, Aho, et al., 2007; Pulvermüller & Shtyrov, 2006 pour des revues). Lorsque des différences au niveau phonologique sont perçues, la MMN se manifeste avec une latence d'environ 100 à 200 ms après le début du son déviant et est générée par une source située dans le cortex temporal supérieur (Näätänen et al., 1997). Pulvermüller et Shtyrov (2006) reconnaissent trois avantages à l'étude de la MMN : son automaticité, sa précocité, et la perspective qu'elle ouvre sur le suivi des réponses du cerveau à des éléments linguistiques isolés.

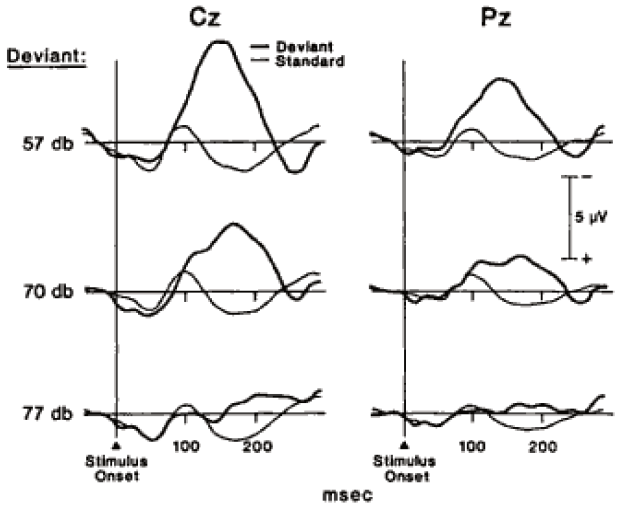

Figure 6 : Les effets de la déviance sur la MMN pour les électrodes Cz et Pz (Image tirée de Näätänen et Picton, 1987). Un stimulus standard (80 dB) était présenté dans 90% des cas et un stimulus déviant (57, 70 et 77 dB dans des blocs différents) dans 10% des cas. Les potentiels évoqués liés aux stimuli standards sont représentés par le trait fin, ceux liés aux stimuli déviants par le trait épais. L'amplitude de la MMN augmente avec la différence acoustique (intensité) entre les standards et les déviants.

Aaltonen et collègues (1987) ont comparé les MMN générées dans un paradigme *oddball* lors de la présentation de paires de stimuli composées soit des voyelles aux extrémités du continuum finnois /i/-/y/ (deux phonèmes), soit d'une de ces voyelles et d'une voyelle à mi-chemin entre les deux pôles du continuum. Ils ont constaté que la MMN présentait une

amplitude plus grande et une latence plus courte dans la condition phonémique (i.e., voyelles aux extrémités du continuum) que dans la condition où la paire de stimuli était formée d'une des deux voyelles extrêmes du continuum et d'une voyelle située à mi-chemin des extrémités du continuum. Ces résultats montrent que pour les sons de parole, la MMN présente une amplitude plus grande et une latence plus courte lorsque le déviant n'appartient pas à la même catégorie phonémique que le standard. Näätänen et collègues (1997) ont comparé lors d'une étude ERP les réponses MMN d'auditeurs finlandais à l'écoute de la voyelle finnoise /ö/ ou de la voyelle estonienne inconnue /õ/ dans une séquence de stimuli standards constitués par la voyelle finnoise /e/. Ils ont pu montrer que l'amplitude de la MMN était plus grande lorsque la voyelle finnoise /ö/ était présentée en tant que stimulus déviant par rapport à la condition dans laquelle le stimulus déviant était la voyelle estonienne /õ/, et ce malgré le fait que la distance acoustique entre les voyelles /õ/ et /e/ soit plus grande que celle entre les voyelles /ö/ et /e/. Ces résultats tiendraient au fait que la voyelle /ö / est présente dans la langue maternelle des sujets contrairement à la voyelle /õ/. De la même manière, Dehaene-Lambertz (1997) a étudié des contrastes consonantiques à l'aide d'un paradigme *oddball*. Elle a utilisé un continuum /ba - da - ɠa/ variant sur le VOT. Elle a inséré au sein d'une série répétée de la même syllabe, des déviants appartenant soit à la même catégorie phonémique que les standards *(e.g.,* /ba/1- /ba/1- /ba/1-**/ba/5**), soit à une catégorie différente (/ba/1 et /ba/5 étant des stimuli différents de la catégorie /ba/ sur le continuum). Pour les déviants appartenant à une catégorie phonémique différente, elle a étudié deux frontières phonémiques, l'une présente en français (*e.g.*,/da/-/da/-/da/- /**ba**/) et l'autre non *(e.g.,* /da/-/da/-/da/-/**ɠa**/, un contraste présent en hindi). Chez les auditeurs français, une MMN était générée seulement pour le contraste inter-catégoriel pertinent dans la langue maternelle *(*/da/-/ba/). La réponse électrophysiologique est donc plus importante lorsque la différence entre le stimulus déviant et le standard correspond à un contraste phonémique présent dans la langue maternelle du sujet.

Du point de vue méthodologique, la plupart des études en MMN menées avec des sons de parole ont utilisé un seul exemplaire de chaque catégorie phonémique qui était acoustiquement constant. Ainsi, la MMN observée pourrait être le reflet soit de la perception du changement phonémique entre catégories, soit simplement celui de la perception d'un changement purement acoustique. En réalité, les deux types de changements contribuent à la MMN générée, les changements acoustiques générant une MMN distribuée bilatéralement alors que les changements purement phonétiques génèrent une MMN dont la source se situe dans l'hémisphère gauche pour la plupart des sujets (Näätänen et al., 1997). Afin d'éliminer la

composante acoustique, des chercheurs ont mis en place des paradigmes innovants dont voici quelques exemples. Ghislaine Dehaene-Lambertz & Pena (2001) ont utilisé des stimuli produits par trois femmes différentes pour constituer les trois exemplaires d'un standard (/ta/) présentés devant un déviant (/pa/ et inversement) qui était produit par une quatrième femme. Shestakova et collaborateurs (2002) ont utilisé un paradigme *roving-standard* (standard volant) dans lequel les stimuli des trois catégories de voyelles /a/, /u/ et /i/ étaient constitués de 150 occurrences produites chacune par un homme différent. Et pour dernier exemple, l'étude de Miglietta, Grimaldi et Calabrese (2013) qui ont mis en place un paradigme *oddball* classique comportant trois exemplaires différents des voyelles /ɛ/, /e/ et /i/ produites par un même homme. Cette étude présente également la particularité d'avoir étudié la iMMN (MMN *identity*) qui, à la différence de la MMN classiquement calculée en faisant la différence entre les réponses évoquées par les stimuli standards et déviants d'un même bloc, est calculée comme la différence entre un même stimulus lorsqu'il est présenté en tant que standard et en tant que déviant. Cette approche est censée éliminer dans la morphologie des réponses évoquées la variation résultant des différences acoustiques pures et doit donc permettre d'observer les contributions des représentations en mémoire (Pulvermüller et Shtyrov, 2006).

1.2.1.3 Les autres potentiels évoqués candidats à l'étude des traitements phonologiques

Les deux potentiels évoqués précédemment décrits sont les plus fréquemment utilisés pour étudier les processus de traitement phonologiques. Cependant d'autres potentiels évoqués se sont avérés sensibles au traitement d'informations phonologiques, comme par exemple la P3a, la LDN (*Late Discriminative Negativity*) ou même la N400.

La P3a est une composante qui apparait parfois après la MMN avec une topographie fronto-centrale et une latence allant de 250 à 400 ms après le début du changement (Picton, 1992). Pour que la P3a soit présente, il faut que le stimulus déviant soit suffisamment différent et parvienne à attirer l'attention quoique à un niveau non contrôlé délibérément. Ainsi, la P3a est supposée refléter une redirection involontaire de l'attention suite à la détection d'un changement pertinent. L'amplitude de la P3a augmente pendant les processus d'acquisition de la L2, suggérant que, quand un nouveau contraste phonémique est établi, il devient plus pertinent et plus à même d'attirer l'attention et l'engagement de ressources de traitement.

La LDN est générée en réponse à un stimulus déviant parmi des standards également, qu'il s'agisse de sons de parole ou non. Elle présente un pic d'amplitude entre 300 et 600 ms après le début du changement et a une distribution fronto-centrale. Différents points de vue quant à son interprétation coexistent dans la littérature : elle pourrait refléter l'évaluation du stimulus à un niveau cognitif pré-attentionnel. Du fait qu'elle présente une latence plus élevée que la MMN, elle ne devrait pas refléter des traitements purement sensoriels. Bien que certaines études portent sur cette onde chez l'adulte (e.g., Jakoby et al., 2011), elle est le plus souvent analysée chez l'enfant ; son amplitude décroit avec l'âge (Cheour, Korpilahti, Martynova et Lang, 2001). Enfin, la N400, bien que surtout associée à l'étude de processus sémantiques, peut refléter l'intégration d'informations phonologiques dans des paradigmes impliquant des paires minimales. La N400 est une onde négative dont l'amplitude maximale est atteinte 400 ms après le début d'un mot, présenté visuellement ou oralement. L'amplitude de cette onde est modulée par la difficulté d'intégration du mot dans le contexte qui le précède (Kutas et Hillyard, 1980).

1.2.2 Potentiels évoqués et perception des sons de L2

Les potentiels évoqués ont également été largement utilisés pour étudier l'acquisition de contrastes en langue seconde. Nenonen, Shestakova, Huotilainen et Näätänen (2003) ont mené une étude pour examiner le traitement de la durée de stimuli de parole et de sons non verbaux chez des enfants finlandais et russes qui étaient arrivés en Finlande après l'âge de 4 ans. Les enfants étaient âgés de 10 à 14 ans au moment de l'expérience. La distinction de durée est phonologique entre les voyelles courtes et longues en finnois ce qui n'est pas le cas en russe. Les auteurs ont présenté les syllabes longues et courtes /ka:/ et /ka/ aux enfants et des tons complexes longs et courts comme standards et déviants dans différents blocs. La différence de durée était de 50 ms dans les deux conditions (parole et ton). Alors que la réponse MMN ne présentait pas de différence entre les deux groupes d'enfants pour la condition avec les tons, elle était de plus faible amplitude pour les stimuli de parole dans le groupe des enfants nativement russophones (voir Figure 7). Les auteurs ont interprété cette différence comme une preuve de similarité de la détection de changements acoustiques entre les participants de L1 et de L2. Cependant, lorsqu'il s'agit de sons de parole, le traitement de la durée est inhibé à un

niveau pré-attentionnel chez les locuteurs dont la langue maternelle ne dispose pas de ce contraste, et ce même si les locuteurs parlent couramment leur L2.

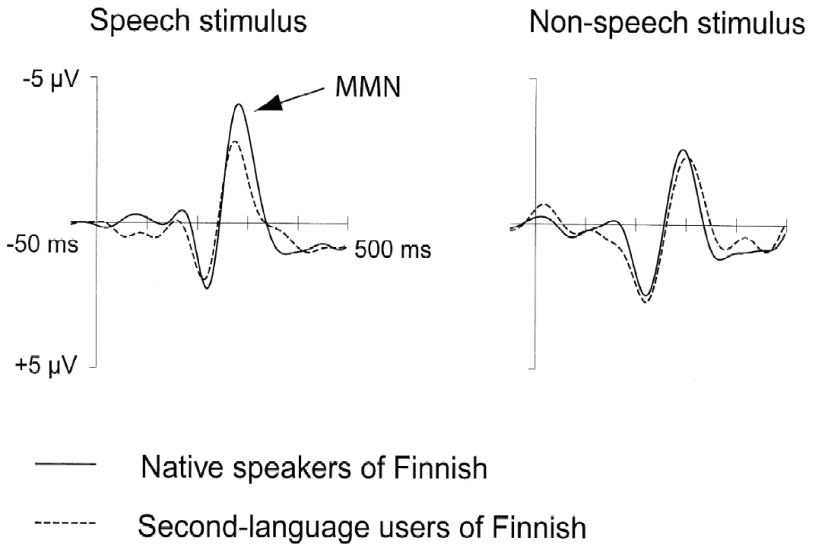

Figure 7 : Réponses MMN observées pour les stimuli langagiers à gauche et les tons à droite. Les réponses des enfants finlandais sont représentées en trait continu, celles des enfants russes en trait pointillé. Image reproduite de Nenonen et al. (2003).

Winkler et collègues (1999) ont montré la nature dynamique des représentations des phonèmes avec une étude sur la MMN. Dans cette étude, des participants hongrois (locuteurs naïfs ou avancés en finnois) et finnois ont été soumis à un paradigme *oddball* permettant de comparer la MMN générée pour un contraste pertinent en finnois mais pas en hongrois (/æ/-/e/) et celle générée pour un contraste pertinent dans les deux langues (/y/-/e/). Ils ont montré que la MMN pour le contraste spécifique au finnois était présente pour les participants hongrois fluents en finnois de la même manière que chez des locuteurs natifs, mais pas pour le groupe naïf en finnois. Ces résultats montrent que les participants hongrois qui parlent couramment le finnois ont développé des représentations corticales pour les phonèmes du système finnois qui leur permet de catégoriser les phonèmes spécifiques à cette langue de manière pré-attentionnelle. Dans cette étude, les analyses portaient seulement sur l'électrode fronto-centrale Fz, ce qui a pu masquer des différences entre les locuteurs fluents en finnois L2 et les locuteurs natifs. En effet, Zevin et collaborateurs (2010) ont mené une étude pour déterminer s'il est possible de trouver des différences sur la MMN entre des locuteurs natifs d'une langue et des apprenants L2 avancés pour des contrastes de cette langue. Ils ont observé la MMN pour le contraste anglais /r/-/l/ chez des locuteurs anglais natifs et des locuteurs japonais parlant couramment l'anglais. Comme dans Winkler et al. (1999), les auteurs n'ont pas montré de différence significative sur la MMN analysée sur le pôle fronto-central classique Fz, alors même que des

mesures comportementales montraient que les locuteurs japonais avaient des performances nettement inférieures aux natifs en discrimination des phonèmes /r/ et /l/ marqués par des années d'exposition à la langue. Cependant une analyse topographique et une analyse de la source sur la fenêtre temporelle de la MMN ont montré des différences entre les locuteurs natifs et les locuteurs de L2 en termes de latéralité des réponses. En effet, ils ont observé que les réponses des locuteurs natifs étaient davantage localisées dans l'hémisphère gauche alors que celles des locuteurs de L2 l'étaient davantage dans l'hémisphère droit.

Peltola et collègues (2003) ont étudié le développement d'une nouvelle catégorie de voyelles à l'âge adulte dans une étude en potentiels évoqués. Pour cela, ils ont observé la MMN chez trois groupes de participants : des participants de langue maternelle finnoise ayant peu de connaissances en anglais, des participants de langue maternelle finnoise étudiants avancés en anglais et des locuteurs anglais natifs. Le signal EEG des participants était enregistré alors qu'ils étaient soumis à un paradigme *oddball* avec une fréquence d'apparition des déviants de 14%. Les stimuli étaient des voyelles de synthèse stables ayant été jugées comme de bons exemplaires par des locuteurs natifs. Quatre blocs ont été présentés : Finnois /i/ - /e/, Anglais /i/ - /e/, Anglais /i/ - /ɪ/ et Anglais /e/ - /ɪ/. La plus petite distance euclidienne entre les voyelles des paires était celle de la paire anglaise /i/ - /ɪ/ (108 Mel), la plus grande celle pour la paire anglaise /i/ - /e/ (272 Mel) et les deux autres paires présentaient des différences presque identiques (185 et 182 Mel). Pendant la présentation des blocs, les participants regardaient un film silencieux et devaient ignorer les stimuli présentés auditivement. Les résultats, présentés dans la Figure 8, montrent que la MMN est plus grande pour les contrastes natifs comparés (i.e., /i/-/e/ finnois pour les Finlandais et /i/-/e/ anglais pour les locuteurs anglais natifs) à ceux de langues étrangères à distance acoustiques comparables (i.e., /i/-/e/ anglais pour les Finlandais et /i/-/e/ finnois pour les locuteurs anglais natifs). Ils montrent également que les étudiants en anglais ne présentent pas des réponses MMN semblables à celles des locuteurs natifs pour le contraste anglais /e/-/ɪ/. Les auteurs concluent que l'enseignement formel reçu par les étudiants n'est pas suffisant pour former une trace en mémoire à long terme comparable à celles de contrastes natifs. Lorsqu'ils ont comparé les réponses aux voyelles natives des étudiants à celles des deux groupes monolingues, les auteurs ont également remarqué qu'elles n'étaient pas similaires. Cela est interprété comme le signe d'un processus d'apprentissage non terminé capable d'affecter néanmoins l'accès aux représentations dans la langue maternelle. En effet, les auteurs interprètent la faible amplitude de la MMN observée dans ce groupe, que ce soit pour les contrastes natifs et étrangers, comme un corrélat neuronal de l'hypothèse de la

confusion (Caramazza, Yeni-Komshian, Zurif et Carbone, 1973) qui réfère à l'étape de l'apprentissage pendant laquelle les deux systèmes sont encore entrelacés. Les distances acoustiques entre les stimuli déviants et standards semblent jouer un rôle plus important chez les apprenants, car le contraste le plus saillant génère une MMN plus grande ce qui n'est pas le cas chez les participants monolingues. En effet, chez ces derniers, la réponse MMN dépend davantage du statut phonémique du contraste.

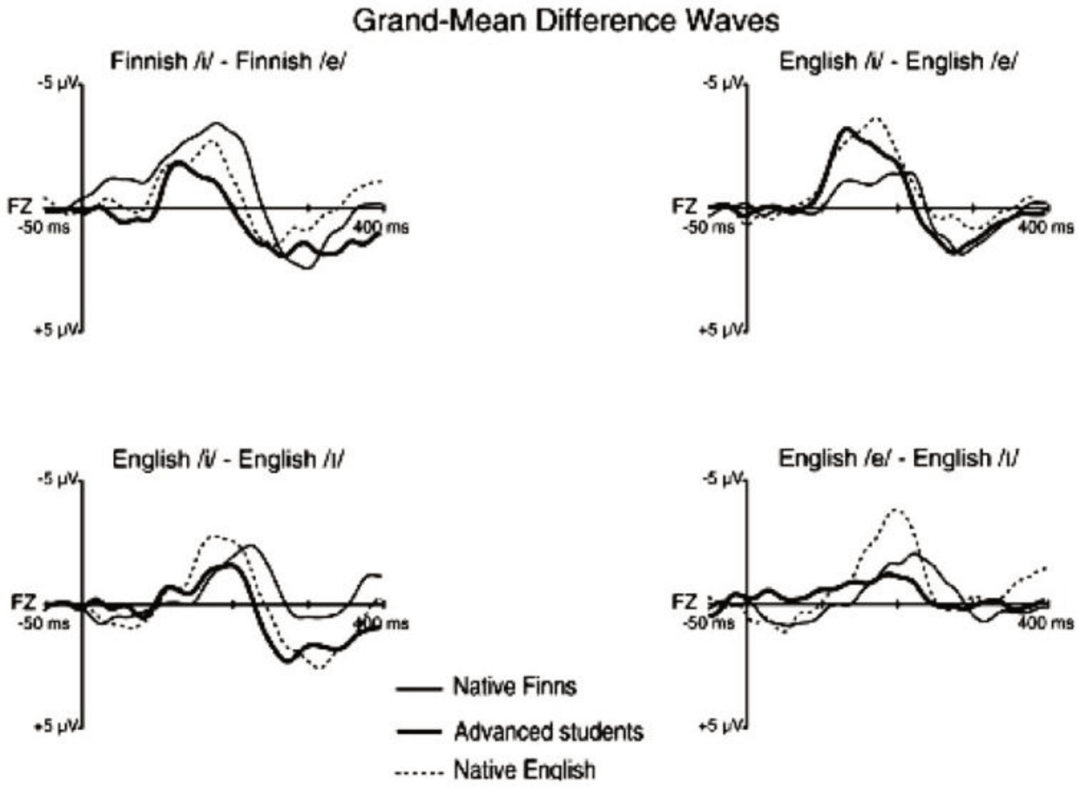

Figure 8 : Grandes moyennes des différences entre déviants et standards dans les quatre blocs de voyelles pour les locuteurs finnois monolingues en ligne fine, les étudiants finnois en ligne épaisse et les locuteurs anglais monolingues en pointillés. Image reproduite de Peltola et al., (2013).

Frenck-Mestre, Peri, Meunier et Espesser (2010) ont étudié la perception des voyelles de l'anglais américain par des locuteurs anglais natifs, français monolingues et bilingues français-anglais ayant appris l'anglais tardivement. Leurs résultats montrent que les participants bilingues présentent des corrélats EEG intermédiaires entre les locuteurs anglais natifs et les locuteurs monolingues français pour la perception de voyelles de l'anglais américain. Ils ont utilisé un paradigme *oddball* dans lequel le standard était la voyelle /ɛ/, le déviant était la voyelle /ɪ/ (15%) et les participants devaient détecter la cible /æ/ présentée dans 10% des items pour une expérience, alors que cible et déviant étaient inversés dans une autre expérience. Leurs résultats montrent que, pour tous les groupes, dans les 100 ms qui suivent la présentation des trois voyelles, les potentiels évoqués (N100) présentent des amplitudes corrélées aux

différences acoustiques entre les voyelles. L'analyse de la P300 a montré des différences entre les groupes. Les participants bilingues montraient une P300 pour la voyelle /ɪ/ présentée comme déviant (i.e., attention dirigée vers /æ/) similaire à celle des locuteurs anglais natifs en termes d'amplitude et de distribution, contrairement aux français monolingues. Quand le /ɪ/ était la cible et /æ/ le déviant, les participants monolingues présentaient une inversion de la P300 pour ces voyelles (i.e., P300 plus grande pour la cible) alors que les bilingues ne pouvaient désengager leur attention et montraient une P300 équivalente pour les deux voyelles. Ces résultats montrent que bien que les apprenants bilingues tardifs établissent des nouvelles catégories pour les voyelles de L2, ces catégories ne sont pas aussi stables que celles de la L1.

2 Les entraînements pour l'acquisition de la L2

2.1 Les entraînements perceptifs

2.1.1 Les débuts

Dès les années 1980, après des expériences menées en perception de la parole inter-langues montrant que des apprenants de L2 ne percevaient pas les contrastes de L2 de la même manière ou avec la même précision que des locuteurs natifs, des recherches ont débuté pour savoir s'il était possible d'entraîner des locuteurs à mieux percevoir des sons de parole non natifs (e.g., Best & Tyler, 2007 ; Jamieson & Morosan, 1986 ; McClaskey, Pisoni, & Carrell, 1983 ; Pisoni, Aslin, Percy, & Hennessy, 1982 ; Strange & Dittmann, 1984).

Lively, Logan et Pisoni (1991) proposent alors un entraînement constitué de stimuli contenant une certaine variabilité des locuteurs et des environnements phonétiques (*High Variability Phonetic Training* ou *HVPT*). Ils formulent l'hypothèse selon laquelle cette variabilité pourrait aider les apprenants à extraire les paramètres acoustico-phonétiques pertinents pour percevoir un contraste étranger. Dans cette étude, ils s'intéressent à la perception du contraste anglais /r/-/l/ par des apprenants japonais. Les stimuli sont composés de 68 paires minimales anglaises variant uniquement sur les sons /r/ et /l/ (e.g., *rock – lock*) enregistrés par six locuteurs natifs différents. L'étude suit un paradigme prétest, entraînement, posttest. Les phases de prétest et posttest sont identiques et proposent la tâche d'identification à deux choix forcés lors de laquelle les apprenants japonais entendent des mots et doivent indiquer si le mot contient un /r/ ou un /l/. Les stimuli de cette tâche sont enregistrés par un locuteur dont les productions ne

sont pas utilisées dans les entraînements. L'entraînement est constitué de 15 sessions sur trois semaines. Lors de chaque session, les apprenants réalisent une tâche d'identification à deux choix forcés avec feedback portant sur les stimuli enregistrés par un seul locuteur. Les résultats montrent que les apprenants ayant suivi l'entraînement parviennent à mieux identifier les sons /r/ et /l/ après plutôt qu'avant l'entraînement, alors que ce n'est pas le cas de ceux n'ayant pas suivi l'entraînement. Les résultats montrent ainsi un transfert de l'entraînement à un nouveau locuteur. Cependant, l'importance de l'amélioration varie en fonction du contexte phonétique : les clusters initiaux et les environnements intervocaliques bénéficient d'une plus grande amélioration que les phonèmes seuls en position initiale ou finale, ce qui s'explique partiellement par un effet plafond en prétest pour les phonèmes en position initiale.

Lively, Logan, & Pisoni (1993) ont répliqué cette étude en ajoutant une tâche de généralisation à de nouveaux mots non entraînés prononcés à la fois par des locuteurs entendus lors de l'entraînement et non entendus. Ils montrent que les apprenants sont capables de généraliser les connaissances à des nouveaux mots produits par d'anciens et de nouveaux locuteurs. Dans une seconde expérience, les apprenants ont été entraînés avec des stimuli produits par un seul locuteur, mais les stimuli étaient constitués de cinq environnements phonétiques au lieu de trois dans la première expérience. Dans ce cas, les apprenants ont amélioré leurs performances entre les phases prétest et posttest, mais n'ont pas été capables de généraliser aux mots produits par un nouveau locuteur. Les chercheurs ont donc conclu que la variabilité des stimuli liée à la variation interlocuteurs aidait à construire de nouvelles catégories robustes.

Une autre question au centre des études menées autour du *HVPT* concerne la variabilité qu'il est nécessaire d'introduire dans ces entraînements. Lively et al. (1993) ont testé deux sources de variations : celle liée au locuteur et celle liée au contexte phonétique. Dans une première étude, des participants japonais étaient entraînés avec une tâche d'identification dont les stimuli avaient été enregistrés par plusieurs locuteurs et étaient constitués de mots anglais contenant le contraste /r/-/l/ dans trois contextes phonétiques (position initiale isolée, cluster initial ou position intervocalique). Ils n'ont pas montré d'effet de la taille de l'environnement phonétique en comparant leurs résultats à leurs premiers résultats dans Logan et al. (1991) qui reposaient sur cinq contextes phonétiques. Dans une seconde étude, les apprenants étaient entraînés à l'aide de stimuli enregistrés par un seul locuteur dans lequel le contraste /r/-/l/ était présenté dans cinq contextes phonétiques différents (les trois mêmes que dans l'expérience 1, plus un contexte de position finale et un contexte de cluster final). Ils ont ici montré que les participants

entraînés avec moins de variabilité en termes de locuteurs n'étaient pas capables de généraliser leur apprentissage à de nouveaux mots. Les chercheurs ont conclu que les entraînements ne proposant pas de variabilité au niveau du locuteur ne permettaient pas aux apprenants de développer des catégories de L2 qui soient suffisamment robustes pour traiter correctement de vrais mots.

Lively, Pisoni, Yamada, Tohkura et Yamada (1994) ont repris ce protocole pour déterminer quels en étaient les effets à long terme. Ils ont montré que les apprenants améliorent non seulement leurs performances significativement après entraînement, mais aussi qu'ils sont meilleurs pour les nouveaux mots produits par des locuteurs familiers par rapport aux nouveaux locuteurs non familiers. Les participants sont revenus 3 et 6 mois après l'entraînement pour réaliser un nouveau posttest (sur les stimuli entraînés) et une nouvelle tâche de généralisation (sur de nouveaux items). Les auteurs rapportent que 3 mois après la fin de l'entraînement, les performances sont toujours meilleures qu'avant entraînement à la fois dans la tâche de posttest et dans la tâche de généralisation. Six mois après l'entraînement, les performances sont toujours meilleures qu'avant l'entraînement dans la tâche de posttest, et les performances en généralisation n'ont pas baissé depuis le test réalisé juste après l'entraînement. Les auteurs interprètent cette absence de différence entre posttest et posttest à 6 mois comme un argument pour la capacité du *HVPT* à produire des changements à long terme dans les représentations des catégories phonétiques.

Depuis ces premières versions du *HVPT* de nombreuses études ont utilisé cette technique qui est aujourd'hui considérée comme l'une des méthodes d'entraînement à la perception de parole en laboratoire les plus efficaces.

L'une des questions qui a été soulevée est de savoir si l'apprentissage perceptif issu de cette méthode est transférable à la production. Bradlow, Pisoni, Akahane-Yamada et Tohkura (1997) ont ainsi mené une recherche portant sur le contraste /r/-/l/ anglais pour des apprenants japonais, avec un programme de 45 sessions de *HVPT* sur 3 à 4 semaines. Les apprenants étaient testés en prétest et posttest dans une tâche d'identification à choix forcés et une tâche de répétition de mot pour évaluer leurs performances en production. La qualité de la prononciation des apprenants était évaluée par des juges anglais natifs, de deux manières. Tout d'abord, les jugent voyaient un mot à l'écrit, puis entendaient les productions en prétest et posttest d'un apprenant. Leur tâche était d'indiquer laquelle des deux productions était la meilleure. Ensuite, ils réalisaient une tâche d'identification /r/-/l/ avec l'ensemble des

productions réalisées avant et après l'entraînement par les apprenants. Leurs résultats montrent que les locuteurs anglais natifs préfèrent non seulement les prononciations que les apprenants ont réalisé après l'entraînement, mais également qu'ils les identifient plus justement. Bradlow, Akahane-Yamada, Pisoni et Tohkura (1999) ont montré dans une étude similaire que les effets de l'entraînement en perception et production étaient maintenus 3 mois après l'entraînement. Les auteurs ont interprété ce résultat comme le signe d'un lien perception-production tel que l'amélioration des capacités en perception induit directement une amélioration en production de la parole. En d'autres termes, le *HVPT* se montre efficace pour aider les apprenants à prononcer des sons difficiles en L2.

2.1.2 Les entraînements perceptifs pour l'acquisition des voyelles

Lambacher, Martens, Kakehi, Marasinghe, et Molholt (2005) ont appliqué le *HVPT* aux voyelles de l'anglais difficiles à produire pour des apprenants de langue maternelle japonaise. L'entraînement était constitué de six séances de 20 minutes de *HVPT* étalées sur six semaines. Les participants réalisaient une tâche de lecture de mots CVC (Consonne-Voyelle-Consonne) à haute voix en prétest et posttest. Ces productions étaient évaluées par des juges de langue maternelle anglaise dans une tâche d'identification à cinq choix forcés. Les résultats montrent une amélioration des productions après entraînement à la fois à travers les réponses des locuteurs natifs ayant effectué l'analyse perceptive des productions, et à travers les résultats d'analyses réalisées de manière objective sur le signal acoustique. Les auteurs concluent de cet ensemble de données que la prononciation de phonèmes de L2 peut être améliorée par un entraînement de type *HVPT* alors même qu'aucune information explicite sur une règle de production n'a été donnée.

Cette recherche est complétée par l'étude conduite par Thomson et Derwing (2016) pour tester si un entraînement orientant l'attention des apprenants sur les traits phonétiques pertinents (présentation des voyelles dans des syllabes ouvertes) est plus efficace qu'un entraînement basé sur de vrais mots. Des participants de différentes L1 étaient ainsi entraînés sur des voyelles anglaises dans ce type de contexte. Les résultats montrent que l'entraînement en contexte syllabique et orienté vers le niveau phonétique est plus efficace que celui portant sur les mots pour améliorer la production de ces mêmes mots. De plus, ils ont également montré que la précision des productions avant et après entraînement était affectée par l'interaction entre la fréquence lexicale des mots contenant les voyelles-cibles et le type de voyelles dans

l'environnement phonétique. Les résultats mitigés de cette étude ont conduit les auteurs à conclure qu'un entraînement basé uniquement sur la perception ne suffit pas à améliorer de manière satisfaisante les performances en production. D'autres études ont montré des effets positifs de ce type d'entraînement sur la production de voyelles de L2 (Rato & Rauber, 2015 ; Thomson, 2011), mais l'effet est souvent plus restreint qu'en perception (Lacabex & Lecumberri, 2010 ; Lengeris & Hazan, 2010 ; Nobre-Oliveira, 2007), et il est même parfois inexistant (Wang & Munro, 2004).

L'influence de divers paramètres relevant de la nature des stimuli ou du type de protocoles a été évaluée pour comprendre ce qui rend l'entraînement plus ou moins efficace. Nishi et Kewley-Port (2007) ont par exemple étudié l'effet du nombre de phonèmes sur lesquels les apprenants sont entraînés. Ils ont testé si des apprenants étaient capables d'apprendre plus de cinq catégories de voyelles à la fois, et si un entraînement sur un nombre restreint de catégories de voyelles pouvait aider à l'acquisition d'autres catégories. Leurs participants de langue maternelle japonaise ont été entraînés soit sur le jeu complet des monophtongues anglaises (/i, ɪ, ɛ, æ, ɑ, ʌ, ɔ, ʊ, u/), soit sur un sous-ensemble composé des trois voyelles les plus difficiles pour les apprenants (/ɑ, ʌ, ʊ/). Ils ont montré que les participants entraînés sur le sous-ensemble montraient une plus grande amélioration des performances sur ces voyelles comparativement à l'autre groupe entraîné sur le jeu complet, mais ne montraient pas cependant pas d'effet de transfert puisqu'il n'y avait pas d'amélioration sur les voyelles non entraînées. De plus, les participants entraînés sur le jeu complet étaient capables de généraliser leur apprentissage à de nouveaux mots et à de nouveaux locuteurs, et ce même trois mois après l'entraînement.

Plusieurs études ont proposé des entraînements composés de stimuli synthétiques exacerbant la saillance des traits acoustiques pertinents pour la perception de contrastes vocaliques en L2 (e.g., Nobre-Oliveira, 2007 ; Wang & Munro, 2004). Wang et Munro (2004) ont ainsi testé l'effet d'un entraînement perceptif pour l'apprentissage de trois contrastes vocaliques anglais (/i/-/ɪ/, /ɛ/-/æ/ et /u/-/ʊ/) par des locuteurs mandarins et cantonais vivant au Canada. Seize participants ont été entraînés durant deux mois au moyen d'une tâche d'identification avec *feedback* immédiat. Les participants géraient eux-mêmes l'agenda de leurs sessions. Les sessions d'entraînement commençaient par les items synthétiques, suivis par des sessions de *fading* et les dernières sessions incluaient des items naturels produits par plusieurs locuteurs de langue maternelle anglaise. Le *fading* est un entraînement progressif qui consiste à diffuser des stimuli présentant de grandes différences acoustiques en début d'entraînement et de réduire

ces différences acoustiques au fur et à mesure que l'entraînement avance, afin d'inciter les apprenants à percevoir des différences de plus en plus fines. Des tâches d'identification avec des continua *heed-hid*, *who'd-hood* et *had-head* de synthèse et avec des paires minimales contenant les trois paires de voyelles ont été utilisées pour tester les performances en perception. L'effet de l'entraînement sur la production a été évalué au moyen d'une tâche d'intelligibilité dans laquelle les productions des participants, obtenues dans une tâche de lecture de mots contenant les voyelles en question, étaient identifiées par des locuteurs anglais natifs et par les mesures acoustiques de la durée des voyelles. Les résultats montrent que les participants améliorent leurs performances en perception après entraînement et que l'apprentissage est généralisé à de nouveaux locuteurs. De plus, l'apprentissage est maintenu trois mois après l'entraînement. Cependant, sur la base des résultats issus des deux méthodes d'évaluation des productions, les auteurs ont conclu que l'entraînement basé uniquement sur la perception n'est pas suffisant pour un transfert de compétences en production.

Un effet plus convainquant sur la production est toutefois possible. Nobre-Oliveira (2007) a comparé un entraînement constitué de stimuli naturels à un autre entraînement composé de stimuli de synthèse pour l'apprentissage des contrastes anglais /iː - ɪ/, /ɛ - æ/ et /uː - ʊ/ par des locuteurs portugais brésiliens. L'effet des entraînements était évalué en perception et en production au moyen d'une tâche d'identification portant sur des stimuli naturels et d'une tâche de lecture de mots à haute voix. L'entraînement était constitué de tâches d'identification et de discrimination réparties sur trois semaines. Les résultats ont permet à l'auteur de montrer une amélioration des performances en perception pour les trois contrastes et une amélioration en production, certes restreinte au contraste /iː - ɪ/, mais maintenue même un mois après la fin de l'entraînement. Bien que les apprenants entraînés avec des stimuli naturels présentent une amélioration plus nette des performances, les différences entre les deux groupes ne sont pas significatives.

Le succès de l'entraînement pourrait aussi dépendre de caractéristiques des langues L1 et L2. Iverson et Evans ont par exemple étudié l'effet de la taille de l'inventaire phonologique de L1 sur l'acquisition de contrastes de L2. Iverson et Evans (2007) ont comparé les performances de locuteurs espagnols (avec un inventaire de cinq voyelles) à celles de locuteurs allemands (avec un inventaire de 18 voyelles) pour l'acquisition de voyelles anglaises (/aɪ/, /eɪ/, /ɛ/, /i/, /ɪ/, /ɑ/, /ɜ/, /aʊ/, /əʊ/, /ɒ/, /ɔ/, /u/, /æ/ et /ʌ/) dans un protocole *HVPT*. Vingt-six participants (13 locuteurs espagnols et 13 allemands) ont suivi cinq sessions de 45 minutes comprenant des

tâches d'identification de 225 items avec *feedback* immédiat. Les stimuli des entraînements ont été enregistrés par cinq locuteurs originaires du Sud-Est de l'Angleterre (*Standard Southern British English*) tous différents de ceux ayant enregistré les items des tests. Les performances des participants étaient évaluées avant et après entraînement dans une tâche d'identification à 14 choix forcés composée de 56 mots /bVt/. Les résultats montrent que les apprenants allemands améliorent plus leurs performances que les apprenants espagnols. Ces résultats suggèrent qu'un inventaire phonologique dense facilite l'acquisition de nouveaux contrastes car il rend vraisemblablement les apprenants de L2 plus sensibles aux différences catégorielles fines entre les voyelles. Dans une autre étude, Iverson et Evans (2009) ont montré que les apprenants espagnols étaient capables d'améliorer leurs performances autant que les allemands s'ils faisaient dix sessions d'entraînement supplémentaires. Plus tard, Iverson, Pinet et Evans (2012) ont testé l'impact d'un entraînement en *HVPT* pour l'acquisition de voyelles de l'anglais par des apprenants francophones de niveaux d'expertise variés en anglais. L'entraînement était constitué sur le même principe que dans l'étude de Iverson et Evans (2009) et les stimuli des tests étaient enregistrés par dix locuteurs britanniques. Les performances en perception ont été évaluées par des tâches d'identification à 14 choix forcés (14 AFC) et de discrimination (*oddity*) et les productions ont été enregistrées dans une tâche de lecture de mot puis évaluées par quatre locuteurs natifs (tâche d'identification à 14 choix forcés). Cette expérience a montré que les deux groupes de participants ont bénéficié l'un et l'autre tout autant de l'entraînement. Cette étude est la seule qui, à notre connaissance, propose des entraînements pour les voyelles orales de l'anglais à des apprenants francophones.

L'efficacité d'un entraînement pourrait aussi être modulée par la nature des tâches proposées dans les exercices d'apprentissage. Plusieurs études proposant des entraînements perceptifs sur les contrastes vocaliques de L2 ont fait varier ces tâches pour tester leurs effets. Par exemple, Cebrian (2015) a comparé l'efficacité d'entraînements basés sur des exercices d'identification ou de discrimination. Elle a montré qu'un entraînement s'appuyant sur une tâche d'identification produisait une amélioration plus nette pour les items ayant été traités pendant l'entraînement, alors que les exercices de discrimination permettaient une meilleure généralisation de l'apprentissage à des items non entraînés. Dans une étude toute récente, Wee, Grenon, Sheppard et Archibald (2019) ont comparé l'utilisation d'une tâche d'identification et de discrimination AX pour entraîner de façons différentes deux groupes de locuteurs nipponophones au contraste vocalique de l'anglais /iː-ɪ/. Les locuteurs japonais se caractérisent par une tendance à utiliser la durée de façon excessive pour discriminer les voyelles. Une tâche

de pondération des traits évaluait l'utilisation des informations spectrales et temporelles par les apprenants des deux groupes avant et après un entraînement de seulement une heure. Leurs performances ont également été comparées à celles de locuteurs anglais natifs. Les auteurs ont montré que le groupe ayant réalisé les exercices de discrimination et celui exercé en identification réduisent leur utilisation de la durée et augmentent leur sensibilité au timbre de manière équivalente ; ils se rapprochent ainsi des mécanismes d'analyse utilisés par des locuteurs natifs. Les participants des deux groupes montrent des performances similaires aux natifs en ce qui concerne l'utilisation de la durée, mais restent tout de même différents des natifs pour le timbre des voyelles, bien que l'utilisation de cet indice devienne plus automatique. Dans une autre étude de Grenon et collaborateurs (2019) ne rapportant que les effets de l'entraînement avec les exercices d'identification, les auteurs ont rapporté qu'après l'entraînement, onze apprenants sur 23 ont créé une nouvelle catégorie sur la dimension spectrale. Ils interprètent ce résultats comme une preuve de la capacité des apprenants de L2 tardifs à créer de nouvelles catégories phonétiques suite à ce type d'entraînement, mais que les nouvelles catégories peuvent faire l'objet d'une dissimilation phonétique catégorielle comme décrite dans le modèle SLM (Flege, 1995) c'est-à-dire que la frontière de la nouvelle catégorie est placée aussi loin que possible de la catégorie de L1 la plus proche, afin de maintenir distincts les deux systèmes phonétiques.

2.2 D'autres types d'entraînements pour l'acquisition de contrastes vocaliques de L2

Bien que le *HVPT* ait montré son efficacité pour l'apprentissage de consonnes mais aussi de voyelles de L2, de nouveaux protocoles n'impliquant pas uniquement la perception auditive ont également été développés.

Certains de ces entraînements se sont distingués des *HVPT* classiques quant à la nature des exercices proposés aux apprenants. Aliaga-García et Mora (2009) ont proposé un programme composé de multiples tâches à des apprenants bilingues catalans/espagnols pour les voyelles anglaises /iː - ɪ/, /æ - ʌ - ɑː/ et /uː - ʊ/. Parmi divers exercices en perception et en production, l'entraînement comptait une tâche de production avec feedback visuel de 15 minutes réalisée avec le logiciel EyeSpeak (Chan, 2004). Ce logiciel permet au participant de voir la réalisation de sa voyelle dans l'espace des deux premiers formants, d'en vérifier la durée, d'obtenir un score de justesse et d'ajuster sa production. Les résultats de cet entraînement sur les voyelles

intégrant des indices visuels montrent que le groupe entraîné améliore ses performances en perception pour les trois contrastes après entraînement, mais qu'il présente toujours des scores plus faibles que ceux de locuteurs natifs. Pour évaluer l'effet de l'entraînement sur la qualité des productions orales, les valeurs des deux premiers formants et la durée des voyelles produites avant et après l'entraînement ont été mesurées pour tester l'effet de l'entraînement. Les analyses montrent que l'entraînement n'a pas d'effet sur les productions des apprenants et que leurs productions diffèrent toujours beaucoup de celles de locuteurs natifs.

D'autres études ont proposé des entraînements audio-visuels pour tester si des apprenants pouvaient bénéficier d'indices visuels consistant cette fois en des traits faciaux pour mieux apprendre de nouveaux contrastes de L2. Inceoglu (2016) a ainsi comparé l'effet d'entraînements auditifs et audio-visuels pour l'apprentissage des voyelles nasales du français [ɔ̃–ɑ̃–ɛ̃] par des apprenants anglophones. Les deux entraînements de six sessions chacun étaient composés d'exercices d'identification avec *feedback* portant sur des mots et non-mots de structure CVC contenant l'une des trois voyelles nasales. Dans l'entraînement audio-visuel, les apprenants voyaient des vidéos d'une femme prononçant les stimuli alors qu'ils entendaient seulement les stimuli dans l'autre entraînement. Les participants étaient évalués en perception et en production avant et après ce programme. Les performances en production étaient mesurées au moyen d'une tâche de répétition de mots isolés entendus avant une phrase en français pour éviter une imitation directe à partir de l'information stockée en mémoire sensorielle. Les enregistrements en prétest et posttest étaient ensuite appréciés par des locuteurs natifs. La perception était évaluée par une tâche d'identification à trois choix forcés (*3AFC*) dans laquelle les stimuli étaient présentés dans trois modalités différences : audio-visuelle, seulement auditive et seulement visuelle. Les résultats de cette étude montrent que les deux groupes ont, l'un comme l'autre, amélioré leurs performances en perception, d'une façon comparable. Cependant, pour les performances en production, il s'avère que les apprenants ayant suivi l'entraînement audio-visuel bénéficient d'une amélioration plus nette que celle des apprenants ayant suivi l'entraînement seulement auditif. L'auteure conclut que le fait de voir les traits faciaux lors de l'apprentissage améliore la prononciation.

Certains entraînements se singularisent par l'association d'exercices de perception et de production. Pour dissocier les effets des deux types d'exercice, Alshangiti et Evans (2014) ont comparé l'influence d'un entraînement en production pour l'apprentissage des voyelles de l'anglais par des apprenants arabophones, à l'influence d'un entraînement *HVPT* classique et

à celle d'un entraînement hybride combinant production et *HVPT*. Dans les exercices de production, pour chaque voyelle, le participant entendait un mot-clé (e.g., *heed*) puis entendait la voyelle isolée (e.g. /iː/). Ensuite, il voyait une vidéo montrant l'articulation de cette voyelle sur une coupe sagittale. C'est-à-dire qu'il voyait la position de la mâchoire, de la langue et des lèvres. Il lui était ensuite demandé de produire la voyelle en isolation, puis de prononcer le mot clé et enfin un mot exemple (e.g. *heat*, *feet*). Enfin, le participant s'enregistrait en produisant ces trois items contenant la voyelle et comparait auditivement ses propres productions à celles de locuteurs natifs avec l'aide d'un instructeur. Les résultats de cette étude montrent que les bénéfices de l'entraînement dépendent du type d'exercices effectués. En effet, les apprenants entraînés en perception ne manifestent une amélioration de leurs performances qu'en perception, alors que ceux entraînés en production s'améliorent essentiellement en production.

L'implication de la production de parole dans les entraînements est parfois plus sophistiquée, et accordent au *feedback* une place importante. Certains programmes ont la particularité de tenter de moduler la production durant les exercices en proposant des *feedbacks* basés sur les propriétés acoustiques des productions des apprenants. Dès 1998, Dowd, Smith et Wolfe ont ainsi utilisé un spectromètre à impédance acoustique pour fournir une représentation du conduit vocal. Proposé à des locuteurs anglais apprenants le français, l'entraînement fournissait en temps réel une représentation visuelle des deux premières résonances des voyelles produites, afin qu'elles puissent être comparées à celles de locuteurs natifs du français. Plus récemment, Kartushina, Hervais-Adelman, Frauenfelder et Golestani (2015) ont aussi proposé un entraînement basé sur l'analyse en temps réel des paramètres acoustiques des voyelles produites par les apprenants afin de leur proposer un *feedback* immédiat, item par item, avec une comparaison possible entre la réalisation de leur voyelle et celle de locuteurs natifs dans le plan des deux premiers formants. Plus précisément, des différences pouvaient être visualisées entre les deux premiers formants et la F0, comme présenté dans la Figure 9(b). Les participants étaient des francophones n'ayant aucune connaissance des langues slaves (danois, suédois, finnois). Il leur était proposé un entraînement pour des voyelles danoises présentant des différences phonétiques avec les voyelles françaises. Les exercices amenaient donc à produire les contrastes de timbre en danois /e-ɛ/ et /y-ø/.

Les participants étaient évalués en perception au moyen d'une tâche de discrimination et en production avant et après l'entraînement. La précision des voyelles produites a été évaluée en calculant des distances de Mahalanobis entre les productions des apprenants et les voyelles cibles dans l'espace F1/F2. Les résultats montrent qu'un entraînement en production de seulement une heure par voyelle permet d'améliorer non seulement la production des voyelles de L2 mais aussi leur perception. Cette fois, le sens classique du lien imaginé entre perception et production est renversé. Les auteurs relèvent également des variations importantes entre les individus quant à l'intensité de l'amélioration.

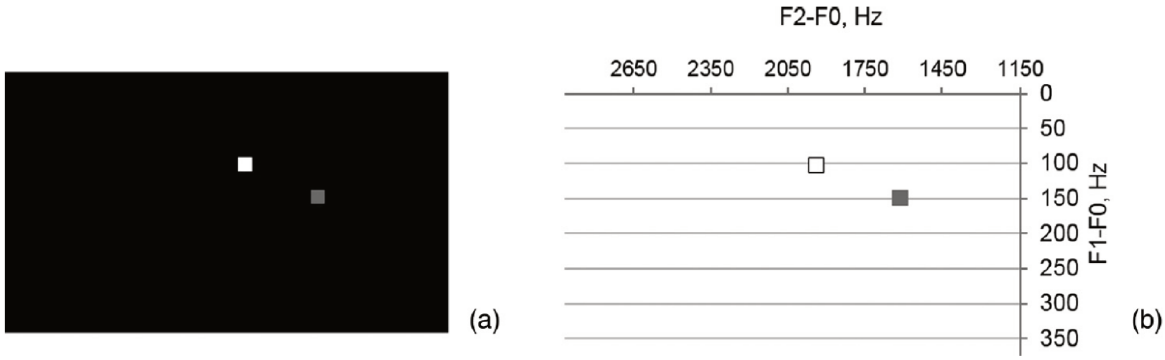

Figure 9 : Exemple de feedback visuel proposé aux apprenants dans l'étude de Kartushina et collègues (2015) à gauche (a). Le carré blanc correspond à la voyelle produite par l'apprenant et le gris à la cible native à atteindre. La représentation de droite (b) montre les unités et l'étendue utilisée qui correspond à l'espace des 4 voyelles de leur étude. Image reproduite de Kartushina et al. (2015).

Un autre type d'image a encore été utilisé pour le feedback visuel. Pillot-Loiseau, Antolík et Kamiyama (2013) ont en effet proposé un entraînement où le feedback consistait à percevoir visuellement une image de la langue par ultrason. Dans cette étude, deux locuteurs de langue maternelle japonaise ont été exercés à produire les voyelles françaises /y/ et /u/. Lors des sessions d'entraînements, ils produisaient ce contraste et observait le contour de leur langue sur des images à ultrasons : pour produire le /u/ la position de langue doit être la plus postérieure possible par rapport à celle du /y/. Leurs performances en production de voyelles isolées montrent qu'après l'entraînement les distances euclidiennes entre /y/ et /u/ dans le plan F1-F2 sont plus élevées et les contours de langues sont plus espacés, traduisant l'efficacité du programme pour permettre aux apprenants de réaliser le contraste plus distinctement.

2.3 Entraînements perceptifs et potentiels évoqués

Les études en EEG ayant montré les effets d'entraînements perceptifs pour l'acquisition de la phonologie en langue seconde sont encore assez peu nombreuses et, conformément à l'histoire

des entraînements perceptifs, celles-ci ont d'abord été menées sur des contrastes consonantiques.

Kraus et al. (1995) ont exercé des participants à discriminer deux variantes de synthèse du phonème /d/ dans /da/. Les résultats après un entraînement d'une semaine à raison d'une heure par jour sont visibles sur la MMN. En effet, la durée et l'aire de cette onde augmentent en phase posttest. Tremblay, Kraus, Carrell et McGee (1997) ont mis en évidence des corrélats EEG signalant le transfert d'un apprentissage à des contextes non différents de celui des exercices. Ils ont proposé un entraînement à des locuteurs de langue maternelle anglaise pour identifier une occlusive labiale prévoisée (/b/) non utilisée en anglais. Les participants ont été évalués avant et après l'entraînement dans des tâches de perception du contraste ayant fait l'objet de l'entraînement, mais aussi le transfert sur d'autres consonnes occlusives alvéolaires prévoisées (/d/). Ils ont montré une amélioration des performances après l'entraînement, y compris pour le contraste de voisement pour cette nouvelle consonne. L'apprentissage a donc été transféré à un autre lieu d'articulation. Des données en EEG témoignent aussi de cet effet, car la durée et l'aire de la MMN augmentent quand cette onde est générée par les sons entraînés, et la latence d'apparition de cette MMN diminue pour les stimuli non traités pendant l'entraînement. Les auteurs ont également relevé que les changements des caractéristiques de la MMN sont plus importants dans l'hémisphère gauche. Ces résultats peuvent être rapprochés de ceux de Tremblay, Kraus, McGee, Ponton et Otis (2001) qui ont mis en évidence des changements de la morphologie du complexe N1-P2 consécutifs à un entraînement en perception du VOT suivant un peu le même principe.

Suite à l'observation de modifications des réactions cérébrales dans l'hémisphère gauche, c'est le cortex auditif gauche qui fut l'objet d'une étude en magnétoencéphalographie (MEG) conduite par Menning (2002). Il s'agissait de relever les indices d'un apprentissage des contrastes temporels des mores du japonais dans cette aire cérébrale connue pour sa dominance pour le traitement du langage. L'évaluation portait sur le MNF (*Mismatch Negativity Field*), équivalent de la MMN en EEG. En japonais, la more est une unité phonétique qui règle la longueur temporelle d'un son ou d'une syllabe. Ainsi, la structure des mores peut modifier le sens des mots. Cette expérience comparait les changements de réponses cérébrales de participants allemands et japonais à des différences dans la structure des mores de mots japonais. Les participants allemands ont suivi un entraînement en discrimination de dix sessions de 1.5 heure. Ils ont appris à discriminer des mots japonais présentés par paire où le

second mot était raccourci d'une more. À l'issue des exercices, une amélioration significative des performances a été observée sur des mesures comportementales et elle s'accompagnait d'une augmentation significative de l'amplitude de la MNF. Les taux de détection correcte de déviations de durée ont considérablement augmenté après l'entraînement. Les temps de réactions et la latence de la MNF ont diminué significativement entre les sessions d'exercices. La MNF des participants japonais a montré une sensibilité à des différences particulièrement fines. Cette étude a contribué à montrer que l'apprentissage perceptif d'un contraste temporel de mores non-natives est possible et qu'il est corrélé à changements mesurables dans les substrats corticaux.

Les données électrophysiologiques sur les effets d'entraînements perceptifs, cette fois sur les voyelles, ont été enrichies par Reinke, He, Wang et Alain (2003) qui ont évalué l'impact d'un tel entraînement basé sur des exercices de discrimination de phonèmes. Dans cette étude, les stimuli étaient les cinq voyelles /i/, /a/, /æ/, /u/ et /ɜ/ de l'anglais américain. Les auteurs ont créé des doubles voyelles en superposant les signaux des voyelles, et chaque voyelle était ainsi associée aux quatre autres. Chaque item comportait donc deux voyelles différentes présentées simultanément de façon binaurale et les participants devaient identifier les deux voyelles de la paire. Aucun *feedback* n'était donné. Deux groupes ont été constitués : l'un recevant un entraînement, l'autre n'en recevant pas. L'entraînement comportait trois séances pendant lesquelles les participants réalisaient la même tâche de discrimination qu'en phases prétest et posttest lors de l'enregistrement en EEG. Il n'est pas précisé si les participants recevaient un *feedback*. Cette fois, les auteurs ont étudié le complexe N1-P2. Leurs résultats montrent que l'amélioration en discrimination des voyelles présentées simultanément s'accompagne d'une diminution de la latence du complexe N1-P2 sur les électrodes fronto-centrales et temporales et d'une augmentation de l'amplitude de la P2 dans le groupe entraîné comparé au groupe non entraîné. L'effet du programme sur les potentiels évoqués renforce l'idée selon laquelle les apprentissages perceptifs sont accompagnés de changements au niveau du cortex sensoriel. Toujours à propos de l'onde P2, Sheehan, McArthur et Bishop (2005) ont testé l'effet des exercices de discrimination de sons de parole du continuum /ga/-/da/ sur l'onde P2. Il s'agit cette fois de consonnes. Alors que seuls les participants ayant suivi l'entraînement montraient des améliorations comportementales en discrimination, les deux groupes (y compris celui n'ayant pas reçu d'entraînement) présentaient une augmentation significative de l'onde P2 en posttest, ce que les auteurs ont imputé à la simple exposition aux stimuli durant l'enregistrement EEG. Cette interprétation est renforcée par l'observation d'une augmentation

significative de l'onde entre le début et la fin de la session. Les auteurs concluent ici qu'une augmentation de l'onde P2 reflète une amélioration des capacités de discrimination du cortex auditif.

Afin de s'assurer que des entraînements en simple perception peuvent avoir une influence sur les corrélats électrophysiologiques des apprenants, Lu, Wayland et Kaan (2015) ont proposé à des locuteurs anglais natifs des exercices où il s'agissait seulement de percevoir, ou bien de percevoir et de produire des tons lexicaux, qui leur étaient totalement inconnus. Les deux types d'entraînement ont été efficaces sur les performances comportementales et les participants des deux groupes ont même généralisé leur apprentissage à des stimuli non entraînés. Les mesures en potentiels évoqués ont montré que l'amplitude de la MMN diminuait après l'entraînement et, entre eux, les deux groupes ne présentaient pas de différences de compétences en traitement des tons ni au niveau attentionnel, ni au niveau pré-attentionnel. Ces résultats suggèrent que l'utilisation du système moteur (dans l'entraînement combinant perception et production) n'apporte pas de bénéfices particuliers pour la perception des tons.

Les indices électrophysiologiques de changements de procédure d'analyse des sons d'une L2 ont été aussi étudiés de manière plus qualitative. Parmi les rares études en potentiels évoqués menées sur les voyelles, Ylinen et collaborateurs (2010) ont testé si une utilisation inadaptée des indices acoustiques due à une assimilation inappropriée entre L1 et L2 pouvait être modifiée par un entraînement strictement phonétique. Dans cette étude, ils se sont intéressés à l'utilisation des indices spectraux pour la perception des voyelles anglaises /ɪ/ et /iː/ par des locuteurs finlandais. L'hypothèse était que, puisqu'en finnois la durée est un indice phonologique pertinent pour discriminer les phonèmes longs et courts, les participants finlandais utilisent sans doute davantage la durée que les locuteurs anglophones pour distinguer ces deux voyelles en anglais L2. L'entraînement proposait des exercices d'identification reposant sur des paires minimales pour lesquelles, dans la moitié des cas, les durées vocaliques des stimuli avaient été modifiées afin d'inciter les locuteurs finnois à utiliser les indices spectraux des voyelles, comme le font les anglophones, plutôt que leurs paramètres temporels. Ils ont comparé avant entraînement l'utilisation des indices spectraux et de durée des voyelles anglaises /iː/ et /ɪ/ entre des locuteurs anglais et finnois natifs. Les effets inter-langues et ceux des entraînements ont été évalués en comportement et en EEG par l'étude de la MMN. L'analyse des performances a montré qu'avant l'entraînement les

locuteurs finnois natifs s'appuient bien davantage que les locuteurs anglais sur l'indice de durée. Après l'entraînement, ils ont montré que les locuteurs finnois utilisaient mieux les indices spectraux pour reconnaitre les voyelles. La combinaison des données de type comportemental et électrophysiologique permet donc d'étudier de manière particulièrement fine la nature des modifications d'analyse du signal de parole à l'issue d'entraînements en L2 ciblant des aspects très précis de la langue, mais aussi les stratégies les plus appropriées pour les traiter. En ce qui concerne les résultats sur la MMN, les comparaisons inter-langues, présentées dans la Figure 10, montrent que chez les Finlandais, la MMN était plus petite pour le /iː/ familier présenté en contexte de /ɪ/ non familier, que celle pour le /ɪ/ non familier présenté en contexte de /iː/ familier, alors que le *pattern* opposé est observé chez les locuteurs anglais. Ce résultat révèle une asymétrie dans le traitement des voyelles chez les Finlandais qui serait, selon les auteurs, due au fait que la mémoire à long terme facilite la formation d'une représentation régulière élaborée pour la voyelle familière /iː/ mais pas pour la voyelle non familière /ɪ/.

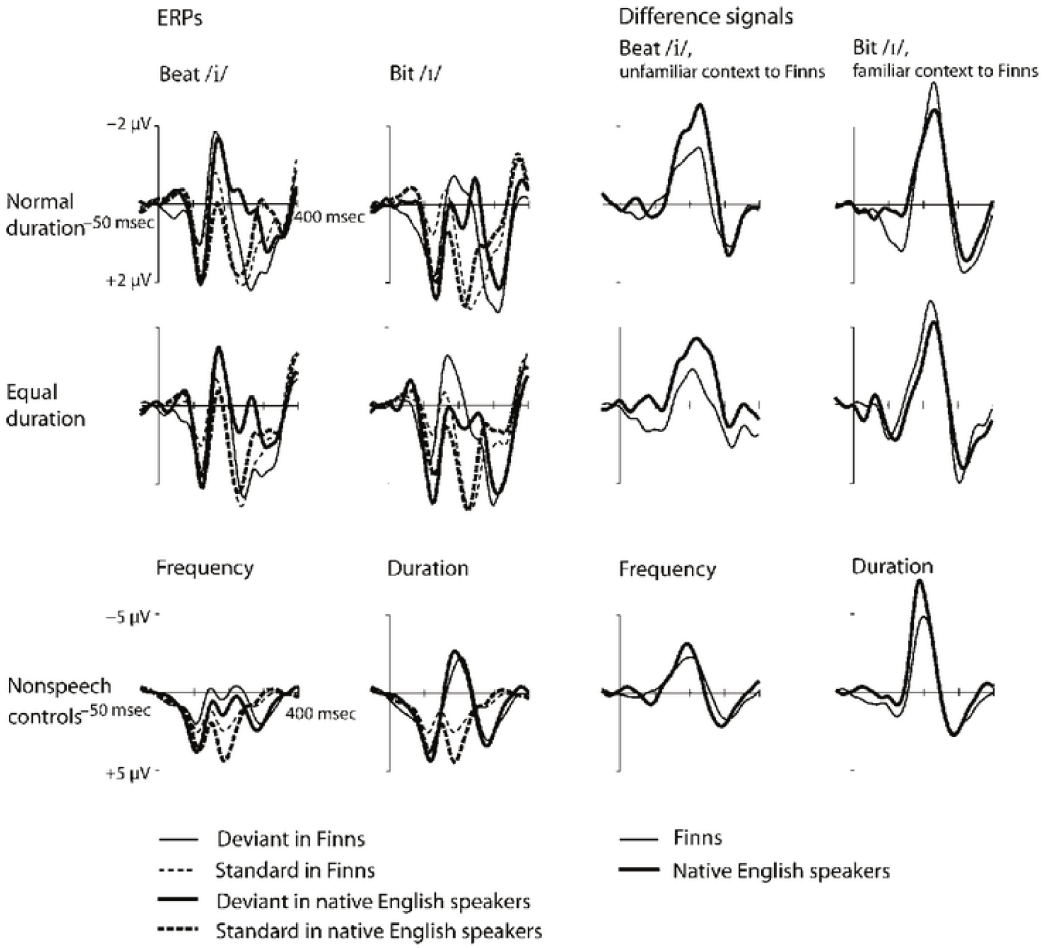

Figure 10 : MMN en Fz chez les locuteurs de langue maternelle anglaise (ligne épaisse) et finnois (ligne fine) avant entraînement. A gauche sont représentées les réponses aux stimuli standards (pointillés) et déviants (ligne continue), à droite la différence (déviant moins standard). Les deux premières lignes montrent les réponses évoquées par les stimuli beat et bit avec des durées normales (donc différentes) ou égales. La ligne du bas montre les réponses pour les changements de fréquence et durée dans la condition contrôle avec les stimuli non langagiers. (Image reproduite de Ylinen et al. 2010)

Par ailleurs, pour ce qui est des effets de l'entraînement, les résultats sur la MMN (voir Figure 11) montrent un effet de l'entraînement chez les apprenants : la MMN était significativement plus ample en phase posttest qu'en phase prétest dans l'hémisphère gauche sur les électrodes centrales. De plus, bien que la différence soit non significative, les auteurs soulignent que l'augmentation de la MMN tend à être plus importante pour le /iː/ présenté dans un contexte non familier, que pour le /ɪ/ présenté dans un contexte de /iː/ familier. Ils présument que sous l'effet d'une exposition perceptive répétée et des efforts actifs fournis pour assimiler les stimuli correctement à des mots anglais pendant l'entraînement, les Finlandais sont devenus plus familiers avec les caractéristiques spectrales du /ɪ/ précédemment non prototypique. Cela a pu permettre la formation d'une représentation plus précise pour les stimuli standards /ɪ/, ce qui a induit une augmentation de la MMN générée par le déviant /iː/. Le fait d'avoir dirigé l'attention des participants sur les indices spectraux en réduisant les indices de durée

disponibles pour catégoriser les stimuli peut aussi avoir réduit l'assimilation des voyelles anglaises avec des représentations des voyelles finnoises en mémoire à long-terme. Ainsi, cela a pu permettre le développement de nouvelles représentations pour la voyelle /ɪ/ en mémoire à long-terme, indépendamment du système phonologique finnois.

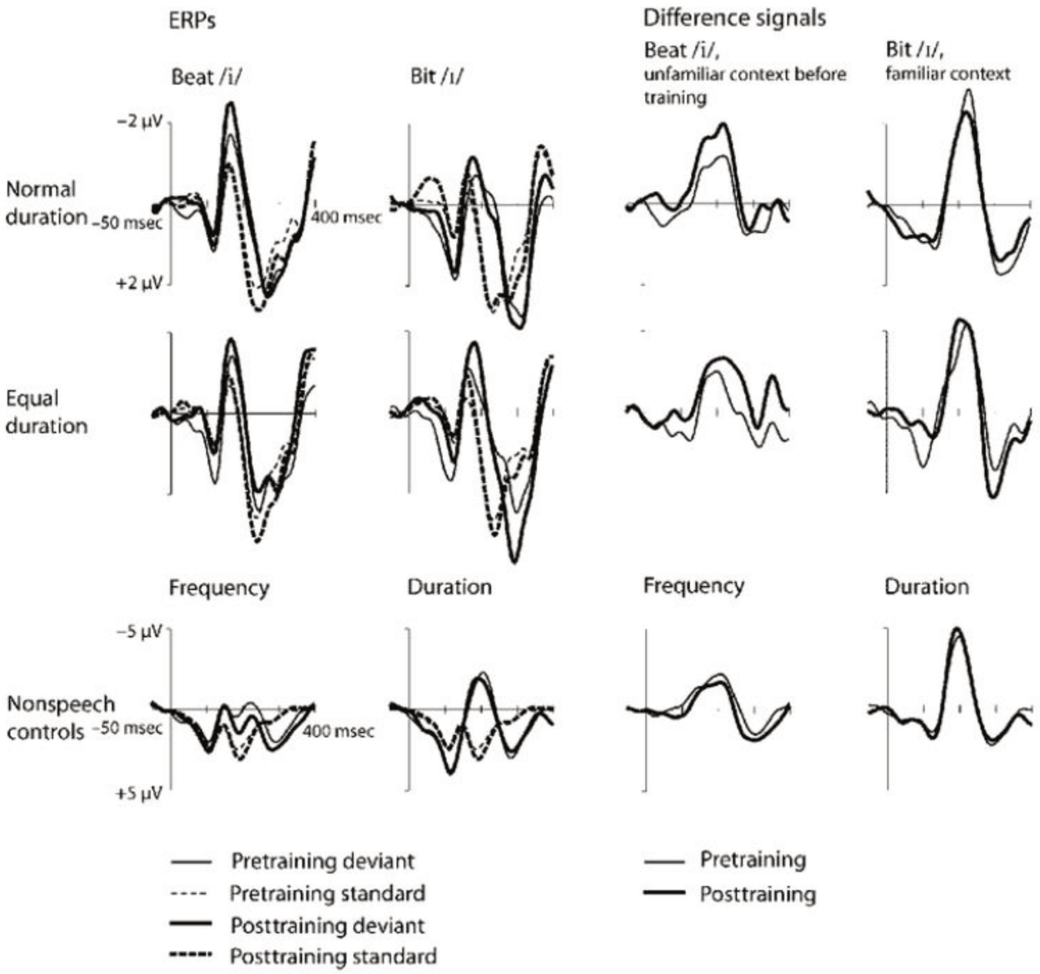

Figure 11 : MMN sur l'électrode Fz en prétest (ligne fine) et posttest (ligne grasse) telle que représentée par les réponses aux stimuli déviants et standard à gauche et la différence déviant – standard à droite. Les deux premières lignes montrent les réponses évoquées par les stimuli beat et bit avec des durées normales ou égales. La ligne du bas montrent les réponses pour les changements de fréquence et durée dans la condition contrôle avec les stimuli non langagiers. (Image reproduite de Ylinen et al. 2010)

Le VOT est lui aussi un indice de durée, et il oppose des consonnes en anglais d'une façon qui peut être nouvelle pour des locuteurs d'une autre langue. Là encore d'après des données en potentiels évoqués, un entraînement approprié semble pouvoir y remédier. Tamminen et al., (2015) ont montré qu'après un entraînement de trois jours consistant à écouter et répéter le contraste de voisement anglais /f/-/v/, des participants finnois devenaient capables de discriminer ce contraste absent en finnois et que cet apprentissage s'accompagnait d'une augmentation significative de la MMN pour ce contraste. Pour revenir aux voyelles, très

récemment, Saloranta et al., (2020) ont étudié l'efficacité d'un entraînement de trois jours pour la perception et la production de contrastes de durée de voyelles. Douze participants de différentes langues maternelles ont ainsi réalisé quatre sessions d'un entraînement où ils devaient écouter et répéter des pseudo-mots choisis pour contenir des voyelles contrastées par leur durée (e.g., /tite/-/ti:te/). Les auteurs ont aussi testé la généralisation de l'apprentissage à un autre contraste /tote/-/to:te/ et à des paires de tons. Ils testaient l'effet de l'entraînement sur des performances en perception, en production, mais également sur des ondes enregistrées en potentiels évoqués (la N1 et la MMN). Ils ont montré une nette amélioration sur toutes les mesures pour le contraste entraîné. L'effet était aussi présent pour le contraste vocalique non entraînés sur la N1 et sur les performances dans une tâche de discrimination de tons. Cependant, pour les stimuli non entrainés, la MMN n'augmentait pas après les exercices. Les auteurs concluent que l'entraînement a permis d'augmenter la sensibilité pré-attentive à la durée, mais que des informations spectrales spécifiques aux phonèmes peuvent également être nécessaires pour façonner la représentation neuronale des catégories de phonèmes.

3 Les systèmes vocaliques du français et de l'anglais britannique

Dans cette thèse, nous nous sommes intéressée aux problèmes rencontrés par les apprenants francophones de France pour la perception et la production des voyelles de l'anglais britannique. Nous avons considéré l'anglais du sud-est de l'Angleterre qui constitue la variété d'anglais enseignée en France (dite *Received Pronounciation*). Dans cette partie, nous présenterons les systèmes vocaliques du français, puis de l'anglais britannique du sud-est de l'Angleterre et nous terminerons en recensant les difficultés rencontrées par les apprenants francophones rapportées dans la littérature en ce qui concerne les voyelles.

3.1 Système vocalique du français

Contrairement aux consonnes, dont le nombre reste stable d'une variété de français à l'autre, le nombre de voyelle peut varier (Valdman, 2000). Le français standard compte entre dix et douze voyelles (i, e, ɛ, a, y, ø, œ, u, o, ɔ, ɑ, ə) orales et trois voyelles nasales /ɛ̃, ɔ̃, ɑ̃ / (Fougeron & Smith, 1993 ; Vaissière, 2009).

Les voyelles orales sont couramment classifiées selon trois paramètres :

- La hauteur, qui correspond au degré d'ouverture de la cavité buccale. On distingue de cette manière :
  - Voyelles fermées
  - Voyelles mi-fermées
  - Voyelles mi-ouvertes
  - Voyelles ouvertes
- La position de la langue dans la cavité buccale sur l'axe antéro-postérieur permettant de distinguer :
  - Voyelles antérieures
  - Voyelles postérieures
- L'arrondissement des lèvres donnant lieu aux :
  - Voyelles arrondies
  - Voyelles non-arrondies

Les voyelles du français ne sont pas phonologiquement discriminables sur la base de leur durée.

L'ensemble des voyelles orales du français est représenté selon ces trois paramètres dans le tableau Tableau I.

Tableau I : Classification des voyelles orales du français (d'après Gess, Lyche et Meisenburg, 2012, p. 4)

	Antérieures		Postérieures	
	Non-arrondies	Arrondies	Non-arrondies	Arrondies
Fermées	i	y		u
Mi-fermées	e	ø		o
Mi-ouvertes	ɛ	œ ə		ɔ
Ouvertes	a		(ɑ)	

Les voyelles orales sont communément représentées en phonétique par le biais d'un trapèze vocalique qui met en relation les paramètres acoustiques (les formants) et articulatoires des voyelles qui sont corrélés (voir Figure 12). Ainsi, dans ce type de représentation, le premier formant (F1) des voyelles est représenté sur l'axe des ordonnées croissant du haut vers le bas, il correspond au degré d'aperture des voyelles de fermé à ouvert. Plus la valeur du premier formant est élevée, plus la voyelle est ouverte. Le second formant (F2) est représenté de gauche à droite sur l'axe des abscisses et correspond au positionnement antéro-postérieur de la langue dans la cavité buccale. Une valeur de F2 élevée correspond à une voyelle antérieure alors qu'une valeur de F2 faible indique une voyelle postérieure. Enfin, l'arrondissement des lèvres

apparait plutôt corrélé au deuxième formant pour les voyelles postérieures et au troisième formant pour les voyelles antérieures, un arrondissement des lèvres faisant baisser respectivement ces deux formants (Vaissière, 2015). Une représentation dans l'espace des formants F2/F3 est donc plus appropriée pour refléter ce paramètre.

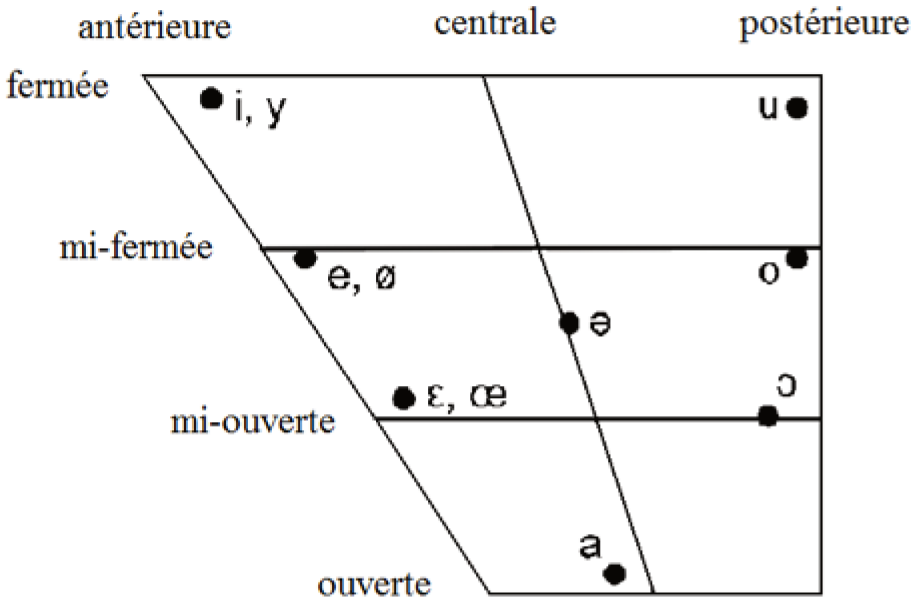

Figure 12 : Trapèze vocalique des voyelles orales du français (reproduit de Capliez, 2011)

3.2 Système vocalique de l'anglais britannique

Le système vocalique de l'anglais standard est composé de douze voyelles orales comme présenté dans le Tableau II. On remarque qu'en plus des positions antéro-postérieures sur l'axe horizontal, il existe une position centrale. Les voyelles frontales du système anglais sont nécessairement non arrondies alors que les voyelles postérieures sont arrondies, sauf la voyelle [ʌ]. Un quatrième paramètre permet de décrire les voyelles de l'anglais, le caractère tendu ou relâché de la voyelle qui est systématiquement lié à la durée vocalique marquée par le symbole ː. En effet, les voyelles tendues sont plus longues que les voyelles relâchées. La Figure 13 présente le trapèze vocalique de l'anglais.

Tableau II : Classification des voyelles orales de l'anglais

	Antérieures		Centrales		Postérieures	
	Tendues	Relâchées	Tendues	Relâchées	Tendues	Relâchées
Fermées	iː				uː	
mi-fermées		ɪ	ɜː	ə	ɔː	ʊ
mi-ouvertes		e		ʌ		
Ouvertes		æ			ɑː	ɒ

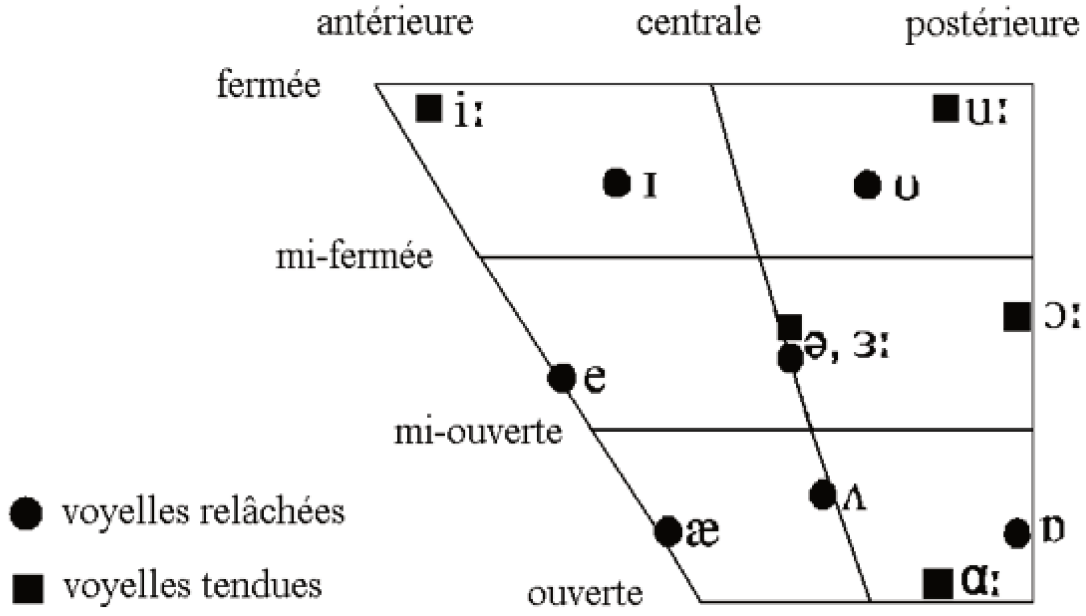

Figure 13 : Trapèze vocalique des monophtongues de l'anglais (reproduit de Capliez, 2011)

3.3 Difficultés des apprenants francophones pour les voyelles de l'anglais

Dans son travail de thèse sur le rôle de traits segmentaux et suprasegmentaux dans l'acquisition et l'enseignement de la phonologie de l'anglais par des locuteurs français, Capliez (2016) a enregistré quatre locuteurs français dans une tâche de lecture de texte afin de déterminer les erreurs qu'ils produisaient notamment au niveau des voyelles anglaises. Il a ainsi noté plusieurs substitutions : l'utilisation du [ɛ] comme substitut du /e/ anglais (e.g., *hello*), [a] à la place de /æ/ (e.g., *have*), [ɔ] au lieu de /ɒ/ (e.g., *hot*), [œ] pour /ʌ/ (e.g., *must*), [i] pour /ɪ/ (e.g., *think*), [u] pour /ʊ/ (e.g., *book*) et même [œ] ou [o] à la place de /ə/ (e.g., asleep). Dans ces cas, les sons de L1 et de L2 partagent certaines caractéristiques bien que des symboles phonétiques

différents soient utilisés. Capliez relève également d'autres types d'erreurs de prononciation dans lesquels la voyelle substituée ne présente pas de ressemblance spécifique avec la voyelle cible mais l'utilisation de cette voyelle peut être attribuée à une influence de l'orthographe. Par exemple, [y] est parfois utilisé à la place de /ʌ/ (e.g., *bus*) car c'est la prononciation classique de la lettre <u> en français. D'autres erreurs de prononciation proviennent de l'utilisation d'emprunts anglais en français, ainsi, il a pu observer le mot anglais *club* prononcé /klœb/, *chewing-gum* prononcé /ʃwiŋgɔm/ ou encore *people* prononcé /pipœl/ ou /pipɔl/. Capliez remarque que, pour certains phonèmes, d'autres substitutions auraient pourtant conduit à une meilleure précision phonétique. Par exemple, la voyelle française [œ] se substitue à la voyelle anglaise /ʌ/, alors que le [a] français est plus proche de cette voyelle cible, notamment dans l'espace des deux premiers formants.

Alors même que le système français ne présente pas de diphtongues, ce ne sont pas les diphtongues mais les monophtongues anglaises qui posent le plus de problèmes aux apprenants francophones (Capliez, 2011). En effet, la distinction entre voyelles tendues et relâchées n'existant pas en français, les francophones ont tendance à n'utiliser qu'un seul exemplaire phonémique. Ainsi, les paires /æ/-/ɑː/, /ɪ/-/iː/, ou encore /ʊ/-/uː/ sont neutralisées et deviennent respectivement /a/, /i/ et /u/. La neutralisation de la distinction entre les voyelles tendues et relâchées est une autre conséquence du transfert du français à l'anglais du fait de l'absence de cette distinction en français. L'exemple le plus typique est l'utilisation systématique du son français [i] comme substitut des deux voyelles /ɪ/ relâchée et /iː/ tendue sur lesquelles repose par exemple la distinction de la paire minimale *live* et *leave*.

En perception, Iverson et Evans, (2007) relèvent que plusieurs voyelles de l'anglais sont assimilées à la même voyelle française. En particulier, le /ɪ/ et le /iː/ qui sont assimilées au /i/ français, les trois voyelles /a/, /aɪ/ et /ʊ/ assimilées au /a/ français, les voyelles /ɜ/ et /ʌ/ assimilées au /ø/ français ou encore les voyelles /ɔ/ et /əʊ/ assimilées au /o/ français.

4 Objectifs de l'étude

À notre connaissance, dans la littérature, seule l'étude de Iverson et collaborateurs (2012) a mis en place des entraînements dont l'objectif était de permettre à des apprenants francophones d'améliorer leurs performances en perception et en production des voyelles orales de l'anglais.

En effet, dans cette étude, les auteurs ont entraîné les participants sur l'ensemble des 14 voyelles du système vocalique britannique. En perception, l'entraînement s'est montré efficace dans la tâche d'identification pour la plupart des voyelles, sauf pour les voyelles /ɪ/, /ʌ/ et /aɪ/, mais a eu très peu d'effet sur la discrimination. De plus, ils n'ont observé une amélioration significative des performances en production (telles que mesurées par une évaluation des productions par des locuteurs natifs) que pour /i/, /eɪ/ et /aʊ/. Les auteurs ont aussi constaté que les participants qui amélioraient le plus leurs performances après l'entraînement, commettaient, en prétest, des erreurs de type orthographique (e.g., dans une tâche de lecture de mots, ils prononçaient *bout* présenté à l'écrit comme si c'était un homophone de *boat* ou *boot*). Ainsi l'amélioration observée après entraînement témoignait plutôt du fait que les participants parvenaient à associer la bonne voyelle lorsqu'ils la voyaient écrite. Le bilan général de cette étude quant à l'efficacité de la méthode utilisée (*HVPT*) étant donc mitigé, notamment en production, nous avons voulu utiliser un autre type d'entraînement qui serait susceptible de présenter une meilleure efficacité sur les performances en production, sans amenuiser le bénéfice apporté sur le plan des performances en perception.

Pour cela, nous avons mis au point un entraînement qui, en plus de la perception, implique également la production orale des voyelles ciblées par l'entraînement. Nous supposions que le fait le réaliser la production des voyelles pendant la phase d'entraînement permettrait d'améliorer la qualité ultérieure de leur production de manière plus efficace qu'après un entraînement ne sollicitant que des processus perceptifs comme le font les programmes de type *HVPT*. Peu d'études ont directement comparé l'effet d'entraînements en production à celui d'entraînements en perception, à la fois sur la perception et la production des voyelles. Si la perception des sons de parole est dirigée par la production, comme le prédit la Théorie Motrice de la perception de la parole (Liberman et Mattingly, 1985), exercer les participants à produire ces sons et non simplement à les percevoir devrait renforcer l'efficacité de la perception de la parole, au moins pour les phonèmes ciblés.

De plus, nous nous attendons à ce que le fait de produire certaines voyelles durant l'entraînement favorise une meilleure production de ces dernières, en amenant à la création de nouvelles cibles articulatoires plus précises, telles que définies dans le modèle DIVA de Guenther (2011). Ce modèle suggère que le *feedback* auditif lors de tâche de répétition peut être utilisé pour réarranger les cartes neuronales afin de modifier des *patterns* de production déjà établis. Le *feedback* auditif est en effet responsable du maintien de l'intelligibilité de la

parole et il ferait pour cela en sorte que les cibles articulatoires soient correctement atteintes. A cette fin, il effectuerait des comparaisons entre les sons produits par le locuteur et la cible prédite par le système d'anticipation. En cas de *mismatch*, le système est supposé réajuster les commandes motrices et les prédictions afin d'atteindre la cible auditive. Liu et Tian (2018) ont montré que la suppression de ce *feedback* auditif dans une tâche de répétition de mots empêchait l'apprentissage de nouvelles voyelles de L2. Ils ont également montré que le *feedback* n'était réellement efficace que si la cible en L2 était suffisamment éloignée de sons existant dans la L1. Ce résultat nous incite à penser que, dans le cas où les voyelles entraînées sont suffisamment discriminables par les apprenants des voyelles de leur L1, des exercices de production permettraient de bénéficier de ce *feedback* auditif pour affiner les cibles articulatoires des voyelles produites. Cette attente est aussi fondée sur les résultats de plusieurs études ayant proposant des entraînements de type écoute-répétition et conduisant à une amélioration des performances en perception comme en production de contrastes en langue étrangère (Saloranta et al., 2020 ; Tamminen et al., 2015). Une autre particularité de l'entraînement testé dans notre recherche est de proposer un *feedback* visuel basé sur l'analyse des paramètres acoustiques des voyelles produites par les apprenants, afin qu'ils puissent ajuster leurs productions pendant les exercices pour se rapprocher de celles de locuteurs natifs. Ce type d'entraînement s'est montré efficace pour l'acquisition de contrastes vocaliques danois absents en français par des apprenants français (Kartushina et al., 2015).

Pour évaluer l'efficacité des entraînements, nous avons utilisé un protocole prétest-entraînement-posttest auprès d'un groupe d'apprenants adultes, mais nous avons en plus comparé leurs performances à celles d'un groupe d'apprenants comparable, testés eux aussi deux fois avec un intervalle équivalent à celui séparant les étapes prétest et posttest, mais ne suivant pas d'entraînement entre ces deux séries de tests. De plus, les performances du groupe entraînés ont été comparées à celles d'un groupe de locuteurs de langue maternelle anglaises pour nous assurer que les éventuelles modifications de performances des participants entraînés tendaient bien à se rapprocher de celles de natifs après l'entraînement.

Nous nous sommes focalisée sur seulement deux régions de l'espace acoustique pour lesquelles le système phonologique anglais comporte plusieurs voyelles, là où le système français n'en présente qu'une. La comparaison des systèmes vocaliques Figure 14 montre que, dans la région du [a], une seule catégorie existe en français alors que l'anglais en présente trois − /æ/, /ʌ/ et /ɑː/ − qui se situent entre le /a/ français et le /œ/. Dans la région du [i], une seule catégorie est

présente en français, mais deux catégories existent en anglais : le /iː/ et le /ɪ/. Outre le timbre des voyelles, la durée a une valeur phonologique en anglais, contrairement au français.

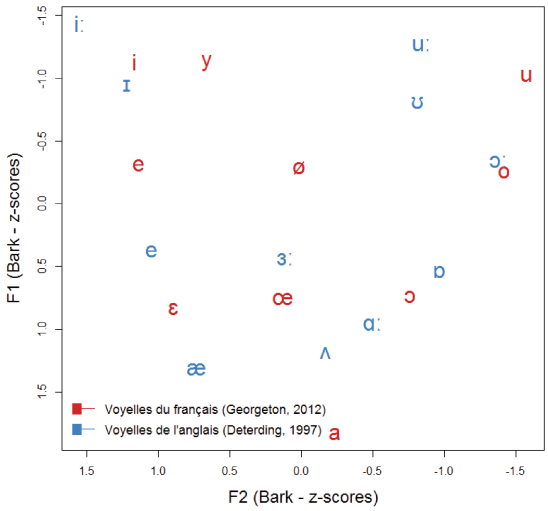

Figure 14 : Voyelles du français (d'après Georgeton et al., 2012, en rouge) et monophtongues de l'anglais (d'après Deterding, 1997, en bleu), pour les femmes.

Une vue générale du protocole expérimental suivi dans les groupes est présentée dans le Tableau III.

Tableau III : Déroulement du protocole expérimental suivi dans les groupes.

		GROUPE			
	TÂCHES	Perception (PE)	Production (PR)	Contrôle (C)	Natif (N)
PRETESTS	- 7AFC identification	✓	✓	✓	✓
	- Oddity discrimination	✓	✓	✓	✓
	- Production français & anglais	✓	✓	✓	✓
	- EEG (oddball)	✓	✓	✓	✓
TRAINING	**Sessions 1 à 3 :** 1 séance par contraste (/æ - ʌ/ et /ʌ - ɑː / et /ɪ - iː/)				
	- 2AFC identification	✓			
	- AX discrimination	✓			
	- Répétition de mots		✓		
	Sessions 5 & 6 : 2 séances avec tous les contrastes				
	- 7AFC identification	✓			
	- Oddity discrimination	✓			
	- Répétition de mots		✓		
POSTTESTS	- 7AFC identification	✓	✓	✓	
	- Oddity discrimination	✓	✓	✓	
	- Production anglais	✓	✓	✓	
	- EEG (*oddball*)	✓	✓	✓	
Nombre de participants		**16**	**16**	**16**	**16**

Ainsi, nous avons suivi l'évolution des performances en prétest et posttest, en perception et en production, pour les cinq voyelles de l'étude. Pour cela, nous avons utilisé des tâches de perception classiques (identification et discrimination) et une tâche de répétition de mots pour évaluer les performances en production. En perception, nous souhaitions répliquer les effets déjà observés dans d'autres études quant à l'efficacité de l'entraînement de type *HVPT* et nous pensions que l'entraînement basé sur la production produirait quant à lui un effet au moins

comparable à l'entraînement *HVPT*, avec l'hypothèse qu'il serait même supérieur. En production, nous avons évalué la parole enregistrée des apprenants de deux manières : en analysant les paramètres acoustiques des voyelles et en faisant juger leur qualité par des locuteurs natifs. En ce qui concerne les paramètres acoustiques des voyelles, nous nous attendions à ce que l'entraînement intégrant des exercices de production produise de plus grands changements que ceux induits par l'entraînement *HVPT*, notamment dans le sens d'un rapprochement vers les paramètres observés dans le groupe de locuteurs natifs. Quant à l'évaluation par des juges natifs, nous pensions que, après l'entraînement basé sur la production, elles seraient meilleures qu'avant l'entraînement et meilleures que celles des productions consécutives à l'entraînement *HVPT*.

En plus de cette évaluation comportementale, nous avons testé si des indices d'acquisition des phonèmes pouvaient faire suite aux entraînements et être observables au niveau cérébral avec la méthode des potentiels évoqués. Comme il a été montré que des traces mnésiques induites par un apprentissage pouvaient être observées à travers les caractéristiques de la MMN et que les paramètres de cette onde peuvent changer pendant l'acquisition d'une L2, que ce soit suite à l'arrivée dans un nouvel environnement linguistique (e.g., Winkler et al., 1999) ou suite à un entraînement (Saloranta et al., 2020; Tamminen et al., 2015; Ylinen, Shestakova, Huotilainen, Alku et Näätänen, 2006), nous avons souhaité observer si des variations de la MMN se produisaient après notre entraînement. Ces modifications pourraient être les indices d'une perception plus automatique des indices pertinents des voyelles entraînées, après la série d'exercices.

Nous étudierons l'effet des entraînements sur le contraste /æ-ʌ/ qui a la particularité de ne pas présenter de frontière phonologique en français. Nous le comparerons à un autre contraste de l'anglais pour lequel une frontière phonologique existe en français, le contraste /æ-ɛ/. Une MMN de belle amplitude devrait être observée chez tous nos participants français pour ce second contraste. En revanche, nous nous attendons à ce qu'en prétest, la MMN pour le contraste /æ-ʌ/ soit de moins grande amplitude et/ou plus tardive. Nous espérons que, pour les groupes ayant suivi un entraînement, cette onde présente des variations entre les étapes prétest et posttest et que ces dernières seront corrélées aux performances observées en perception. Nous testerons également un groupe de locuteurs natifs en EEG pour voir si les locuteurs ayant suivi un entraînement présentent des *patterns* de réponses évoqués se rapprochant des leurs, soit en termes temporel et d'amplitude (Jakoby et al., 2011), soit en terme de latéralisation de

la réponse comme l'ont observé Zevin et al., (2010) décrite au Paragraphe 1.2.2. En effet, dans cette étude, les auteurs n'avaient pas mis en évidence de différence entre les locuteurs natifs et les locuteurs de L2 avancés sur la MMN analysée seulement sur les électrodes fronto-centrales, mais ils avaient montré des différentes en termes de topographie. Bien qu'il convienne d'être prudent sur l'interprétation des données spatiales en EEG, il s'agit d'une piste intéressante pour affiner l'analyse des données.

Partie 2. Contribution expérimentale

Dans ce chapitre, nous présenterons l'étude qui a été menée. Dans une première partie, nous présenterons le matériel expérimental qui a été construit pour les besoins des tests et des entraînements. Une deuxième partie décrira la méthodologie expérimentale des deux entraînements créés, puis nous présenterons l'évaluation des performances des apprenants en perception, puis en production et enfin en EEG.

1 Matériel expérimental

Le matériel expérimental était constitué de paires minimales anglaises de type consonne-voyelle-consonne (CVC) portant sur deux régions de l'espace vocalique de l'anglais, la région /ɪ - iː/ et la région /æ - ʌ - ɑː/ pour laquelle les deux contrastes /æ - ʌ/ et /ʌ - ɑː / ont été considérés. Quatre paires minimales (bVd, bVt, hVd et hVt, i.e., quatre contextes CVC) ont servi dans les tâches comportementales des phases de tests et dix autres, non utilisées dans les phases tests, ont servi à l'élaboration des entraînements (voir Tableau IV). L'ensemble des stimuli utilisés étaient des mots, aucun pseudo-mot n'a été utilisé.

Tableau IV : Paires minimales utilisées dans les phases test et les phases d'entraînement

	VOYELLE				
PHASE	/iː/	/ɪ/	/æ/	/ʌ/	/ɑː/
TEST (prétest et posttest)	bead	bid	bad	bud	bard
	beat	bit	bat	but	bart
	heed	hid	had	hudd	hard
	heat	hit	hat	hut	heart
ENTRAÎNEMENT	beak	bik	back	buck	Bark
	bean	bin	ban	bun	barn
	cheap	chip	cad	cud	card
	deed	did	cap	cup	carp
	keep	kip	cat	cut	cart
	leak	lick	lack	luck	lark
	mean	min	mat	mutt	mart
	peak	pick	pack	puck	park
	seek	sick	pat	putt	part
	teak	tick	tat	tut	tart

Ces paires minimales ont été enregistrées par dix-huit locuteurs (neuf hommes et neuf femmes) de langue maternelle anglaise. Dans la mesure du possible, les locuteurs ont été sélectionnés pour présenter un accent du Sud-Est de l'Angleterre (assez proche de la *Received Pronunciation – RP*). Quatorze locuteurs étaient étudiants en échange à Lyon pour un semestre ou une année, deux étaient enseignants, un autre était pilote et un dernier pilote retraité. L'un de nos locuteurs était originaire de région parisienne (Saint Cloud) mais de parents anglais et avait l'anglais comme langue maternelle. Nous avons conservé ses enregistrements après écoute de ceux-ci par E. Ferragne qui a confirmé qu'il présentait bien un accent proche de RP. Les origines géographiques des locuteurs sont présentées dans le Tableau V, leurs âges dans le Tableau VI. Les locuteurs en gris sont ceux dont les enregistrements ont été utilisés dans les tests, les autres ceux dont les enregistrements ont été utilisés dans les entraînements. Chaque

locuteur a rempli un formulaire de consentement et a été défrayé 10€ pour une heure d'enregistrement environ[1].

Tableau V : Origine géographique des locuteurs et locutrices ayant enregistrés les stimuli expérimentaux (en gris pour les stimuli utilisés dans les tests, en blanc pour ceux des entraînements)

Locuteurs	Origine géographique	Locutrices	Origine géographique
H01	London	F01	SouthHampton
H02	Rochester, UK	F02	London
H03	King's Lynn	F03	Chelmsford
H04	Sidcup	F04	London
H05	London	F05	Guildford
H06	High Wycombe	F06	Essex
H07	Saint Cloud	F07	London
H08	Stockport	F08	London
H09	Colchester	F09	Croydon

Tableau VI : Âge des locuteurs anglais ayant enregistré le matériel expérimental

Moyenne	27
Écart-type	12,83
Médiane	20
Min	20
Max	67

Les enregistrements ont été réalisés dans une salle isolée acoustiquement de l'Institut des Sciences de l'Homme avec un micro USB Audio-Technica 2020. Le signal était converti au format numérique PCM mono avec un taux d'échantillonnage de 44,1 kHz et une résolution de 16 bits. Le logiciel ROCme! a été utilisé pour réaliser les enregistrements. Celui-ci permet de présenter visuellement les stimuli à produire et d'enregistrer les productions mot par mot dans des fichiers .wav séparés (Figure 15).

[1] L'ensemble des défraiements des locuteurs et participants ayant contribué dans cette thèse a été financé par le LabEx ASLAN (ANR-10-LABEX-0081) de l'Université de Lyon dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) du gouvernement français géré par l'Agence Nationale de la Recherche française.

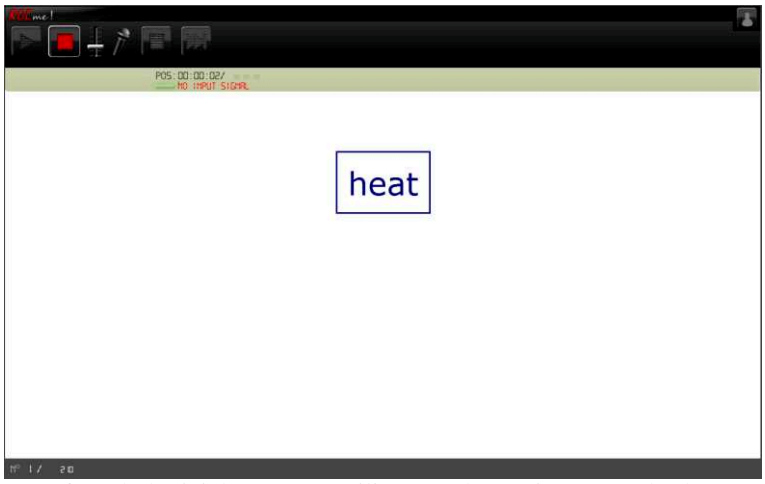

Figure 15 - Interface du logiciel ROCme! utilisé pour l'enregistrement des locuteurs natifs

L'ensemble des mots des phases tests et des entraînements étaient présentés trois fois chacun dans un ordre aléatoire. Chaque locuteur a ainsi enregistré 210 mots[2].

Pour la constitution des stimuli expérimentaux, une seule occurrence de chaque mot a ensuite été sélectionnée pour chaque locuteur selon deux critères : i) la qualité de l'enregistrement et ii) la qualité de la réalisation de la voyelle, qui était déterminée à partir de l'écoute des fichiers avec l'aide d'Emmanuel Ferragne et de l'inspection des spectrogrammes. Au total, le corpus était constitué de 1260 mots, soit 252 occurrences (14 contextes CVC × 18 locuteurs) pour chaque voyelle.

Le logiciel Praat (Boersma et Weenink, 2018) a été utilisé pour l'élaboration des stimuli expérimentaux et leurs analyses acoustiques. Les fichiers ont été segmentés manuellement pour ne contenir que le mot produit. Tous les stimuli ont été normalisés en amplitude de pic au seuil de 0.99. Les trois premiers formants ainsi que la durée de la voyelle ont été mesurés par estimation semi-automatique : l'estimation formantique de Praat superposée au spectrogramme était ajustée en faisant varier le nombre de formants estimés et la fréquence maximale de détection des formants (en Hz) jusqu'à adéquation pour les trois premiers formants sur la durée de la voyelle (voir Figure 16). Pour chaque formant, une valeur toutes les 6.25 millisecondes a été extraite puis afin de minimiser les erreurs provenant de l'extraction

[2] Chaque locuteur a en réalité enregistré au total 294 mots, car le projet initial prévoyait de tester également le contraste /ʊ-uː/, mais les enregistrements pour ce contraste n'étaient pas de bonne qualité et plusieurs locuteurs ne produisaient pas le contraste. Ce contraste a donc été retiré de l'étude.

formantique de Praat (saut de formant) la valeur médiane sur la durée de la voyelle a été calculée. Nous avons considéré que cette unique valeur était suffisante pour caractériser les valeurs formantiques des voyelles car mis à part le /iː/, qui peut présenter une faible diphtongaison chez certains locuteurs de cette variété d'anglais, l'ensemble des voyelles de notre étude présentent des trajectoires formantiques stables (Ferragne, 2008).

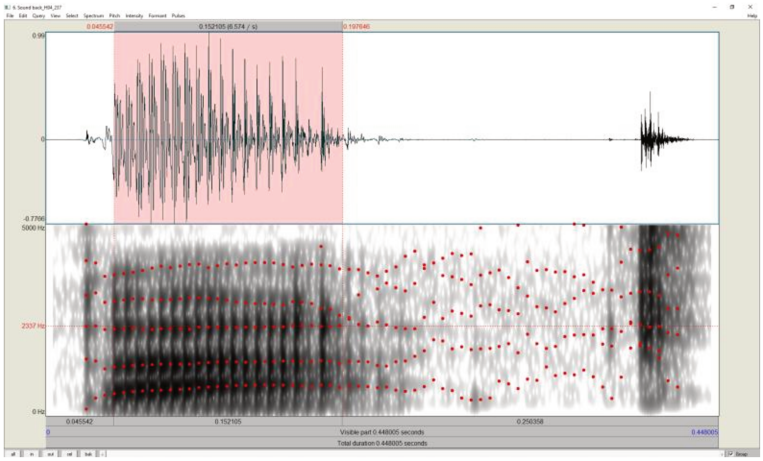

Figure 16 : Estimation formantique réalisée sous Praat

Le matériel des phases tests, dont les paramètres acoustiques sont présentés dans le Tableau VII et la Figure 17, a été sélectionné parmi les productions de six des locuteurs (trois hommes et trois femmes). Au total, le matériel des phases tests était constitué de 120 mots CVC, i.e., 5 voyelles × 4 contextes CVC × 6 locuteurs (3 hommes et 3 femmes).

Tableau VII : Valeurs moyennes des trois premiers formants (en Hz) et de la durée des voyelles (en ms) des stimuli utilisés dans les phases tests. Les écart-types sont indiqués encore parenthèses.

Voyelle	Sexe	N	F1	F2	F3	Durée
/æ/	F	12	999 (99)	1585 (138)	2810 (110)	170 (74)
	H	12	774 (78)	1472 (88)	2508 (87)	169 (62)
/ʌ/	F	12	753 (73)	1418 (164)	3030 (181)	124 (29)
	H	12	675 (64)	1224 (59)	2519 (118)	124 (21)
/ɑː/	F	12	767 (79)	1255 (140)	3043 (158)	262 (64)
	H	12	671 (64)	1092 (57)	2531 (113)	243 (83)
/ɪ/	F	12	550 (61)	2327 (165)	3099 (111)	129 (27)
	H	12	425 (30)	2003 (141)	2591 (208)	118 (24)
/iː/	F	12	381 (35)	2877 (121)	3468 (221)	196 (72)
	H	12	274 (13)	2283 (60)	3059 (91)	207 (79)

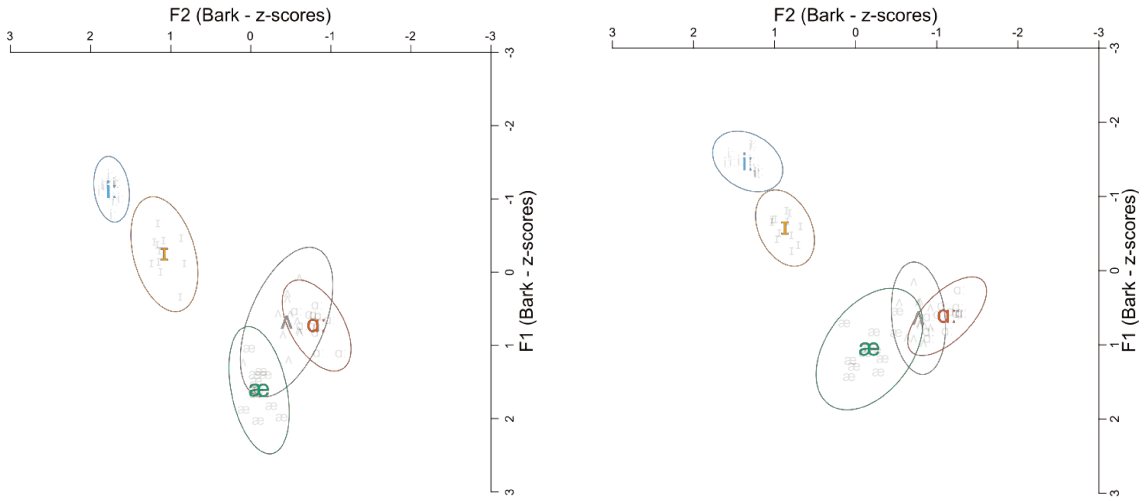

Figure 17 : Représentation dans l'espace F1/F2 (en Bark[3] centrés-réduits) des voyelles des stimuli utilisés dans les tests, pour les femmes à gauche et les hommes à droite. Les labels indiquent les centres de catégorie, les ellipses représentent un intervalle de confiance de la moyenne de la catégorie à 95%.

Les productions de douze autres locuteurs (six hommes et six femmes) ont été sélectionnées pour constituer le matériel des entraînements. Les paramètres acoustiques de ces stimuli sont présentés dans le Tableau VIII et la Figure 18. Au total, le matériel des entraînements était constitué de 600 mots CVC, i.e., 5 voyelles × 10 contextes CVC × 12 locuteurs (6 hommes et 6 femmes).

Tableau VIII : Valeurs moyennes des trois premiers formants (en Hertz) et de la durée des voyelles (en ms) des stimuli utilisés dans les entraînements

Voyelle	Sexe	N	F1	F2	F3	Durée
/æ/	F	60	988 (154)	1695 (109)	2947 (198)	157 (62)
	H	60	734 (103)	1536 (87)	2552 (207)	145 (48)
/ʌ/	F	60	818 (111)	1428 (106)	2989 (280)	118 (31)
	H	60	674 (52)	1192 (119)	2667 (158)	122 (31)
/ɑː/	F	60	755 (67)	1239 (69)	3081 (244)	273 (49)
	H	60	624 (55)	1093 (100)	2726 (153)	251 (64)
/ɪ/	F	60	555 (60)	2356 (178)	3115 (232)	123 (34)
	H	60	424 (39)	1961 (115)	2697 (169)	116 (33)
/iː/	F	60	371 (37)	2830 (129)	3486 (266)	186 (71)
	H	60	314 (44)	2243 (166)	2993 (297)	178 (66)

[3] Nous avons réalisé la conversion en Bark suivant l'équation : $z = [26.81/(1 + 1960/f)] - 0.53$ (où z la fréquence sur l'échelle des Barks et f la fréquence à transformer en Hertz). (Traunmüller, 1990)

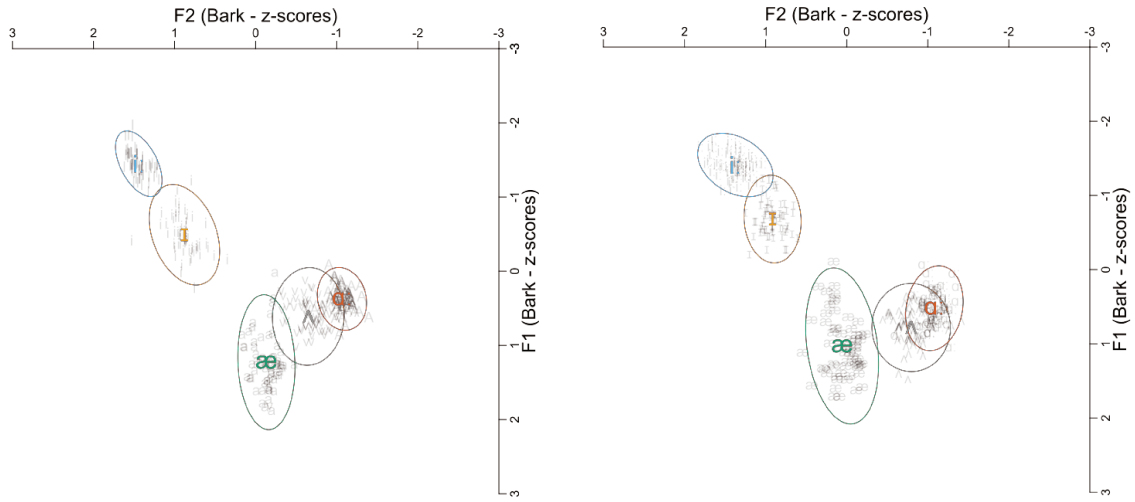

Figure 18 : Représentation dans l'espace F1/F2 (en Bark centrés-réduits) des voyelles des stimuli utilisés dans les entraînements, pour les femmes à gauche et les hommes à droite. Les labels indiquent les centres de catégorie, les ellipses représentent les intervalles de confiance à 95% du barycentre des catégories.

Des analyses discriminantes linéaires (*LDA*) ont été réalisées avec la méthode du *leave-one-out* pour tester le caractère discriminable des catégories. Ces analyses portaient sur les valeurs médianes de F1, F2, F3 (en Hz) et la durée des voyelles de l'ensemble des stimuli utilisés dans les tests et les entraînements, indépendamment pour les hommes et pour les femmes. Les deux sets comportaient chacun 630 voyelles. Les modèles discriminants ont été validés avec des taux de classification correcte de 95% pour les femmes et 93% pour les hommes. En ce qui concerne les durées des voyelles (représentées dans la Figure 19), un modèle d'analyse linéaire mixte avec comme facteurs fixes le Set (Test ou Entraînement) et la voyelle, et comme facteur aléatoire le locuteur n'a pas montré d'effet principal du Set, ni d'interaction Set ×Voyelle. L'effet de la Voyelle est significatif, $F(4, 694) = 109.57$, $p < .001$, indiquant suite à des analyses post-hoc que les voyelles courtes sont significativement plus courtes que les voyelles longues dans les deux sets.

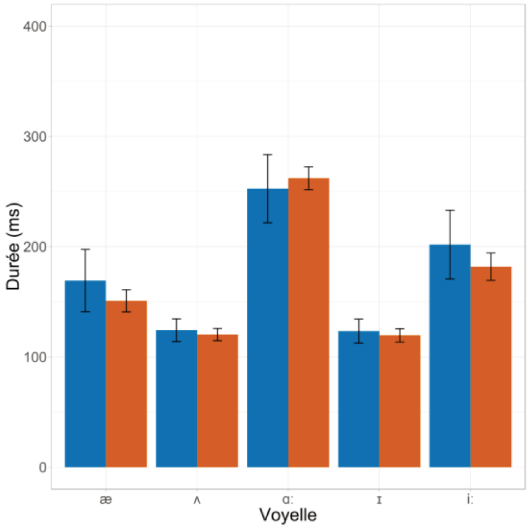

Figure 19 : Durées moyennes des voyelles utilisées dans les tests (en bleu) et les entraînements (en orange).

2 Méthodologie expérimentale des entraînements

2.1 Participants

Quarante-huit participants de langue maternelle française non bilingues, inscrits en première année d'anglais (LEA ou LLCER) à l'Université Lyon 2 ou Lyon 3 ont participé à cette étude. Les participants étaient tous droitiers, sans troubles de l'audition ni du langage et ne devaient pas avoir séjourné plus d'un mois en pays anglophone. Les âges des participants sont présentés dans le Tableau IX. Ces participants ont été dispersés dans trois groupes différents au fur et à mesure des candidatures à l'expérience de façon à incrémenter le nombre de participants dans les groupes de manière simultanée. Cette filière universitaire rassemble beaucoup plus de femmes que d'hommes, il n'a donc pas été possible d'équilibrer les groupes en genre. Cependant, ces trois groupes comportaient une même proportion entre femmes et hommes, soit un quart d'hommes. Deux groupes ont reçu un entraînement, soit en perception (groupe PE, n = 16), soit en production (groupe PR, n = 16), et un troisième groupe constituait le groupe contrôle (groupe C, n = 16) et ne recevait pas d'entraînement mais écoutait des livres audio en anglais pendant une durée équivalente aux entraînements, soit 1 heure chaque jour pendant 5 jours. Nous avons utilisé cette méthodologie sur les conseils de Nathalie Bedoin, chercheur au laboratoire, pour ne pas que ce groupe ne fasse rien. Dans la littérature sur les entraînements à des contrastes vocaliques similaires à notre étude, soit les études n'ont pas de groupe contrôle (e.g., Iverson et al., 2012), soit le groupe contrôle ne fait rien entre prétest et posttest (e.g., Aliaga-García & Mora, 2009) soit un entraînement sur d'autres contrastes était proposé aux

participants (e.g., sur des contrastes consonantiques dans Rato & Rauber, 2015), soit le groupe contrôle réalise le même entraînement que le groupe expérimental mais ne recevait pas de feedback (e.g., Kartushina et al., 2015). Dans la mesure où, pour des contraintes de temps, nous n'avons pas pu construire un matériel élaboré à proposer à notre groupe contrôle, il nous a semblé pertinent de suivre le conseil de proposer des livres audio à notre groupe contrôle. Seize participants de langue maternelle anglaise originaires du Sud-Est de l'Angleterre ont également participé à l'étude ; ils constituaient le groupe natif (Groupe N). Ce groupe était constitué de 9 femmes et 5 hommes. La distribution des âges des participants selon leur groupe est décrite dans le Tableau IX.

Tableau IX : Âges des participants selon les groupes

Groupe	PE	PR	C	N
Moyenne	18.88	18.63	19.31	24.56
Ecart-type	1.62	1.36	1.99	4.90
Min	17	17	17	20
Max	24	23	24	38
Médiane	18	18	19	22

Tous les participants signaient un formulaire de consentement au début de l'étude et étaient défrayés 100€ pour les groupes PE et PR dont la participation exigeait de venir sept fois au laboratoire pour une durée total de 9h et 40€ pour les participants du groupe C dont la participation consistait à ne venir au laboratoire que deux fois pour un total de 4h. Les participants du groupe N ne venaient qu'une seule fois au laboratoire pour une session de 2h étaient défrayés 20€. Le protocole a été soumis à un comité de protection des personnes et été considéré hors-champ Loi Jardé (décret 2017-884 du 9 mai 2017).

2.2 Déroulement des entraînements

Les séances d'entraînement, au nombre de cinq, se sont déroulées dans la salle d'expérience de l'Institut des Sciences de l'Homme à Lyon. Les participants venaient jusqu'à 4 simultanément pour le groupe entraîné en perception (PE), et individuellement pour le groupe qui suivait l'entraînement en production (PR). Les séances duraient une heure maximum.

Les entraînements selon les groupes suivaient la procédure présentée dans le Tableau X et les tâches elles-mêmes sont décrites dans les sections suivantes.

Tableau X : Déroulement des entraînements selon les groupes

		GROUPE			
	TÂCHES	Perception	Production	Contrôle	Natif
TRAINING (~1h par session)	**Sessions 1 à 3 :** 1 séance par contraste (/ɪ - iː/, /æ - ʌ/ et /ʌ - ɑː /)				
	2AFC identification	✓			
	AX discrimination	✓			
	Répétition de mots		✓		
	Écoute d'*ebook* en anglais			✓	
	Sessions 5 & 6 : 2 séances avec tous les contrastes				
	5AFC identification	✓			
	Oddity discrimination	✓			
	Répétition de mots		✓		
	Écoute d'*ebook* en anglais			✓	
Nombre de participants		**16**	**16**	**16**	**16**

Pour les deux types d'entraînements (PE et PR), chacune des trois premières séances était focalisée sur un seul contraste (/ɪ - iː/, /æ - ʌ/, ou /ɑː - ʌ/) et les deux dernières séances portaient sur l'ensemble des contrastes. L'ordre des 5 séances était pseudo randomisé entre les sujets, de telle sorte que les deux séances focalisées sur la région /æ - ʌ - ɑː/ soient consécutives.

Les stimuli présentés dans les séances 1 à 3 contenaient toutes les paires minimales choisies pour constituer les stimuli d'entraînement produites par six hommes et six femmes.

Pour les séances 4 et 5, les stimuli ont été divisés en deux listes contenant chacune une moitié des paires minimales pour chaque contraste. Ces listes sont présentées dans le Tableau XI.

Tableau XI : Listes de stimuli utilisés dans les séances 4 et 5

	/iː/	/ɪ/	/æ/	/ʌ/	/ɑː/
Liste 1	beak	bik	back	buck	Bark
	bean	bin	cap	cup	carp
	cheaps	chip	lack	luck	lark
	leak	lick	mat	mutt	mart
	mean	min	tat	tut	tart
Liste 2	deed	did	ban	bun	barn
	keep	kip	cad	cud	card
	peak	pick	cat	cut	cart
	seek	sick	pack	puck	park
	teak	tick	pat	putt	part

Les deux types d'entraînement ont été constitués de telle sorte que la quantité et le contenu du signal de parole reçu par les participants des deux groupes soient les plus similaires possibles, de même que les durées des séances. Ainsi, ce qui différencie profondément les deux entraînements proposés est la tâche réalisée par les participants : dans un cas, la tâche repose uniquement sur des mécanismes perceptifs (identification et discrimination), dans l'autre cas, la tâche demandée implique en plus des mécanismes de production de parole (imitation de mots).

L'ensemble des entraînements a été programmé avec l'interface *Demo window* du logiciel Praat[4]. Pour chacune des tâches, le participant pouvait voir en haut à droite de l'écran un compteur indiquant le nombre d'essais réalisés sur le nombre d'essais total à réaliser. L'ensemble des consignes étaient rédigées en anglais.

2.2.1 Entraînement en perception (Groupe PE)

L'entraînement en perception comportait des tâches d'identification (*2AFC* pour les séances 1 à 3 et *5AFC* pour les séances 4 et 5) et de discrimination (*AX* pour les séances 1 à 3 et *oddity* pour les séances 4 et 5) avec *feedback*. Deux types de tâches ont été mises en place car elles

[4] Les scripts des entraînements sont disponibles en libre accès à l'adresse suivante : https://github.com/emmanuelferragne/Entraînements_Phono

développent chacune des compétences différences. La tâche d'identification permet le développement de nouvelles catégories perceptives qui sont robustes à la variabilité, c'est-à-dire qu'elle incite les participants à regrouper perceptivement les traits phonétiques similaires en une même catégorie. La tâche de discrimination quant à elle oriente l'attention des participants sur les indices qui permettent de distinguer les catégories en les incitant à traiter les différences inter-catégorielles. Ainsi, les entraînements en identification favorisent l'acquisition des similarités alors que ceux en discrimination favorisent l'acquisition des différences. À la fin de chaque tâche, le nombre d'items total, le temps de réalisation de la tâche, le nombre d'items réussis et le nombre d'erreurs commises étaient présentés à l'écran.

2.2.1.1 Identification 2AFC

À chaque essai, le participant entendait un mot et devait indiquer quelle voyelle parmi deux affichées à l'écran était prononcée dans le mot entendu en cliquant sur un bouton à l'écran. L'intitulé des boutons contenait la voyelle en phonétique ainsi qu'un mot fréquent comportant cette voyelle. Les participants pouvaient réécouter le mot autant de fois qu'ils le souhaitaient, puis un *feedback* visuel leur était donné immédiatement après leur réponse. Si la réponse était incorrecte, l'indication « ***Try again :-(*** » apparaissait à l'écran, le participant réentendait le mot et devait choisir la bonne réponse. Lorsque la réponse était correcte, le message « ***Correct :-)*** » apparaissait à l'écran puis le mot suivant était présenté. Pour chacune des trois sessions, la tâche comportait 240 items (2 voyelles × 10 contextes CVC × 12 locuteurs (6 hommes et 6 femmes)). La Figure 20 présente les interfaces de la tâche d'identification à deux choix forcés.

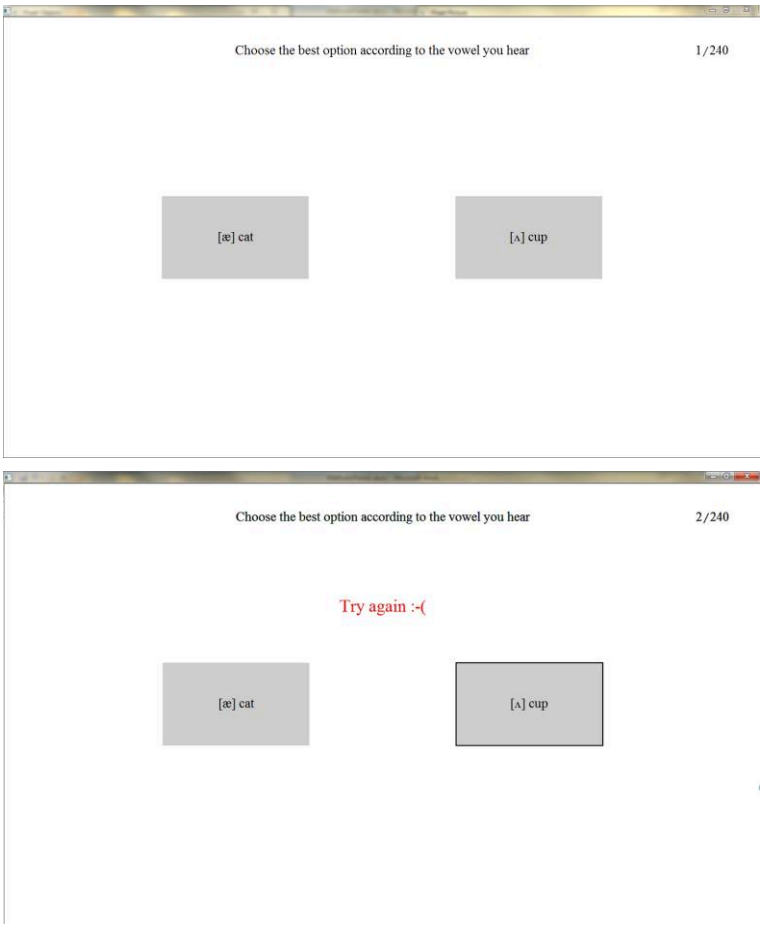

Figure 20 - Interfaces de la tâche d'identification à deux choix forcés de l'entraînement en perception

2.2.1.2 Discrimination AX

À chaque essai, le participant entendait une séquence de deux mots prononcés par deux locuteurs distincts mais de même sexe qui étaient présentés avec un intervalle interstimuli (ISI) de 500 ms. Le participant devait indiquer si les mots étaient identiques ou différents en cliquant sur le bouton approprié à l'écran. Le bouton « *Same* » était toujours positionné à gauche, alors que le bouton « *Different* » était positionné à droite. Il pouvait réécouter la séquence autant de fois qu'il voulait, et un *feedback* visuel lui était donné immédiatement après leur réponse. Si la réponse était incorrecte, l'indication « ***Try again :-(*** » apparaissait à l'écran, le participant réentendait la séquence et devait choisir la bonne réponse. Lorsque la réponse était correcte, le message « ***Correct :-)*** » apparaissait à l'écran puis la séquence suivante était présentée. La moitié des items étaient identiques, l'autre moitié, différents. Chaque paire minimal était présentée selon 4 combinaisons (e.g., *bik-bik, beak-beak, bik-beak* ou *beak-bik*) répétées 6 fois chacune, 3 fois par des voix d'hommes, 3 fois par des voix de femmes. Au total, le participant répondait à 240 items (10 contextes CVC × 4 combinaisons × 6 répétitions).

2.2.1.3 Identification 5AFC

Dans la tâche d'identification à cinq choix forcés (5 *Alternative Forced Choice*) les participants devaient indiquer quelle voyelle parmi les cinq proposées était prononcée en cliquant sur un bouton à l'écran. Les boutons à l'écran contenaient le symbole phonétique de la voyelle (/ɪ/, /iː/, /æ/, /ʌ/, ou /ɑː/) ainsi qu'un mot fréquent contenant la voyelle en question (*pig, feet, cat, cup et card*). Le participant recevait un *feedback* immédiatement après chaque réponse. Le *feedback* était le même que dans la tâche d'identification 2AFC des séances 1 à 3 (Correct vs. Incorrect). Le participant devait changer sa réponse jusqu'à ce qu'elle soit correcte pour passer à l'item suivant. Pour chacune des deux sessions, la tâche comportait 300 items (5 voyelles × 5 contextes CVC (liste1 ou liste2) × 12 locuteurs (6 hommes et 6 femmes)).

2.2.1.4 Discrimination *Oddity*

À chaque item de la tâche de discrimination du type *Oddity*, les participants entendaient une séquence de trois mots anglais prononcés par 3 locuteurs différents mais de même sexe avec un intervalle interstimuli (i.e., durée entre la fin d'un mot et le début du mot suivant) de 500 ms ; parmi ces trois mots, deux étaient identiques, un était différent. Le participant devait décider lequel des trois mots était l'intrus (*oddity task*) en cliquant sur l'un des trois boutons numérotés de 1 à 3 affichés à l'écran. Pour chaque paires de voyelles testées (/ɪ - iː/, /æ - ʌ/, ou /ɑː - ʌ/), les cinq paires minimales de la liste 1 ou liste 2, selon la session, étaient utilisées six fois chacune (e.g, la moitié avec *bik*, l'autre moitié avec *beak* comme intrus en première, deuxième ou troisième position). Les six combinaisons étaient prononcées par des hommes et des femmes. Pour chacune des deux sessions, la tâche comportait 180 items (3 contrastes × 5 paires minimales (liste1 ou liste2) × 6 combinaisons × 2 sexes). Après sa réponse, le participant recevait *feedback* le même type de *feedback* que dans la tâche de discrimination AX proposée lors des séances 1 à 3.

2.2.2 Entraînement en production (Groupe PR)

L'entraînement en production consistait en une tâche de répétition avec *feedback*. Les sessions 1 à 3 comportaient 240 items (i.e., 2 voyelles × 10 contextes CVC × 6 locuteurs × 2 sexes), alors que les séances 5 et 6 en comportaient 300 (5 voyelles × 5 contextes CVC × 6 locuteurs × 2 sexes).

À chaque item, le participant appuyait sur un bouton « *Click here to listen* » à l'écran pour entendre le mot à répéter. Le mot était présenté par l'intermédiaire d'un casque audio à un niveau d'écoute choisi par le participant. En haut de l'écran, les voyelles entrainées dans la session étaient affichées, accompagnées chacune d'un mot fréquent contenant cette voyelle. Lors de la présentation auditive du mot à répéter, le mot contenant effectivement la même voyelle était encadré en rouge (voir Figure 21). Le participant devait appuyer sur le bouton « *Click here to record* » afin d'enregistrer sa production. Deux secondes de signal était alors enregistrées à l'aide d'un micro AT2020-PRO. Le logiciel Praat ne permettant d'arrêter l'enregistrement dès qu'il n'y a plus de bruit, comme le ferait un dispositif équipé d'un algorithme VAD (*Voice Activity Detection*), comme il en existe sur certains dictaphones modernes, nous avons choisi d'enregistrer plus de signal que nécessaire afin d'être sûr d'avoir l'enregistrement nécessaire.

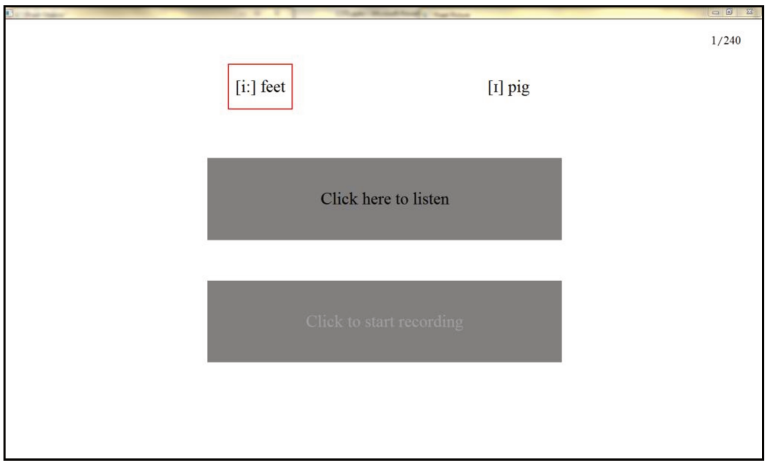

Figure 21 : Écran présenté dans la tâche de répétition pour écouter le mot à répéter. Dans cet exemple, le participant va entendre et répéter un mot contenant la voyelle /i:/ comme dans le mot « feet » encadré en rouge en haut de l'écran.

Un *feedback* sur la qualité de la voyelle était ensuite présenté à l'écran (voir Figure 22). Les 3 premiers formants de la voyelle produite, ainsi que sa durée, étaient automatiquement mesurés selon la procédure suivante. Les mots produits étant tous des CVC, la zone voisée a été considérée comme correspondant à la voyelle. Ainsi, une détection automatique du voisement permettant de repérer la voyelle était appliquée. Les valeurs des trois premiers formants étaient ensuite extraites au milieu de la zone voisée et la durée de la voyelle calculée.

Trois formes de *feedback* étaient simultanément présentées au participant :

1. une indication sur le timbre de la voyelle
2. une indication sur la durée de la voyelle
3. une indication sur la justesse de la production de type « correct » ou « incorrect ».

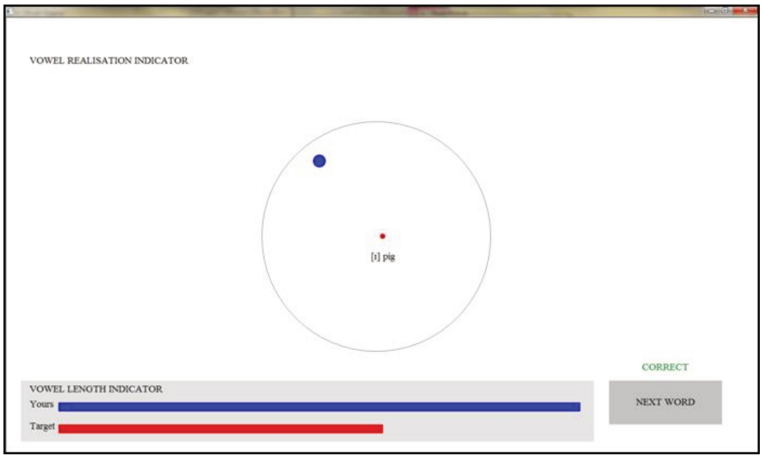

Figure 22 : *Feedback* visuel présenté lors de la tâche de répétition de mots dans l'entraînement en production. Dans cet exemple, le participant a produit une voyelle correctement catégorisée. On peut voir que la durée de la voyelle produite est plus longue que la moyenne des natifs.

La voyelle produite était donc représentée à l'écran dans le plan des deux premiers formants (point bleu). Au centre de l'écran le point rouge représentait la moyenne dans l'espace F1/F2 en Hz des productions pour cette voyelle de l'ensemble des locuteurs natifs enregistrés pour constituer le matériel des tests et entraînements dans la totalité des contextes consonantiques présentés dans le Tableau IV (p. 60). Ainsi, pour chaque voyelle et par sexe, 126 occurrences ont été utilisées (14 contextes consonantiques × 9 locuteurs). Ces valeurs moyennes sont présentées dans le Tableau XII. En bas de l'écran, la durée de la voyelle était représentée (barre bleue) ainsi que la durée moyenne des voyelles correspondantes enregistrées par des locuteurs natifs de même sexe que le participant (barre rouge). La taille de la barre rouge était fixe, et la longueur de la barre bleue représentait la distance au modèle de durée en pourcentage. À partir des paramètres acoustiques extraits (F1, F2, F3 en Hz et la durée), un *feedback* sur la qualité de la voyelle était de plus donné à partir de la classification s'appuyant sur le modèle d'analyse discriminante linéaire (*LDA*) appris sur les productions des locuteurs natifs de même sexe que le participant. De plus, un cercle gris autour de la cible indiquait une distance euclidienne à la cible de 200 Hz dans l'espace des deux premiers formants. Cette distance maximale, déterminée empiriquement, a été indiquée car l'analyse discriminante ne prenant en compte que 5 voyelles dans la totalité de l'espace vocalique, nous voulions éviter qu'une production réalisée avec des valeurs formantiques aberrantes soit correctement classifiée du seul fait de sa plus grande proximité avec la catégorie cible qu'avec les autres.

Tableau XII : Valeurs moyennes des trois premiers formants (en Hz) et de la durée des voyelles (en ms) utilisées pour représenter les valeurs cibles en fonction du sexe du participant dans l'entraînement en production. Les écart-types sont indiqués entre parenthèses.

Voyelle	Sexe	N	F1	F2	F3	Durée
/æ/	F	126	990 (146)	1676 (120)	2924 (192)	159 (64)
	H	126	741 (100)	1526 (90)	2545 (192)	149 (51)
/ʌ/	F	126	807 (108)	1426 (117)	2995 (266)	119 (30)
	H	126	674 (54)	1197 (112)	2642 (161)	123 (29)
/ɑː/	F	126	757 (68)	1242 (84)	3075 (231)	271 (52)
	H	126	632 (59)	1093 (94)	2694 (164)	250 (67)
/ɪ/	F	126	554 (60)	2351 (175)	3112 (216)	124 (33)
	H	126	424 (38)	1968 (120)	2679 (179)	116 (31)
/iː/	F	126	373 (36)	2838 (128)	3483 (257)	188 (71)
	H	126	308 (43)	2250 (154)	3004 (274)	183 (69)

Lorsque la voyelle produite était correctement classifiée et que la distance euclidienne à la cible ne dépassait pas 200 Hz, la mention « **CORRECT** » était affichée à l'écran et le mot suivant était présenté. Dans le cas contraire, « **Please Try again** » était affiché si le participant n'avait pas dépassé deux essais et le même mot était présenté à nouveau. Après le troisième essai échoué, « **INCORRECT, go to the next word** » apparaissait et le participant passait au mot suivant. La durée maximale de la session était d'une heure.

À la fin de chaque session, la durée de la session, le nombre de mots présentés et les pourcentages de réussites total et par voyelle étaient présentés au participant.

3 Évaluation en perception

Les performances des participants en perception ont été évaluées au moyen d'une tâche d'identification à 5 choix forcés et d'une tâche de discrimination de type *oddity* en adaptant la méthodologie d'Iverson et collègues (2012).

Les scores des trois groupes d'apprenants (C, PE et PR) ont été comparés entre eux et aux scores des locuteurs natifs avant et après entraînement séparément. Puis l'efficacité des entraînements a été évaluée en comparant les scores de chacun des groupes avant et après entraînement.

3.1 Hypothèses

Nos hypothèses générales concernant les performances des participants dans les tâches de perception étaient les suivantes :

H1 : Avant entraînement, les performances globales des trois groupes d'apprenants devraient être équivalentes et toutes inférieures à celles des locuteurs natifs. Nous n'attendons pas de différence entre voyelle pour le groupe N.

H2 : Après entraînement, les performances des trois groupes d'apprenants devraient toujours rester plus faibles que celles des locuteurs natifs. Les performances globales des groupes d'apprenants ne devraient plus être équivalentes. Nous prévoyons tout d'abord que les performances des participants ayant suivi un entraînement (groupes PE et PR) seront supérieures à celles du groupe contrôle n'ayant pas suivi d'entraînement spécifique. De plus, nos deux types d'entraînement ayant été constitués de telle sorte que la quantité et le contenu du signal de parole reçu lors des entraînements soient le plus similaire possible, nous espérons que l'entraînement basé sur l'imitation de mot soit plus efficace que celui portant sur des tâches de perception uniquement. Ainsi, les performances globales du groupe PR devraient être les meilleures, suivies de celles du groupe PE et enfin de celles du groupe C.

H3 : Les comparaisons des performances avant et après entraînement devraient montrer une plus grande amélioration des performances pour le groupe PR, que pour le groupe PE, que pour le groupe C.

3.2 Méthodologie

3.2.1 Description des tâches utilisées

Les deux tâches de perception (5AFC identification et *oddity* discrimination) ont été programmées avec l'interface *Demo window* du logiciel Praat.

3.2.1.1 Tâche d'identification à 5 choix forcés

Les participants réalisaient une tâche d'identification à cinq choix forcés (5 *Alternative Forced Choice*). Les mots étaient présentés au moyen d'un casque audio (BeyerdynamicDT 770 PRO) à un niveau d'écoute confortable. La tâche comportait 120 essais (5 voyelles × 4 contextes CVC × 6 locuteurs (3 hommes et 3 femmes)). Les participants devaient indiquer quelle voyelle parmi les cinq proposées était prononcée en cliquant sur un bouton à l'écran. Les boutons à l'écran contenaient le symbole phonétique de la voyelle (/ɪ/, /iː/, /æ/, /ʌ/, ou /ɑː/) ainsi qu'un mot fréquent contenant la voyelle en question (*pig, feet, cat, cup et card*) (voir Figure 23). Ils ne recevaient aucun *feedback* et ne pouvaient pas réécouter le mot. Après cinq items d'essai, les stimuli étaient présentés dans un ordre aléatoire.

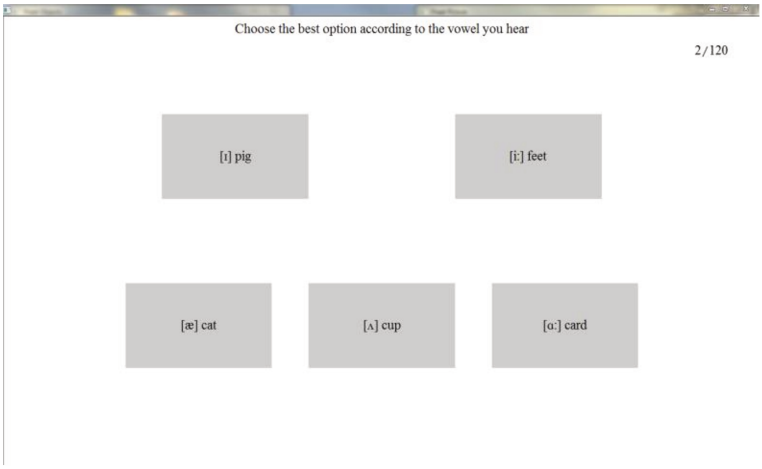

Figure 23 : Écran présenté lors de la tâche d'identification à 5 choix forcés

3.2.1.2 Tâche de discrimination *oddity*

Les participants réalisaient ensuite une tâche de discrimination de détection d'intrus (*Oddity).* À chaque item les participants entendaient une séquence de trois mots anglais prononcés par 3 locuteurs différents de même sexe avec un intervalle interstimuli (i.e., durée entre la fin d'un mot et le début du mot suivant) de 500 ms ; parmi ces trois mots, deux étaient identiques et un était différent. Le participant devait décider lequel des trois mots était l'intrus (*oddity task*) en cliquant sur l'un des trois boutons numérotés affichés à l'écran comme présenté dans la Figure 24. Ils ne recevaient aucun *feedback* et ne pouvaient pas réécouter la séquence. Pour chaque paire de voyelles testées (/ɪ - iː/, /æ - ʌ/, ou /ɑː - ʌ/), quatre paires minimales (e.g., *bid – bead, bit – beat, hid – heed, hit – heat*) étaient présentées six fois chacune (e.g, la moitié avec *bit*, l'autre moitié avec *beat* comme intrus en première, deuxième ou troisième position) et prononcées soit par des voix d'hommes ou de femmes. La tâche comportait 144 items (3 contrastes × 4 paires minimales × 6 combinaisons × 2 sexes). Après six items d'essai identiques pour tous les participants, les stimuli étaient présentés dans un ordre aléatoire variable d'un participant à l'autre.

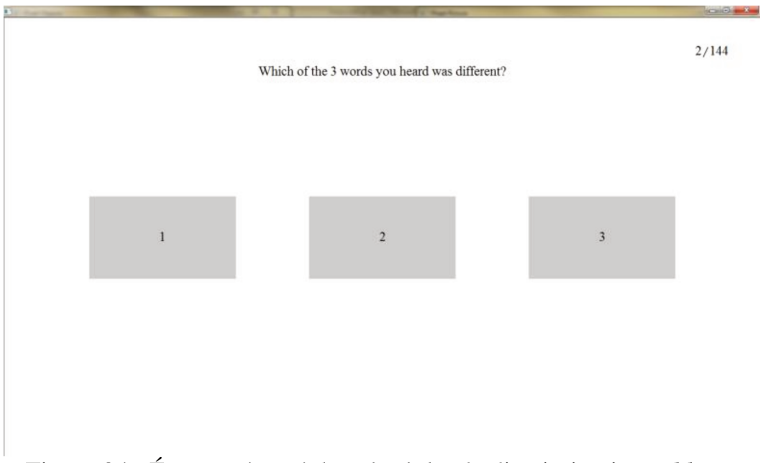

Figure 24 : Écran présenté dans la tâche de discrimination *oddity*

3.2.2 Traitement des données

Pour chaque participant, les scores de sensibilité *dprime d'* (Abdi, 2007 ; Macmillan & Creelman, 2005) ont été calculés en pré-test et en post-test, pour chaque voyelle pour la tâche d'identification, pour chaque paire de voyelles testées (/ɪ - iː/, /æ - ʌ/, ou /ɑː - ʌ/) pour la tâche de discrimination. Les scores de *d'* de la tâche d'identification ont été calculés avec la fonction `dprime.mAFC` du package `psyphy` développé pour le logiciel R. Les *d'* pour la tâche de discrimination ont été calculés avec la fonction `discrim` du package `sensR` en précisant la méthode triangle (Christensen et Brockhoff, 2010). Les analyses statistiques ont été réalisées

avec les packages `lme4` (Bates, Maechler, Walker et Matrix, 2015) et `psycho`. Des modèles linéaires mixtes ont été utilisés et des ANOVA de type III[5] ont été appliquées sur ces modèles pour obtenir les valeurs de *p*. Des comparaisons multiples ont été appliquées aux plus hauts niveaux d'interactions significatifs avec la méthode de Tukey. Les tailles d'effets sont rapportées par les étas carrés partiels (Lakens, 2013)[6].

3.3 Résultats

L'observation préliminaire des résultats en identification a montré que deux fichiers sons enregistrés par une même locutrice pour la voyelle /ɑː/ (A_heart_F03.wav et A_bart-F03.wav) étaient respectivement mal identifiés par 16 natifs sur 16 et 12 natifs sur 16 alors que tous les autres fichiers étaient bien identifiés par 11 natifs sur 16 au minimum. Pour ces deux fichiers, les locuteurs natifs répondaient préférentiellement en faveur de la voyelle /ʌ/. Pour ne pas abaisser les performances des apprenants francophones à cause de ces prononciations ambiguës, nous avons décidé de les retirer des analyses. Ainsi, les scores de *d'* pour la voyelle /ɑː/ ont été calculés sur 22 items au lieu de 24 dans les tâches d'identification et sur 36 items au lieu de 48 pour le contraste /ʌ - ɑː/ dans la tâche de discrimination.

Les résultats seront présentés de la manière suivante pour les tâches d'identification et de discrimination : d'abord les performances des participants seront comparées entre les différents groupes (C, PE, PR et N) indépendamment en prétest et posttest, puis l'effet des entraînements sera évalué en comparant les différences entre prétest et posttest entre les groupes d'apprenants.

---

[5] Les anovas de type III testent la présence d'un effet principal après les autres effets principaux et interaction. En effet, la somme des carrés permettant de calculer l'effet d'un facteur est ajustée pour prendre en compte l'effet des autres facteurs. Ainsi, cette approche est valide en présence d'interactions significatives (Myers, Well et Lorch, 2010).

[6] Les étas carrés partiels se calculent en faisant le rapport de la somme des carrés de l'effet à la somme de la somme des carrés de l'effet et de la somme des carrés de l'erreur : $\eta_p^2 = \frac{SC_{effet}}{SC_{effet}+SC_{erreur}}$. Comparativement aux étas carrés, qui se calculent comme le rapport de la somme des carrés de l'effet à la somme des carrés totale : $\eta^2 = \frac{SC_{effet}}{SC_{totale}}$ et sont donc ainsi dépendant du *design* de l'étude, les étas carrés partiels augmentent la comparabilité des tailles d'effets entre études à design équivalents.

3.3.1 Comparaison avec les locuteurs natifs

3.3.1.1 Résultats en identification

Les indices de sensibilité *d'* des apprenants des trois groupes dans la tâche d'identification en prétest et en posttest ont été comparés séparément à ceux des locuteurs natifs du groupe N avec deux modèles linéaires mixtes incluant les facteurs Groupe (N, C, PE et PR) et Voyelle (/ɪ/, /iː/, /æ/, /ʌ/, et /ɑː/) et l'interaction Groupe × Voyelle comme effets fixes et le facteur Sujet comme facteur aléatoire. Les scores du groupe Natif sont présentés dans la Figure 25.

Les scores des groupes C, PE et PR selon le test et la voyelle sont représentés dans la Figure 26. Les résultats sont représentés sous forme de diagrammes en violon. Sur ces figures, la barre horizontale en gras représente la moyenne, la boîte représente l'intervalle interquartile et le « violon » indique la dispersion des données. Chaque point représente la moyenne d'un participant.

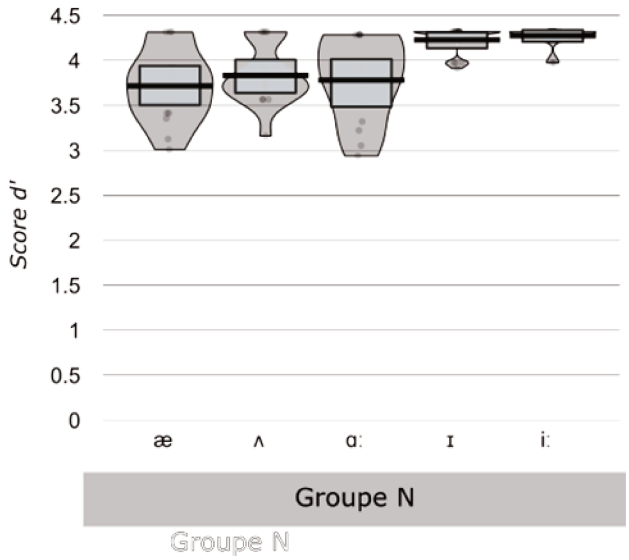

Figure 25 : Scores en identification pour le groupe natif pour chacune des voyelles testées. Un point représente le score *d'* d'un participant pour une voyelle donnée.

Chez les natifs, on observe que les performances sont très élevées. Elles plafonnent pour les voyelles de la région du [i] et sont plus dispersées pour les voyelles de la région du [a]. Cette observation est cohérente avec la dispersion des catégories de voyelles selon les deux régions : en effet nous avons pu observer que les voyelles de la région du [i] ne présentaient aucun chevauchement dans l'espace des deux premiers formants et qu'elles étaient de plus produites avec des durées vocaliques bien différentes. En revanche, les voyelles de la région du [a] étaient produites avec plus de chevauchement intervocalique, surtout les voyelles /ʌ/ et /ɑː/ ce qui a

pu conduire, malgré des différences de durées bien présentes à plus de confusion que dans la région du [i]. De manière générale, on constate que les scores *d'* sont relativement élevés, y compris dans les groupes d'apprenants, ce qui indique que la tâche qu'ils réalisent n'est pas trop difficile. Cependant, la tâche nous permet quand même d'observer des différences entre les groupes et voyelles.

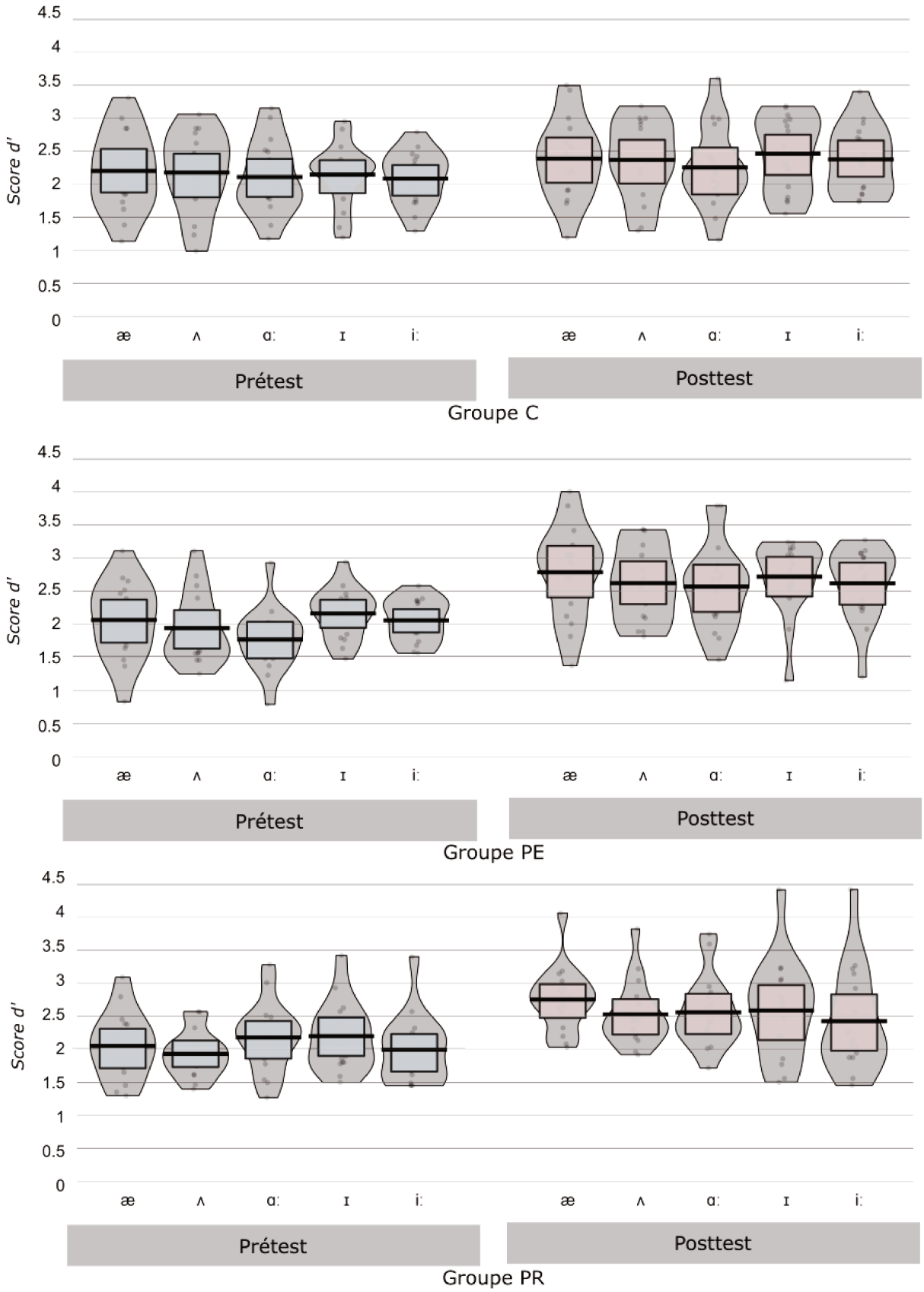

Figure 26 : Scores des 3 groupes C, PE et PR en identification en prétest (en bleu) et posttest (en rouge) pour chacune des voyelles testées. Un point représente le score *d'* d'un participant pour une voyelle donnée.

Le Tableau XIII montre les résultats de l'analyse des scores d'identification **en prétest :**

Tableau XIII : Table d'analyse du modèle linéaire mixte pour les scores *d'* en identification en prétest[7]

Source	SC	CM	ddl_{effet}	ddl_{erreur}	*F*	*p*	η^2_p
Groupe	54.497	18.166	3	60	145.127	0.000***	0.631
Voyelle	2.300	0.575	4	240	4.594	0.001**	0.067
Groupe:Voyelle	4.445	0.370	12	240	2.960	0.001***	0.122

En prétest, l'ANOVA montre un effet principal du Groupe ($\eta^2_p = .63$) qui reflète les scores plus élevés du groupe N par rapport aux trois groupes C, PE et PR (tous les $p < .001$). Les trois groupes C, PE et PR ne présentent pas de différences entre eux. L'effet principal de la Voyelle est significatif ($\eta^2_p = .07$) et l'interaction Groupe × Voyelle est également significative ($\eta^2_p = .12$). Des comparaisons post-hoc par paires pour cette interaction montre que i) toutes les voyelles sont mieux identifiées par les locuteurs natifs (tous les $p < .001$), ii) il n'y a pas de différences significatives dans les scores d'identification des trois groupes C, PE et PR pour toutes les voyelles (tous les $p > .5$) et iii) on peut noter que /ɪ/ est mieux identifié que /iː/ dans les groupes PE, $M_{/ɪ/-/iː/} = 0.71$, $t(240) = 3.64$, $p = .042$ et PR, $M_{/ɪ/-/iː/} = 0.94$, $t(120) = 4.82$, $p < .001$. Le /ɪ/ est également significativement mieux identifié que le /ɑː/ dans le groupe PE, $M_{/ɪ/-/ɑː/} = 0.78$, $t(240) = 4.03$, $p = .011$, et que le /ʌ/ dans le groupe PR, $M_{/ɪ/-/ʌ/} = 0.95$, $t(240) = 4.86$, $p < .001$. Il n'y a pas de différences significatives dans les groupes pour toutes autres voyelles.

Le Tableau XIV montre les résultats de l'analyse des scores en identification **en posttest :**

Tableau XIV : Table d'analyse du modèle linéaire mixte pour les scores *d'* en identification en posttest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	*F*	*p*	η^2_p
Groupe	21.193	7.064	3	60	50.949	0.000^{***}	0.380
Voyelle	1.614	0.404	4	240	2.910	0.022^{*}	0.045
Groupe:Voyelle	4.475	0.373	12	240	2.689	0.002^{**}	0.115

En posttest, l'ANOVA montre qu'il y a toujours un effet principal du Groupe ($\eta^2_p = .38$), avec le groupe N ayant des scores plus élevés que tous les autres groupes (tous les $p < .001$). L'effet principal de la Voyelle est également significatif ($\eta^2_p = .04$) mais cette fois seulement du fait

[7] Dans toutes les tables d'analyses, les astérisques indiquent les seuils de signification comme suit : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

de la différence entre les voyelles /ɑː/ et /ɪ/. De nouveau l'interaction Groupe × Voyelle est significative (η^2_p = .08). Les comparaisons post-hoc montrent tout d'abord que le groupe N a des performances supérieures aux autres groupes pour toutes les voyelles (tous les p <.05) sauf pour la voyelle /æ/ qui, contrairement au prétest, n'est plus significativement moins bien identifiée par les groupes PE, $M_{N\text{-}PE}$ = 0.66, $t(275) = 3.01$, $p = .23$ et PR, $M_{N\text{-}PR}$ = 0.67, $t(275) = 3.06$, $p = .21$ comme nous l'observions en prétest. De nouveau, il n'y a pas de différences significatives entre les trois groupes d'apprenants (C, PE et PR) pour toutes les voyelles (tous les $p > .9$). On peut noter que le /ɪ/ est toujours mieux identifié que le /iː/ pour les groupes PE, $M_{/ɪ/\text{-}/iː/} = 0.77$, $t(240) = 3.81$, $p = .024$, et PR, $M_{/ɪ/\text{-}/iː/} = 0.91$, $t(240) = 4.54$, $p = .001$, mais aussi par le groupe C, $M_{/ɪ/\text{-}/iː/} = 0.82$, $t(240) = 4.08$, $p = .009$, ce qui n'était pas le cas lors du prétest. On peut enfin noter que le groupe PR présente des scores d'identification supérieurs pour le /æ/ par rapport au /iː/, $M_{/æ/\text{-}/iː/} = 0.80$, $t(240) = 3.97$, $p = .014$ et que le groupe C identifie significativement mieux le /ɪ/ par rapport au /ɑː/, $M_{/ɪ/\text{-}/ɑː/} = 0.73$, $t(240) = 3.61$, $p = .045$. Les autres voyelles ne sont pas identifiées de manière significativement différente entre les groupes.

3.3.1.2 Résultats en discrimination

Les scores *d'* en discrimination ont été analysés de la même manière que ceux en identification à la différence que le facteur Voyelle a été remplacé par le facteur Contraste (/ɪ - iː/, /æ - ʌ/, et /ɑː - ʌ/). Les scores du groupe N en discrimination sont présentés dans la Figure 27. Ceux des groupes C, PE et PR sont présentés dans la Figure 28.

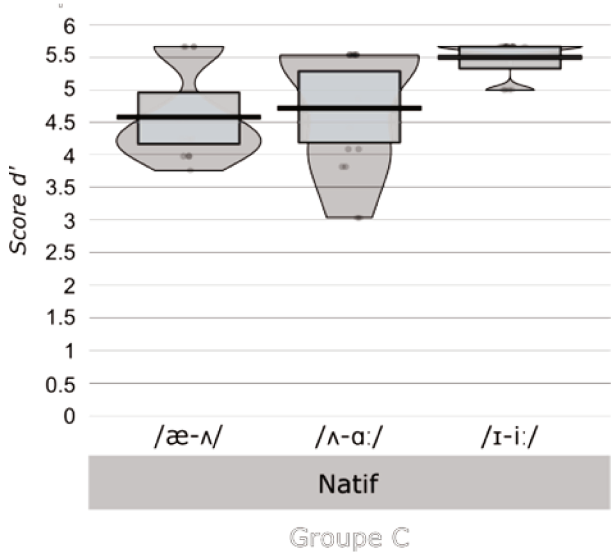

Figure 27 : Scores en discrimination du groupe N

Les locuteurs natifs présentent des scores homogènes qui plafonnent pour le contraste /ɪ-iː/, ce qui n'est pas le cas pour les deux autres contrastes dont la distribution est plus dispersée avec des scores qui ne plafonnent pas, ce qui suit le schéma de leurs scores en identification.

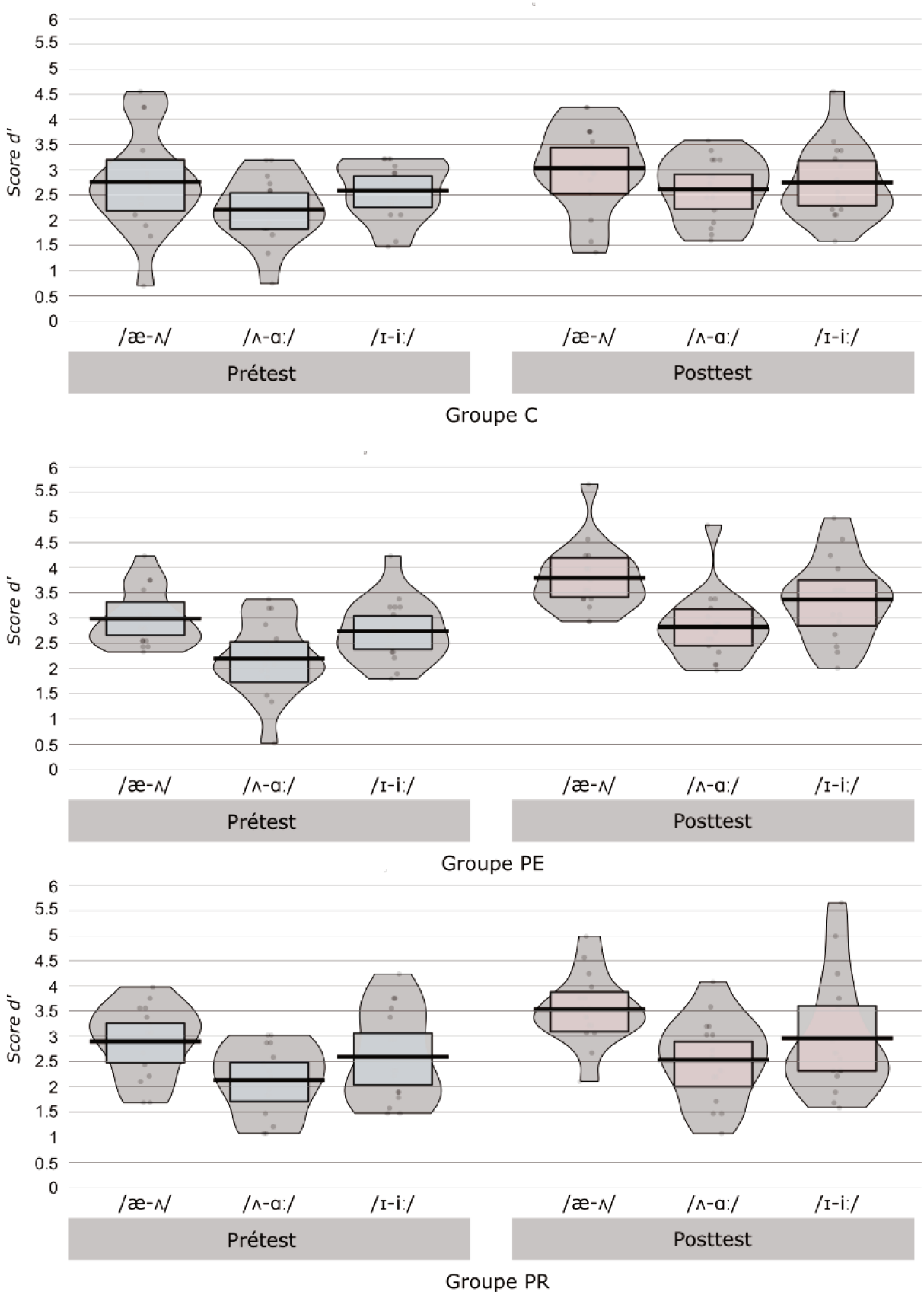

Figure 28 : Scores dans la tâche de discrimination pour chacun des groupes selon le contraste. Un point représente le score moyen d'un participant.

En prétest, comme le montre le Tableau XV, l'effet principal du Groupe est significatif ($\eta^2_p = .63$), reflétant les me illeurs scores dans le groupe N par rapport aux autres groupes (tous les $p < .001$). L'effet principal du Contraste est également significatif ($\eta^2_p = .17$), avec le contraste /ɑː - ʌ/ qui est significativement moins bien discriminé que les deux autres contrastes /æ - ʌ/ et /iː - ɪ/. L'interaction Groupe × Contraste est significative ($\eta^2_p = .13$). Les analyses post-hoc montrent que le groupe N a de meilleures performances que les trois autres groupes pour tous les contrastes (tous les $p < .001$) et que les trois groupes d'apprenants ne diffèrent pas entre eux et ce pour tous les contrastes (tous les $p > .99$). On peut remarquer que le contraste /ɪ - iː/ est significativement mieux discriminé par les natifs que les deux autres contrastes, $M_{/ɪ\text{-}iː/\text{-}/æ\text{-}ʌ/} = 0.92$, $t(120) = 4.16$, $p = .003$ et $M_{/ɪ\text{ - }iː/\text{-}/ɑː\text{ - }ʌ/} = 0.78$, $t(120) = 3.54$, $p = .027$. Le contraste /æ - ʌ/ est mieux discriminé que le contraste /ɑː - ʌ/ par les groupes PE et PR, $M_{/æ\text{-}ʌ/\text{-}/ɑː\text{-}ʌ/} = 0.79$, $t(120) = 3.58$, $p = .024$ et $M_{/æ\text{-}ʌ/\text{-}/ɑː\text{-}ʌ/} = 0.77$, $t(120) = 3.49$, $p = .031$ respectivement.

Tableau XV : Table d'analyse du modèle linéaire mixte pour les scores *d'* en discrimination en prétest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	F	p	η^2_p
Groupe	96.132	32.044	3	60	82.727	0.000***	0.625
Contraste	11.442	5.721	2	120	14.770	0.000***	0.166
Groupe:Contraste	8.871	1.478	6	120	3.817	0.002**	0.133

Les effets principaux du Groupe ($\eta^2_p = .47$) et du Contraste ($\eta^2_p = .18$) sont significatifs **en posttest** également (voir Tableau XVI). Le groupe Natif a toujours des performances plus élevées que les groupes C, PE et PR (tous les $p < .001$). Au sujet de l'effet du Contraste, le post-hoc montre que le contraste /ɑː - ʌ/ est moins bien identifié que les deux autres contrastes testés. L'interaction Groupe × Contraste est également significative ($\eta^2_p = .21$). Le post-hoc pour cette interaction montre que les scores du groupe Natif restent significativement meilleurs que ceux des autres groupes pour tous les contrastes, à l'exception du contraste /æ - ʌ/ qui n'est plus moins bien identifié par le groupe PE que par les natifs comme c'était le cas en prétest, $M_{\text{N-PE}} = 0.78$, $t(120) = 2.82$, $p = .18$. Comme en prétest, les scores des trois groupes d'apprenants ne sont pas significativement différents entre eux pour les trois contrastes (tous les $p > .05$). Le contraste /æ - ʌ/ est toujours significativement mieux discriminé que le contraste /ɑː - ʌ/ par les groupes PE et PR, $M_{/æ\text{-}ʌ/\text{-}/ɑː\text{-}ʌ/} = 0.97$, $t(138) = 4.46$, $p = .001$ et $M_{/æ\text{-}ʌ/\text{-}/ɑː\text{-}ʌ/} = 1.00$, $t(138) = 4.62$, $p < .001$ respectivement.

Tableau XVI : Table d'analyse de modèle linéaire mixte pour les scores *d'* en discrimination en posttest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	F	p	η^2_p
Groupe	46.892	15.631	3	60	41.465	0.000***	0.469
Contraste	11.714	5.857	2	120	15.537	0.000***	0.181
Groupe:Contraste	13.259	2.210	6	120	5.862	0.000***	0.200

3.3.1 Évaluation de l'effet des entraînements

Les scores **d'identification** des trois groupes C, PE et PR ont été comparés avec un modèle linéaire mixte incluant les facteurs Test (prétest vs. posttest), Groupe (C, PE et PR), Voyelle comme effets fixes et le facteur Sujet comme effet aléatoire. Toutes les interactions entre les facteurs ont été spécifiées dans le modèle.

Une ANOVA sur ce modèle (voir Tableau XVII) montre un effet principal du Test ($\eta^2_p = .26$) et une interaction Test × Groupe significative ($\eta^2_p = 0.24$). Les analyses post-hoc indiquent que seuls les deux groupes PE, $M_{T2\text{-}T1} = 0.55$, $t(405) = 9.42$, $p = < .001$ et PR, $M_{T2\text{-}T1} = 0.40$, $t(405) = 4.22$, $p = < .001$, ont amélioré significativement leurs performances générales entre prétest et posttest. L'effet principal du facteur Voyelle est significatif ($\eta^2_p = .02$), mais aucune des interactions incluant ce facteur (Groupe × Voyelle, Test × Voyelle et l'interaction à trois niveaux) n'est significative. Il semble donc que les progressions entre prétest et posttest ne soient pas différentes entre les voyelles.

Tableau XVII : Table d'analyse du modèle linéaire mixte des scores *d'* en identification des groupes d'apprenants en prétest et posttest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	*F*	*p*	η^2_p
Test	23.946	23.946	1	405	146.785	0.000***	0.264
Groupe	0.046	0.023	2	45	0.142	0.868	0.001
Voyelle	1.650	0.412	4	405	2.528	0.040*	0.024
Test:Groupe	3.653	1.826	2	405	11.196	0.000***	0.052
Test:Voyelle	0.214	0.053	4	405	0.328	0.859	0.003
Groupe:Voyelle	1.167	0.146	8	405	0.894	0.521	0.017
Test:Groupe:Voyelle	0.877	0.110	8	405	0.672	0.716	0.013

Afin de déterminer si les trois groupes ont progressé de la même manière, les différences entre prétest et posttest ont été calculées pour chaque voyelle et chaque participant (Figure 29, à gauche). Un modèle linéaire mixte avec pour variable expliquée les différences entre prétest et posttest et comme facteurs fixes le Groupe et la Voyelle et le facteur aléatoire Sujet a été généré (voir Tableau XVIII).

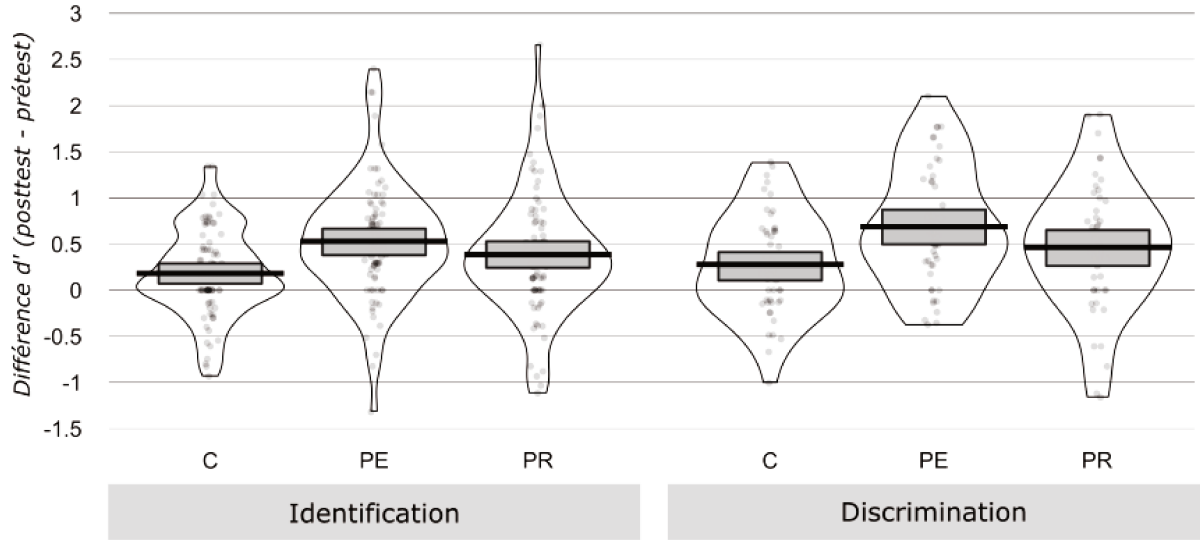

Figure 29 : Amélioration entre prétest et posttest dans la tâche d'identification à gauche et de discrimination à droite. À gauche, un point indique la différence entre prétest et posttest d'un individu pour une voyelle, à droite pour un contraste

Tableau XVIII : Table d'analyse de la différence des scores *d'* entre prétest et posttest en identification

Source	SC	CM	ddl_{effet}	ddl_{erreur}	*F*	*p*	η^2_p
Groupe	2.671	1.336	2	45	7.383	0.002**	0.070
Voyelle	0.428	0.107	4	180	0.591	0.669	0.012
Groupe:Voyelle	1.754	0.219	8	180	1.212	0.294	0.047

Une ANOVA sur ce modèle montre un effet significatif du Groupe ($\eta^2_p = 0.07$). Les analyses post-hocs montrent que le groupe PE s'améliore significativement plus que le groupe C, $M_{PE\text{-}C} = 0.34$, $t(45) = 3.68$, $p = < .005$, et que le groupe PR s'améliore de manière marginale par rapport au groupe C, $M_{PR\text{-}C} = 0.20$, $t(45) = 2.14$, $p = < .09$. Les différences ne sont pas significatives entre les groupes PE et PR. L'effet du facteur Voyelle et l'interaction Groupe × Voyelle ne sont pas significatifs.

L'analyse des scores en **discrimination** a été effectuée de la même manière en remplaçant le facteur Voyelle par le facteur Contraste. Des résultats similaires ont été observés puisqu'aucune interaction incluant le facteur Contraste n'était pas significative. L'interaction Test × Groupe est marginale, avec le groupe PE, qui semble améliorer davantage ses

performances en discrimination que les groupes PR puis C. Comme pour les données d'identification, les différences entre prétest et posttest en discrimination (voir Figure 29 droite) ont été modélisées dans un modèle linéaire mixte avec les facteurs fixes Groupe et Contraste, et le facteur aléatoire Sujet dont les résultats sont présentés dans le Tableau XIX.

Tableau XIX : Table d'analyse du modèle linéaire mixte des scores *d'* en discrimination des groupes d'apprenants en prétest et posttest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	*F*	*p*	η^2_p
Test	16.427	16.427	1	225	47.728	0.000***	0.170
Groupe	0.987	0.493	2	45	1.433	0.249	0.012
Contraste	27.068	13.534	2	225	39.322	0.000***	0.252
Test:Groupe	2.015	1.007	2	225	2.927	0.056.	0.024
Test:Contraste	0.432	0.216	2	225	0.628	0.535	0.005
Groupe:Contraste	1.825	0.456	4	225	1.325	0.261	0.022
Test:Groupe:Contraste	0.330	0.083	4	225	0.240	0.915	0.004

Le modèle, dont le résultat est présenté dans le Tableau XX, montre un effet significatif du Groupe ($\eta^2_p = 0.07$), signe d'une différence significative entre les deux groupes C et PE, $M_{PE-C} = 0.41$, $t(45) = 3.12$, $p = < .01$. L'effet du facteur Contraste et l'interaction Groupe × Contraste ne sont pas significatifs.

Tableau XX : Table d'analyse de la différence des scores *d'* entre prétest et posttest en discrimination

Source	SC	CM	ddl_{effet}	ddl_{erreur}	*F*	*p*	η^2_p
Groupe	4.029	2.015	2	45	4.881	0.009**	0.067
Contraste	0.864	0.432	2	135	1.047	0.354	0.015
Groupe:Contraste	0.661	0.165	4	135	0.400	0.808	0.012

Enfin, des tests de corrélation réalisés entre les scores en identification et en discrimination montrent que ces mesures sont positivement corrélées en prétest, $r(46) = 3.73$, $p < .001$, et en posttest, $R(46) = 4.63$, $p < .0001$. De plus, des tests de corrélations réalisés sur les différences entre prétest et posttest des scores en identification et en discrimination montrent une corrélation positive entre l'amélioration en identification et en discrimination, $r(46) = 2.18$ $p < .05$, c'est-à-dire que les participants qui améliorent le plus leurs performances en identification sont ceux qui améliorent le plus celles en discrimination.

3.4 Discussion des résultats en perception

Les performances en perception des apprenants francophones ont été évaluées au moyen d'une tâche d'identification et d'une tâche de discrimination avant et après qu'ils aient suivi soit un entraînement de type *High Variability Phonetic Training* classique (groupe PE), soit un

entraînement plus spécifiquement orienté sur la production (groupe PR). Ces performances ont été comparées à celles d'un groupe d'apprenants n'ayant pas suivi d'entraînement spécifique (groupe C) ainsi qu'à celles de locuteurs natifs (groupe N) réalisant les mêmes tâches.

Nous avions formulé une première hypothèse selon laquelle les performances des apprenants seraient inférieures à celles des natifs pour toutes les voyelles/contrastes en prétest et nous pensions que les groupes d'apprenants ne présenteraient pas de différences au prétest. Selon cette première hypothèse, nous nous attendions également à ce que toutes les voyelles soient perçues de manière équivalente par les locuteurs natifs.

Selon notre première hypothèse H1, nous nous attendions tout d'abord à ce que les performances des locuteurs natifs plafonnent pour toutes les voyelles/contrastes. Les résultats en identification pour le groupe de locuteurs natifs vont bien dans ce sens, puisqu'aucune des cinq voyelles ne se distingue significativement des autres. Cependant, en discrimination, on observe que le contraste de la région du [i] est significativement mieux discriminé que ceux de la région du [a]. L'effet n'est pas significatif sur les scores en identification mais on voit sur les données qu'elles semblent plafonner et présenter une faible variabilité pour les deux voyelles du [i] alors que celles de la région du [a] présentent une plus grande variabilité et sont identifiées avec moins de justesse. Une hypothèse explicative à cette différence serait que les voyelles de la région du [a] utilisées dans les stimuli présentent des proportions de chevauchement des catégories dans l'espace des deux premiers formants plus importants que celle de la région du [i]. En effet, dans les stimuli utilisés dans les tests, représentés dans la Figure 17 p. 64, les voyelles /iː/ et /ɪ/ se distribuent sur des espaces acoustiques très distincts dans ces deux dimensions alors que les paires /æ/-/ʌ/ et /ʌ/-/ɑː/ surtout présentent des chevauchements inter-catégoriels. Pour expliquer ce phénomène, nous pouvons nous tourner vers des théories postulant que les inventaires phonologiques des langues dans leurs réalisations acoustiques sont constitués de telle sorte qu'ils s'adaptent à la perception. Or plusieurs théories s'accordent à dire que la relation entre l'espace acoustique et l'espace perceptif n'est pas linéaire. Nous pouvons par exemple citer la théorie de l'aimant perceptif, que nous avons décrite dans la Partie 1, selon laquelle les prototypes des catégories, c'est-à-dire les meilleurs exemplaires, contractent l'espace perceptif autour d'eux et rendent ainsi les exemplaires qui leur sont proches difficiles à discriminer (Iverson & Kuhl, 1995). En plus de cette théorie, nous pouvons citer la théorie quantique (Stevens, 1989 ; Stevens & Keyser, 2010) selon laquelle les relations entre les gestes articulatoires et leurs résultats acoustiques

n'auraient pas la même conséquence selon la région de l'espace acoustique dans lequel ils sont produits. En particulier, cette théorie prédit que les voyelles cardinales /i/, /a/, et /u/ parfois /y/ et /æ/ (Clements et Ridouane, 2006) devraient montrer moins de variations que les voyelles plus centrales car ces voyelles sont produites à des endroits du conduit vocal où des perturbations articulatoires produisent seulement de faibles changements dans les fréquences de formants. Dans notre cas, les voyelles /ʌ/ et /ɑː/ ne font pas parties de ces voyelles quantiques ce qui expliquerait pourquoi elles sont produites avec moins de stabilité que les voyelles de la région du [i]. Enfin, les voyelles /ɪ/ et /iː/ font partie des 5 voyelles de l'anglais qui portent la plus grande charge fonctionnelle, c'est-à-dire le rôle qu'elles jouent dans la langue pour distinguer des mots (Oh, Coupé, Marsico et Pellegrino, 2015; Oh, Pellegrino, Coupé et Marsico, 2013). Le contraste /ɪ - iː/ aurait donc une charge fonctionnelle plus élevée que le contraste /ʌ - ɑː/. De fait, le contraste /ɪ - iː/ présente une charge fonctionnelle double de celle du contraste /ʌ - ɑː/ (Pellegrino, communication personnelle). On peut imaginer que les différences de charges fonctionnelles amènent à ce que le niveau de distinction entre les timbres impliqués dans les oppositions diffère également, rendant moins « dramatique » le fait qu'elles puissent davantage être confondues acoustiquement lorsque leur rôle fonctionnel est plus limité. Ainsi, si l'on se place dans une optique exemplariste (Pierrehumbert, 2003 ; Pierrehumbert, 2001) et que l'on considère que les catégories phonémiques se forment selon un apprentissage probabiliste, il n'est pas difficile d'imaginer qu'une catégorie présentant un plus fort rendement (ou une fréquence plus élevée), par exemple le /ɪ/, soit formée avec une plus grande précision, qu'une autre catégorie moins fréquente, par exemple le /ʌ/. En d'autres termes, un apprentissage réalisé sur un plus grand nombre d'occurrences devrait sans doute permettre d'établir le centre de la catégorie avec plus de précision, de la même manière que l'estimation par intervalle de confiance d'une moyenne résulte d'un intervalle plus étroit lorsque celle-ci est réalisée sur un échantillon de plus grande taille.

Au sujet des apprenants francophones, nous avons observé que les performances des apprenants en prétest sont plus faibles que celles des locuteurs natifs pour toutes les voyelles dans la tâche d'identification et tous les contrastes dans la tâche de discrimination ce qui confirme les difficultés de perception des voyelles choisies pour notre étude par les apprenants francophones. Bien que l'ensemble des voyelles ne soit pas perçu de manière significativement différente entre les trois groupes, nous avons pu néanmoins noter des différences entre voyelles dans les groupes, notamment dans les groupes PE et PR : ils obtenaient de meilleurs taux d'identification pour la voyelle /ɪ/ que pour la voyelle /iː/. Bien que la différence ne soit pas

significative dans le groupe C, leurs scores vont globalement également dans ce sens. Ces résultats sont congruents avec ceux obtenus par Iverson et collègues (2012). En effet, dans leur étude, les participants présentaient également plus de difficulté à identifier le /iː/ que le /ɪ/, ce qui reflète le fait qu'ils ont plus tendance à juger le /iː/ comme /ɪ/ plutôt que l'inverse. Ce résultat n'avait pas été discuté par les auteurs bien qu'il soit plutôt surprenant. En effet, selon l'étude d'Iverson et Evans (2007) ces deux voyelles sont assimilées au /i/ français par les apprenants francophones donc il n'y a là pas de raison que l'une des deux conduise à plus d'erreurs que l'autre. Ce résultat semble bien illustrer les difficultés qu'ont les apprenants francophones à traiter les contrastes de durée (Boulenger et al., 2011 ; Iverson & Evans, 2007) puisqu'en plus du contraste de timbre, ces deux voyelles présentent un contraste de durée. Cependant, nous prenons ce résultat avec précaution puisque bien que nous ayons pris soin d'utiliser des mots fréquents en plus du symbole phonétique de la voyelle comme label pour les boutons de réponses (*pig* pour /ɪ/ et *feet* pour /iː/), la plus grande consistance orthographique de *pig* avec le français aurait pu conduire les apprenants à davantage utiliser ce bouton de réponse. En effet, il a été montré que l'orthographe de la L1 influençait à la fois les productions de L2 (e.g., Piske et al., 2002) mais aussi la perception (Bassetti, 2008 ; Bassetti et al., 2015). Iverson et collègues (2012) étaient dans la même situation puisqu'ils ont utilisé les labels *bit* et *beat*. De plus, sur nos données, l'analyse des résultats en discrimination séparément pour les items dont l'intrus était /iː/ ou /ɪ/ ne montre pas de différence dans les capacités à discriminer l'une ou l'autre des deux voyelles dans aucun des trois groupes d'apprenants. Il apparait donc ici que la tâche d'identification mérite d'être améliorée pour éviter ce biais supposé. Autre hypothèse explicative, il se pourrait que les participants français ayant appris à l'école que /iː/ était le *i long* et /ɪ/ le *i court* surestime le contraste de durée pour identifier/discriminer ces deux voyelles et ne s'appuient pas suffisamment sur le timbre des voyelles (Flege et al., 1997).

Pour les voyelles de la région du [a], nos données ne montrent aucune différence significative dans les scores entre voyelles en identification contrairement à ce qu'avait observé Iverson et collègues (2012). Dans leur étude, la voyelle /æ/ était mieux identifiée que la voyelle /ʌ/ suivie de la voyelle /ɑː/ par des locuteurs francophones. Ces différences entre voyelles semblaient en accord avec leurs résultats présentés dans Iverson et Evans (2007). Dans cette autre étude, les auteurs ont demandé aux participants d'indiquer pour chacune des voyelles de l'anglais laquelle des voyelles de leur langue maternelle présentées à l'écrit était la plus proche. Les participants devaient ensuite évaluer sur un continuum le degré de proximité de cette voyelle avec la voyelle de L1 choisie. Ils ont montré que ces trois voyelles (æ, ʌ et ɑː) sont

respectivement assimilées à trois voyelles françaises différentes, /a/, /ø/ et /ɑ/ par des locuteurs français avec des proximités évaluées décroissantes : /a/ est jugé comme plus proche de /æ/ que /ʌ/ ne l'est de /ø/ et que /ɑː/ de /ɑ/. Nous pensons que dans cette région de l'espace vocalique, les correspondances L1-L2 entre les voyelles de l'anglais et celles du français ne sont pas tout à fait telles que décrites par Iverson et Evans (2007). En effet, nous pensons tout d'abord que la voyelle /œ/ est un meilleur candidat d'assimilation de la voyelle /ʌ/. Le fait que le /ø/ ait été préférentiellement choisi comme plus proche exemplaire de L1 dans la tâche d'assimilation proposée par les auteurs peut être dû au fait que des pseudo-mots ont été utilisés par ces auteurs, notamment le pseudo-mots « beute » pour /ø/ et « boeute » pour la voyelle /œ/. La consistance grapho-phonémique du premier n'est pas très évidente et la *bizarrerie* orthographique du second a pu conduire les participants à ne pas choisir cette voyelle. On peut aussi noter que le /ɑ/ phonémique tend à ne plus être utilisé en français. En effet, Tranel (2003) indique que cette voyelle ne représente que 2.4% des occurrences de /ɑ-a/ dans la langue française alors que le /a/ est utilisé dans 97.6% des cas et que l'opposition /ɑ-a/, comme dans l'opposition pâte/patte a un très faible rendement ce qui devrait conduire à une disparition de cette opposition. La base de données LAPSyD qui répertorie les inventaires phonologiques des langues n'inclut d'ailleurs pas cette voyelle dans le système phonologique du français (Fougeron & Smith, 1993 ; Maddieson et al., 2014). Si l'on considère les résultats obtenus en discrimination, on observe que les participants des groupes PE et PR sont significativement moins bons pour le contraste /ʌ - ɑː/ comparativement au contraste /æ - ʌ/. Ici, on peut de nouveau voir un signe de la difficulté des français à utiliser le contraste de durée pour traiter les voyelles. En effet, le fait que les stimuli utilisés présentent plus de chevauchement entre espaces vocaliques pour le contraste /ʌ - ɑː/ comparativement au contraste /æ - ʌ/ a pu les conduire à produire plus d'erreurs s'ils ne tiennent pas compte de la durée des voyelles.

Notre seconde hypothèse supposait que les scores des locuteurs natifs seraient toujours supérieurs à ceux des apprenants en posttest mais que les apprenants présenteraient des différences entre groupes suite aux entraînements suivis. Nous nous attendions à ce que le groupe PR présente les scores les plus élevés, suivi du groupe PE puis du groupe C. Comme attendu, les apprenants n'ont pas atteint les performances des natifs après entraînement, sauf pour la voyelle /æ/ en identification et le contraste /æ - ʌ/ en discrimination pour lesquels les deux groupes PE et PR ne présentent plus des performances significativement plus faibles que les natifs. Ainsi, il semblerait que les deux entraînements aient été bénéfiques pour améliorer les capacités de perception de la voyelle /æ/. Notre hypothèse H2 n'est cependant pas

totalement validée puisque la comparaison des performances entre groupes en posttest ne permet pas de distinguer les groupes entre eux.

En troisième hypothèse, nous pensions que seuls les groupes PE et PR amélioreraient leurs performances en perception et que l'entraînement proposé au groupe PR aurait été le plus efficace. La comparaison des performances entre prétest et posttest montre en effet que seuls les groupes ayant suivi des entraînements améliorent leurs performances entre prétest et posttest de manière significative, cependant il semblerait que l'entraînement reçu par le groupe PR n'ait pas été plus efficace que l'entraînement *HVPT* reçu par le groupe PE. Les analyses statistiques n'ont ensuite pas permis de déterminer si les entraînements avaient plus bénéficié à certaines voyelles qu'à d'autres, en d'autres termes, les participants des groupes ayant reçu des entraînements ont amélioré leurs performances pour toutes les voyelles concernées dans notre étude.

Notre étude est la deuxième à avoir proposé des entraînements à des apprenants francophones pour les voyelles de l'anglais. Contrairement à l'étude d'Iverson et collègues (2012), dans laquelle les participants avaient été entrainés sur l'ensemble des monophtongues de l'anglais (14 voyelles au total) pendant 10 séances, nous avons focalisé les entraînements proposés sur cinq voyelles de l'anglais pendant cinq séances d'entraînement. Les résultats d'Iverson et collègues (2012) ont montré une amélioration des performances en identification pour l'ensemble des voyelles sauf pour les voyelles /ʌ/ et /ɪ/ que nos deux entraînements ont permis d'améliorer. Les auteurs n'avaient pas considéré les paires de voyelles de notre étude en discrimination mais l'entraînement qu'ils avaient proposé n'avaient été bénéfiques qu'à trois contrastes parmi dix-sept en discrimination. Les deux types d'entraînements que nous avons proposés ont été efficaces pour l'ensemble des voyelles et des contrastes chez nos deux groupes ayant reçu un entraînement comparativement au groupe C n'ayant pas reçu d'entraînement. Il est possible qu'un entraînement plus focalisé (sur seulement cinq voyelles), ou tout du moins comprenant des séances spécifiquement focalisées sur un contraste seulement, ait aidé par rapport à l'entraînement plus « dispersé » d'Iverson et collègues (2012). En termes de nombre de contrastes entrainés et durée des entraînements, nos entraînements sont similaires à ceux proposés dans les études portant sur l'acquisition de contrastes vocaliques de l'anglais par des apprenants hispanophones ce qui pousse à pencher en faveur de ce type d'entraînement pour améliorer les capacités perceptives de contrastes vocaliques de L2.

Bien que la méthodologie suivie ait été adaptée de la plupart des études ayant évalué l'effet d'entraînement pour l'acquisition de voyelles de l'anglais, nous notons qu'un certain nombre d'améliorations pourraient être apportées au protocole. Tout d'abord, bien que nous ayons pris le plus de précautions possibles dans la constitution des stimuli expérimentaux, il s'est avéré que certains stimuli des tests n'étaient pas optimaux puisqu'ils n'étaient pas tous correctement identifiés par des locuteurs natifs. Il aurait été profitable de tester les stimuli expérimentaux dans une tâche de perception destinée à des locuteurs natifs avant de les soumettre aux apprenants afin de nous assurer qu'aucun d'entre eux ne présentait d'ambiguïté trop importante. Nous avons également remarqué comme dans l'étude d'Iverson et collègues (2012) qu'un éventuel biais de consistance orthographique sur les labels des boutons de réponse existait. Il serait souhaitable de lever ce doute en proposant une tâche supplémentaire aux participants pour s'assurer qu'ils connaissent bien les symboles phonétiques et ne produisent pas des biais de réponse en faveur des items qui seraient plus consistants dans leur langue maternelle.

4 Évaluation en production

4.1 Objectifs et hypothèses

Les performances en production ont été évaluées au moyen d'une tâche de lecture de mots isolés. Nous avions deux objectifs :

1. Caractériser la manière dont les apprenants francophones produisent avant entraînement les voyelles de l'anglais de notre étude comparativement aux locuteurs natifs ;
2. Évaluer l'effet des entraînements que nous avons proposés sur les voyelles produites par les apprenants.

Pour le premier objectif, nous avons mené plusieurs analyses : 1) nous avons comparé les paramètres acoustiques (dispersion dans l'espace F1/F2 et durée) des voyelles anglaises produites par des locuteurs anglais natifs (L1-anglais), à ceux des voyelles françaises produites par des locuteurs français natifs (L1-français) afin de comparer les systèmes des deux langues. Puis nous avons comparé ces mêmes paramètres acoustiques entre les voyelles anglaises produites par les apprenants français (L2-anglais) et les voyelles françaises produites par ces

mêmes locuteurs (L1-français) afin d'observer comment les deux systèmes « cohabitent ». Enfin, nous avons comparé les paramètres acoustiques des voyelles anglaises produites par des locuteurs anglais natifs (L1-anglais) à ces mêmes voyelles produites par les apprenants francophones (L2-anglais) afin de déterminer les différences acoustiques qui caractérisent les productions des apprenants en L2 par rapport aux voyelles qu'ils tentent de produire. Pour le /i:/, qui est parfois diphtongué dans la variété d'anglais qui nous intéresse (Ferragne, 2008), nous avons cherché à savoir si l'analyse de paramètres décrivant la trajectoire des formants pouvait être pertinente pour caractériser cette voyelle selon qu'elle est produite par des locuteurs natifs ou des locuteurs de L2. Enfin, afin de déterminer les voyelles qui sont produites de manière particulièrement différente chez les apprenants, nous avons soumis les voyelles des apprenants à une classification automatique reposant sur un modèle appris sur les caractéristiques acoustiques des voyelles produites par des locuteurs anglais natifs.

Afin de déterminer l'effet des entraînements, qui était notre deuxième objectif, nous avons utilisé deux méthodologies, une méthodologie que nous avons appelée « objective » basée sur l'analyse des paramètres acoustiques des voyelles et une méthodologie que nous qualifions de plus « *subjective* » et qui a consisté à demander à des locuteurs de langue maternelle anglaise d'évaluer les productions des apprenants en prétest et en posttest, ainsi que celles des locuteurs du groupe natif (groupe N). En ce qui concerne la méthodologie « *objective* », nous avons observé les changements qui se sont opérés sur les paramètres acoustiques des voyelles (valeurs formantiques et durées) et nous avons également soumis les voyelles produites avant et après entraînement à une classification automatique dont le modèle a été appris à partir des paramètres acoustiques des voyelles enregistrées par les locuteurs anglais qui ont produit le matériel expérimental. Nous pensions que si les apprenants produisent en posttest des voyelles qui sont plus proches des voyelles natives du point de vue acoustique, celles-ci devraient être mieux classifiées après entraînement à l'aide d'un modèle de classification appris sur des données enregistrées chez des locuteurs natifs.

Pour les deux méthodes utilisées, nous nous attendons à des résultats suivant un même schéma et avons donc postulé les hypothèses suivantes :

H1 : Avant entraînement, nous nous attendions à ce que les voyelles produites par les apprenants soient moins bien classifiées par la classification et moins bien identifiées par des juges natifs que celles produites par les locuteurs du groupe natif (Groupe N). Nous n'attendions pas de différence entre les groupes C, PE et PR avant entraînement.

H2 : Après entraînement, il nous semble raisonnable de penser que les voyelles des apprenants seraient toujours moins bien classifiées/identifiées que celles des locuteurs natifs. Nous nous attendons à ce que les groupes PE et PR bénéficient des entraînements contrairement au groupe C et nous nous attendons à ce que le groupe PR, ayant suivi l'entraînement en production soit celui dont les scores soient les meilleurs.

H3 : Les comparaisons des scores de classification/identification des voyelles produites devraient montrer une plus grande amélioration dans le groupe PR, que dans le groupe PE suivi du groupe C.

4.2 Méthodologie

4.2.1 Description des tâches utilisées

Deux tâches de production, en français et en anglais, ont été programmées avec le logiciel RocMe!. Lors de la tâche de production en français, les participants produisaient des mots français monosyllabiques de type bV(r) contenant les voyelles /a/ (*bas*), /ɛ/ (*baie*), /œ/ (*beurre*), /i/ (*bi*), et /e/ (*bée*). Les mots à produire étaient présentés visuellement à l'écran, un par un et en deux phases : les cinq mots à produire étaient d'abord présentés une fois chacun comme items d'essai pour se familiariser avec la tâche puis ils étaient présentés trois fois chacun dans un ordre aléatoire. Seuls les participants des groupes PE, PR et C réalisaient cette tâche et seulement une fois avant entraînement. Les participants du groupe N ne passaient pas cette tâche. La tâche de production en anglais commençait par cinq items d'essai composés de mots CVC fréquents contenant les cinq voyelles à produire (*cat, cut, card, feet* et *pig*). Ensuite, les participants enregistraient, dans un ordre aléatoire, trois occurrences de chaque mot anglais de type bVd contenant les voyelles /æ/ (*bad*), /ʌ/ (*bud*), /ɑː/ (*bard*), /iː/ (*bead*), et /ɪ/ (*bid*). Afin de minimiser les erreurs de type orthographique, les mots à produire étaient présentés à l'écran accompagnés chacun du mot fréquent qui avait été présenté lors des items d'essai et qui contenait la voyelle en question. Ce mot fréquent était accompagné du symbole phonétique de la voyelle à produire (voir Figure 30). Les participants des groupes d'apprenants PE, PR et C

passaient cette tâche avant et après entraînement alors que le groupe de natifs ne réalisait cette tâche qu'une seule fois.

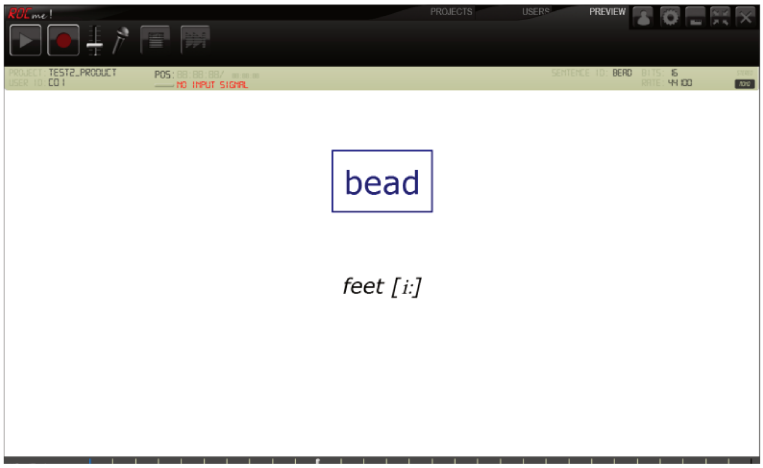

Figure 30 : Écran présenté lors de la tâche de production en anglais. Dans cet exemple, le participant doit prononcer le mot « *bead* » ; il lui est donné comme indice que ce mot contient la voyelle /i:/ qui est la même que dans le mot fréquent « *feet* »

4.2.2 Évaluation *objective* des paramètres acoustiques

Les enregistrements ont été analysés avec le logiciel Praat (Boersma et Weenink, 2018). Les voyelles ont été segmentées manuellement puis les deux premiers formants ont été mesurés semi-automatiquement sur toute la durée de la voyelle : l'estimation formantique de Praat superposée au spectrogramme était ajustée jusqu'à ce que l'estimation soit cohérente avec le spectrogramme. Une valeur de formant a été prise toutes les 0.05 secondes. Cette étape a été réalisée à l'aide d'un programme écrit par E. Ferragne, `cp_formants.praat`[8]. Les valeurs à 50% de la durée de la voyelle des trois premiers formants ont ensuite été extraites, transformées en Bark puis centrées-réduites par locuteur indépendamment pour chaque formant. Les durées des voyelles ont également été calculées.

Pour la voyelle /i:/ qui s'est avérée être parfois diphtonguée chez certains locuteurs de l'anglais du Sud-est de l'Angleterre selon Ferragne (2008), nous avons étudié la dynamique des formants. Pour cela, nous avons suivi la procédure décrite dans Williams & Escudero, (2014) : les 80% centraux de chaque courbe de F2 ont été rééchantillonnés afin d'avoir 30 points et les

[8] Le script est disponible à l'adresse suivante : https://github.com/emmanuelferragne/CminR-Praatik

coefficients 2 et 3 (C1 et C2) d'une transformée en cosinus discrète ont été calculés. Ces coefficients permettent de caractériser les trajectoires de formants. Le deuxième coefficient, C1, correspond aux changements dans les fréquences de résonances tout au long de la voyelle reflétant ainsi la magnitude et la direction (le signe) du changement. Le troisième coefficient, C2 de la DCT représente des aspects plus complexes de la forme de la trajectoire de formant.

4.2.3 Évaluation *subjective* des productions

Six locuteurs natifs originaires du Sud-Est de l'Angleterre ont jugé les productions en pré- et post-tests des apprenants ainsi que celles du groupe natif. Ces locuteurs réalisaient une sorte de tâche d'identification à 5 choix forcés dans laquelle il devait indiquer quel mot parmi les cinq mots *bad*, *bud*, *bard*, *bid*, et *bead* était prononcé selon eux. La tâche était programmée avec l'interface *Demo window* de Praat comme le tâche d'identification des phases tests aux différences que les mots précédemment cités constituaient les labels des boutons et qu'un bouton supplémentaire indiquant « *none of them* » était proposé. La consigne présentée dans la Figure 31 était donnée aux juges.

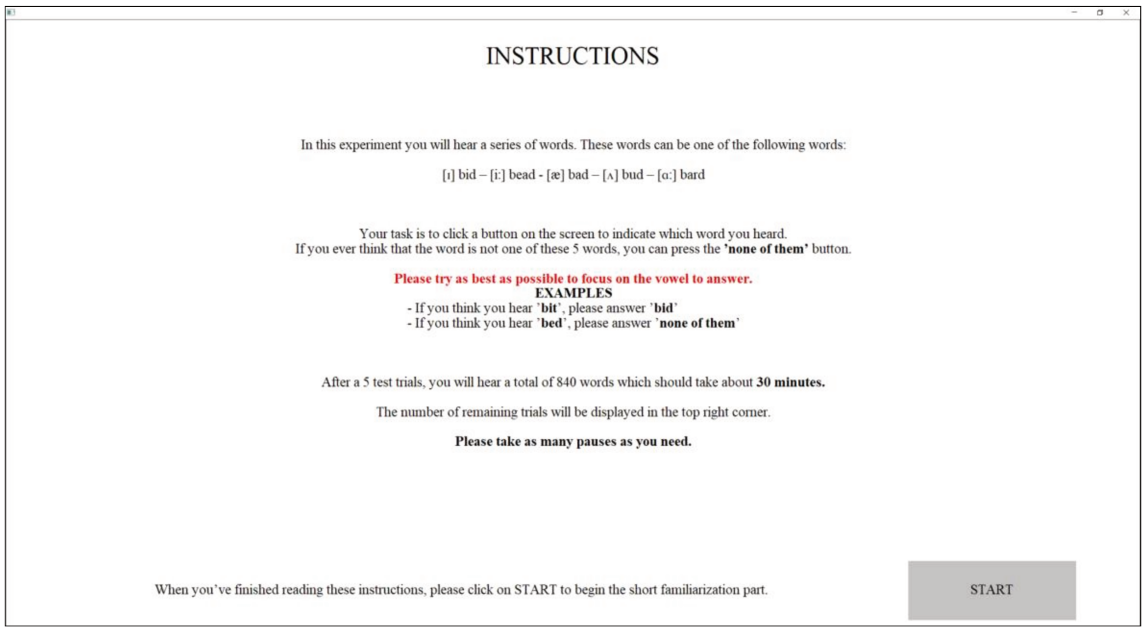

Figure 31 : Écran d'instructions présenté aux locuteurs natifs ayant évalué les productions des groupes expérimentaux (N, C PE et PR)

Chaque juge évaluait la totalité des productions de la moitié des participants de telle sorte que les productions de chaque participant étaient évaluées par trois juges différents et que les juges disposaient tous d'une liste de mots différente à évaluer (voir Tableau XXI). Au total chaque juge évaluait 840 mots composés des productions en prétest et posttest de 8 participants par

groupe d'apprenants et de 8 participants natifs. L'ordre de présentation des stimuli était *pseudo-randomisé* de telle sorte que deux productions d'un même participant ne soient pas présentées consécutivement.

Tableau XXI : Productions évaluées par chacun des six juges natifs

N° du groupe de juges	Numéro des participants dans les groupes C, PE, PR et N	Numéros des juges évaluant les productions de ces participants
1	1-4	1-2-3
2	5-8	1-4-5
3	9-12	2-4-6
4	13-16	3-5-6

Les juges ont été recrutés à Reading (UK) avec l'aide de Marie Martel et étaient rémunérés £10 pour leur participation d'environ 30 minutes.

4.3 Résultats

4.3.1 Production des voyelles par les apprenants francophones avant entraînement : comparaison avec les productions natives

Les voyelles produites par les apprenants francophones avant entraînement et tous groupes confondus ont d'abord été comparées à celles de l'ensemble des locuteurs de langue maternelle anglaise que nous avons enregistrés, i) pour réaliser les stimuli expérimentaux plus ii) ceux qui constituaient le groupe Natif (groupe N). Nous souhaitions déterminer les différences acoustiques qui caractérisent les voyelles produites par les apprenants francophones avant qu'ils ne soient soumis aux entraînements et ainsi déterminer comment les voyelles de notre étude posent problème aux apprenants. En termes de propriétés acoustiques, nous nous sommes intéressés aux valeurs des deux premiers formants, à la durée des voyelles, ainsi qu'aux coefficients C1 et C2 de la transformée en cosinus discrète du second formant pour la voyelle /iː/ puisque celle-ci peut être diphtonguée en anglais.

4.3.1.1 Résumé du corpus utilisé

Le jeu de données utilisé pour déterminer les difficultés des apprenants francophones en production des voyelles /æ/, /ʌ/, /ɑː/, /ɪ/ et /iː/ était constitué :

1. Des productions en anglais de mots bVd (3 occurrences par mot) de l'ensemble des locuteurs ayant servi à créer le matériel expérimental (n = 18 ; 9 femmes)[9] et du groupe expérimental de locuteurs natifs (n = 16 ; 11 femmes), soit un total de 102 occurrences par voyelle. Ces productions seront représentées en bleu dans les graphiques qui suivent.
2. Des productions en anglais de mots bVd (3 occurrences par mot) des locuteurs francophones (les trois groupes expérimentaux PE, PR et C confondus ; n = 48 ; 12 femmes) en prétest uniquement. Ces productions seront représentées en orange dans les graphiques.
3. Des productions en français de mot bV(R) (3 occurrences par mot) des locuteurs francophones (Groupes PE, PR et C confondus). Ces productions sont représentées en vert dans les graphiques.

Ce premier sous-ensemble contient au total 1940 voyelles puisque 5 occurrences de voyelles, indiquées en gras dans le Tableau XXII, n'ont pas pu être traitées car les formants n'étaient pas visibles dans les spectrogrammes. De plus cinq autres voyelles ont été retirées des analyses dans les voyelles anglaises produites par les apprenants français car ils ont prononcé [bɛd] le mot *bead*.

Tableau XXII : Répartition des voyelles dans le premier sous-ensemble en fonction de la langue maternelle des locuteurs.

	Voyelles anglaises					Voyelles françaises				
	æ	ʌ	ɑː	ɪ	iː	a	œ	ɛ	e	i
Locuteurs français	144	144	144	144	139	**143**	144	**143**	144	**142**
Locuteurs anglais	**101**	102	102	102	102	-	-	-		-

[9] Les productions de mots bVd des locuteurs ayant servi à constituer le matériel expérimental ont été utilisées ici afin d'avoir un échantillon plus important de données. Il faut noter que la tâche réalisée par ces locuteurs est un peu différente puisqu'ils ne disposaient ni du symbole phonétique de la voyelle, ni du mot indice leur indiquant la voyelle à produire (voir Figure 15, p 62).

4.3.1.1 Analyse statique des formants vocaliques

Les analyses du timbre des voyelles ont été réalisées séparément pour les hommes et pour les femmes. Dans les représentations graphiques, les graphiques de gauche sont ceux issus des données des femmes, ceux de droite ont été réalisés à partir de celles des hommes. Pour chaque graphique, les symboles phonétiques indiquent le barycentre de la catégorie, et les ellipses représentent l'intervalle de confiance au seuil de 95% des valeurs moyennes des formants pour le groupe de participants concerné. Des proportions de chevauchement entre les catégories voisines ont été calculées en faisant le rapport entre l'aire de l'intersection par l'aire de l'union des deux catégories considérées, i.e., la surface totale occupée par les deux catégories (Aire Catégorie A + Aire Catégorie B – Aire A∩B). Ces proportions ont été calculées à l'aide du package `Siber` du logiciel R. Nous comparerons d'abord les voyelles anglaises et françaises produites respectivement par des locuteurs anglais et français natifs. Puis nous comparerons les voyelles anglaises aux voyelles françaises produites par des locuteurs français. Enfin, nous comparerons les voyelles anglaises produites les apprenants à celles des locuteurs natifs.

4.3.1.1.1 Comparaison des voyelles anglaises et françaises produites par des locuteurs natifs

La Figure 32 présente les ellipses de dispersion des voyelles anglaises et française enregistrées par les locuteurs dans leurs langues maternelle, c'est-à-dire en français et en anglais.

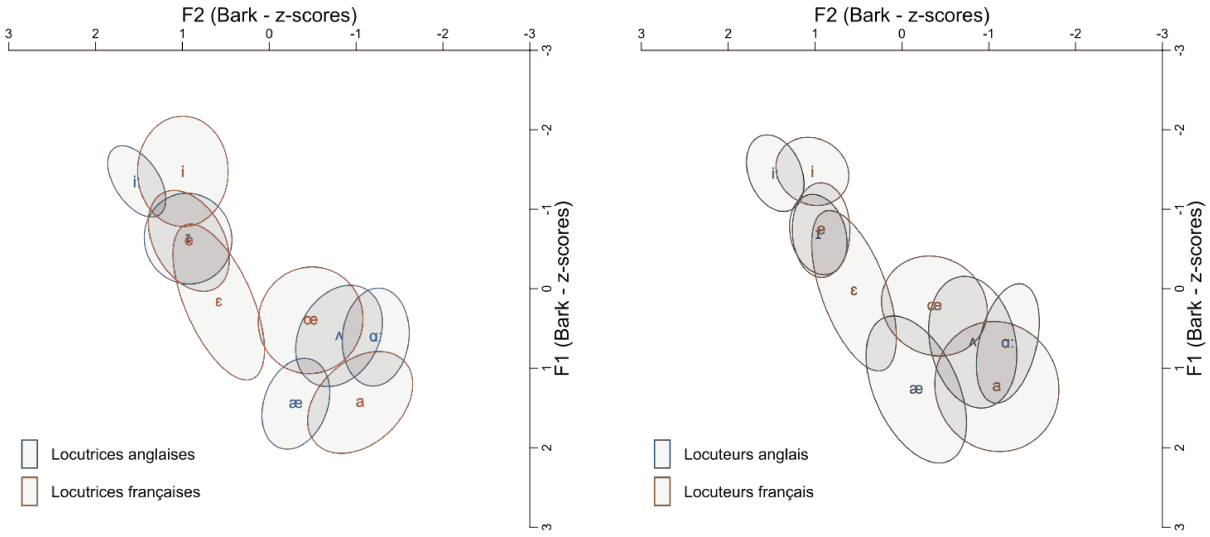

Figure 32 : Voyelles de l'anglais (en bleu) et du français (en orange) produites par des natifs, pour les femmes à gauche, pour les hommes à droite

La comparaison des voyelles anglaises produites par des locuteurs natifs anglais et des voyelles françaises produites par des locuteurs natifs français dans l'espace des deux premiers formants montrent que, pour les voyelles étudiées, la variabilité inter-catégorielle est similaire dans les deux langues pour les hommes et les femmes. On constate en effet, pour les femmes comme pour les hommes que les aires des ellipses sont équivalentes en français et en anglais puisqu'une analyse de la variance sur les aires des ellipses de chaque catégorie vocalique ne montre pas d'effet de la Langue ($F_{(1,8)} = 2.52$, *NS*), du Sexe ($F_{(1,8)} = 0.02$, *NS*), ni d'interaction Langue × Sexe ($F_{(1,8)} = 0.92$, *NS*).

Dans la région acoustique du [i], on observe en français qu'il y a peu de chevauchement entre les catégories /i/ et /e/, soit 12% de chevauchement entre ces deux catégories chez les femmes et 7% chez les hommes. On observe cependant, que la voyelle /ɛ/ occupe une partie plus importante de l'espace acoustique du /e/, chez les femmes (24%) comme chez les hommes (17%). L'écoute des enregistrements pour ce contraste révèle que 9 participants français ont tendance à prononcer le mot baie avec la voyelle /e/ comme c'est le cas en position finale chez des locuteurs présentant un accent méridional (Dufour, Nguyen et Frauenfelder, 2007) et, étonnamment, que deux d'entre eux prononcent parfois le mot bée avec la voyelle /ɛ/. La Figure 33 montre les dispersions des voyelles /e/ et /ɛ/ chez les locuteurs produisant le contraste /e-ɛ/ comparativement à ceux ne le produisant pas.

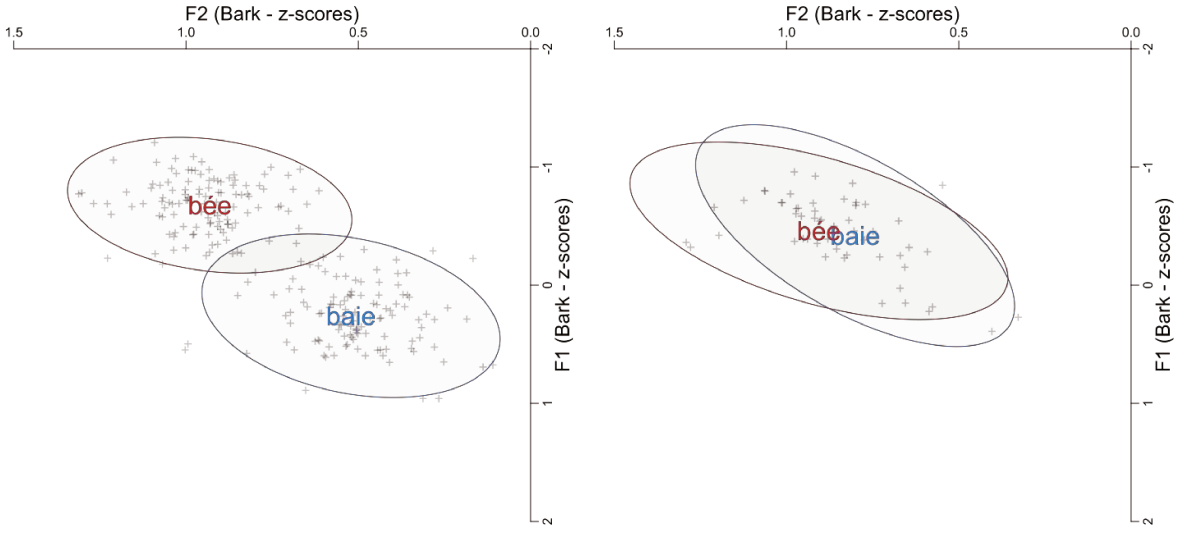

Figure 33 : Représentation dans l'espace F1/F2 en Bark centrés réduits des voyelles produites dans les mots bée (en rouge) et baie (en bleu). Les mots représentent les barycentres des voyelles. À gauche sont représentées les productions des 39 participants prononçant le mot baie avec la voyelle /ɛ/. À droite sont représentées les productions des neuf participants repérés à l'écoute des enregistrements comme prononçant le mot baie avec la voyelle /e/

Si l'on considère les voyelles anglaises, on observe peu de chevauchement entre les deux catégories /iː/ et /ɪ/, tant pour les femmes (1%) que pour les hommes où la part partagée de l'espace acoustique entre les deux catégories est négligeable. Ceci indique que bien que portant sur un contraste de durée également, ces deux voyelles sont réalisées de manière très distincte du point du vue du timbre. De manière générale, dans les deux langues, pour cette région de l'espace vocalique, on constate, mis à part le cas particulier du [ɛ] soumis à des variations régionales, que les voyelles sont produites de manières distinctes du point de vue du timbre avec peu de partage de l'espace acoustique.

Si l'on observe les partages d'espaces acoustiques entre les deux langues, on peut noter que la voyelle française /i/ partage peu d'espace avec les deux voyelles anglaises /iː/ (11% chez les femmes, 16% chez les hommes) et /ɪ/ (11% chez les femmes et 2% chez les hommes). Le /iː/ anglais est réalisé du point du vue du timbre de manière plus antérieure que le /i/ français et le /ɪ/ anglais est quant à lui plus ouvert que le /i/ français. En revanche, les voyelles anglaises /ɪ/ et françaises /e/ partagent une grande partie de leur espace acoustique que ce soit pour les femmes (77%) ou pour les hommes (78%). Ces deux voyelles ne présentant pas théoriquement par ailleurs de différences de durée, on peut imaginer que les apprenants francophones utilisent le /e/ français comme cible articulatoire pour la réalisation du /ɪ/ en anglais. On peut alors supposer que la production du /iː/ pose davantage de problèmes aux apprenants puisqu'il est possible que ces derniers utilisent le /i/ français comme cible articulatoire alors qu'une réalisation adéquate nécessiterait une hyper-articulation de celui-ci.

Dans la région du [a], on constate que les deux voyelles françaises considérées, /a/ et /œ/, sont réalisées de manière tout à fait distincte avec des taux de chevauchement négligeables chez les femmes et inférieurs à 2% chez les hommes. Pour les voyelles présentes dans cette région en anglais, on observe deux phénomènes : d'une part les voyelles /æ/ et /ʌ/, qui théoriquement ne sont contrastées que par le timbre, ne partagent quasiment pas d'espace acoustique (moins de 5%) ; d'autre part, les voyelles /ʌ/ et /ɑː/, qui en plus d'un contraste de timbre présentent un contraste de durée, ont environ un quart de leur espace acoustique en commun que ce soit chez les hommes et chez les femmes. On peut donc penser que pour le contraste /ʌ - ɑː/, la réalisation du contraste de durée constitue un indice important.

Dans cette région vocalique, les catégories anglaises et françaises partagent leurs espaces de manière plus complexe que ce n'est le cas dans la région du [i]. En effet, aucune catégorie du français ne peut très clairement servir de support pour la réalisation de voyelle de l'anglais.

Tout d'abord, on peut observer que le /æ/ anglais est réalisé de manière plus antérieure que le /a/ français. Tout comme le /iː/ qui vient se situer dans une région extrême du triangle vocalique non occupée par les voyelles françaises, le /æ/ ne partage pratiquement pas d'espace acoustique avec les voyelles françaises de cette région (3% pour les femmes, 5% chez les hommes avec la voyelle /œ/, 5 et 6% avec la voyelle /a/ pour les femmes et les hommes respectivement). Pour cette voyelle (le /æ/), on peut s'attendre à deux scenarios d'après le modèle SLM de Flege : soit celle-ci est assimilée au /a/ français et, en production, on peut faire l'hypothèse qu'elle sera réalisée avec des paramètres acoustiques similaires au /a/ français, soit elle est perçue comme bien différente et dans ce cas, on se situe plutôt dans un cas de dissimilation. Dans cette situation, on peut faire l'hypothèse que les productions seront réalisées de manière très différente au /a/ français. Cependant, il se peut qu'une très grande variabilité soit observée dans les productions de cette voyelle du fait de l'absence de cible articulatoire préexistante en français. Si l'on considère la voyelle /ʌ/, celle-ci se situe à mi-chemin entre les /a/ et /œ/ français et l'on observe des différences entre les femmes et les hommes. Chez les femmes, cette voyelle partage une forte proportion de son espace acoustique avec le /œ/ français (38%) et quasiment pas d'espace avec le /a/ français (5%) alors que chez les hommes, cette même voyelle partage un espace acoustique à parts presque égales avec les voyelles /a/ (31%) et /œ/ (25%) du français. La voyelle /ɑː/ est quant à elle réalisée de manière assez distincte des voyelles françaises chez les femmes (11% de chevauchement avec /a/ et 7% avec /œ/) alors que chez les hommes elle partage un espace acoustique conséquent avec la voyelle /a/ (28%). En termes de prédictions, on peut s'attendre à ce que les réalisations du contraste /ʌ - ɑː/ repose d'une part davantage sur la durée que sur le timbre et d'autre part qu'en termes de timbre, ce contraste soit davantage effectué sur le plan de l'ouverture, en suivant l'opposition française /œ - a/ que sur le plan antéro-postérieur comme c'est le cas en anglais pour ce contraste.

4.3.1.1.2 Comparaison des voyelles de l'anglais et du français produites par des locuteurs français

La dispersion des voyelles anglaises et françaises produites par les locuteurs français sont présentées dans la Figure 34.

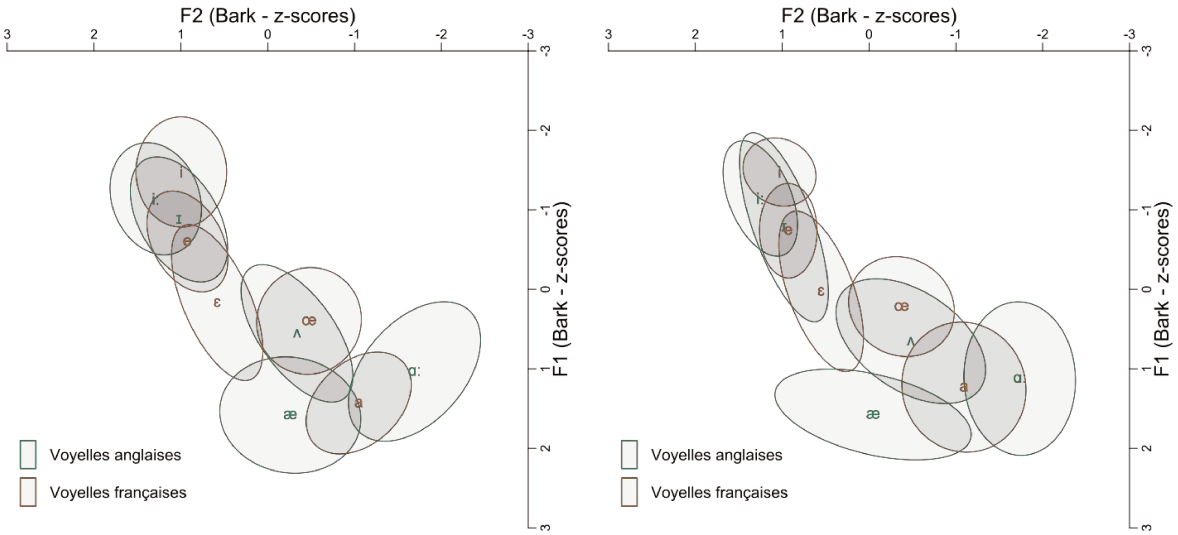

Figure 34 : Voyelles de l'anglais (en vert) et du français (en orange) produites par des apprenants français, pour les femmes à gauche, pour les hommes à droite

L'observation de la dispersion des ellipses des voyelles selon que les voyelles sont des voyelles de L1 (en orange) ou L2 (en vert) semble indiquer une plus grande variabilité en L2 qu'en L1. Cependant, cet effet n'est pas significatif à l'analyse de la variance sur les aires des ellipses ($F_{(1, 8)} = 3.78$, $p = .09$). L'ANOVA ne montre pas non plus d'effet ni du Sexe des participants ($F_{(1, 8)} = 0.26$, *NS*), ni d'interaction Langue × Sexe ($F_{(1, 8)} = 0.19$, *NS*).

Pour les voyelles de la région du [i], la comparaison des espaces vocaliques des voyelles anglaises et françaises produites par les locuteurs francophones montre que les voyelles anglaises sont produites avec plus de confusion que celles du français. En effet, les catégories /ɪ/ et /iː/ présentent un taux de chevauchement plus important (49 % pour les femmes et 40 % pour les hommes) que les voyelles du français /i/ et /e/ (12% pour les femmes et 7% pour les hommes). L'absence de contraste de durée pour les deux voyelles du français impose que celles-ci soient produites avec des timbres distincts. La forte confusion entre les deux [i] de l'anglais peut être provoquée par deux facteurs : le premier est la très grande variabilité des productions pour ces voyelles, en particulier pour la voyelle /ɪ/ chez les hommes, le deuxième est la grande proximité des voyelles produites du point de vue du timbre des voyelles. Il est en effet possible, que les locuteurs français s'appuient davantage sur la durée de la voyelle plutôt que sur le timbre pour produire ces deux voyelles et donc ne réalisent pas des cibles très distinctes dans l'espace des deux premiers formants. Concernant la voyelle /ɪ/, on constate, bien qu'elle soit produite par les apprenants avec une très grande variabilité, que le barycentre de cette catégorie est assez proche pour les deux sexes de celui de la /e/ qui avait été repéré

comme un potentiel bon support articulatoire pour la réalisation de cette voyelle. En revanche, en ce qui concerne le /iː/, on observe qu'elle présente, certes une forte zone de chevauchement avec le /i/ français (34% pour les femmes, 23% pour les hommes), mais qu'elle est réalisée dans un espace intermédiaire au /i/ et au /e/ français. En effet, celui-ci est plus postérieur et plus ouvert que le /i/ français et tend à se rapprocher de la cible /iː/ native.

Dans la région du [a], on observe que l'ensemble des voyelles, françaises et anglaises, sont produites avec une plus grande variabilité que celles de la région du [i], $F(1, 8) = 12.20$, $p = .008$ et ce pour les deux sexes puisque l'on n'observe ni d'effet du sexe $F(1, 8) = 0.38$, *N.S*, ni d'interaction Région × Sexe, $F(1, 8) = 4.38$, $p > .05$. Cette plus grande dispersion dans les réalisations peut être due au plus grand espace disponible dans cette région pour produire des voyelles distinctes. On observe en effet que les barycentres des catégories voisines sont espacés par de plus grandes distances acoustiques dans l'espace des deux premiers formants (en rouge dans le Tableau XXIII) que celles de la région du [i] (en vert dans le Tableau XXIII) que ce soit en français ou en anglais. De plus, on constate que les catégories des voyelles de l'anglais sont produites dans des espaces distincts avec très peu de chevauchement entre elles.

Tableau XXIII : Distances euclidiennes moyennes en Hertz entre les voyelles d'une même région acoustique dans l'espace F1/F2 produites par les locuteurs français pour les deux langues, pour les femmes et les hommes

Voyelles anglaises (L2)	æ		ʌ		ɪ	
	F	M	F	M	F	M
ʌ	218	216				
ɑː	545	560	503	384		
iː					152	139

Voyelles françaises (L1)	a		i	
	F	M	F	M
œ	301	286		
e			144	105

Les apprenants français parviennent à placer la voyelle /æ/ dans une région de l'espace acoustique non occupée en français. En effet, celle-ci est plus antérieure que la /a/ français et ne partage avec lui qu'un espace restreint (18% pour les femmes et 10% pour les hommes). En ce qui concerne la voyelle /ɑː/, celle-ci est produite par les apprenants dans une région plus postérieure que le /a/ français n'ayant en commun qu'environ 25% de l'espace acoustique des deux premiers formants avec la voyelle /a/ française. Enfin, la voyelle /ʌ/ est produite par les femmes avec un fort chevauchement avec la voyelle /œ/ (51%) et un faible chevauchement avec /a/ (6%). Alors que chez les hommes celle-ci est produite dans un espace qui recouvre fortement les catégories /a/ (24%) et /œ/ (39%).

4.3.1.1.3 Comparaison des voyelles anglaises produites par des locuteurs natifs et des locuteurs francophones

La dispersion des voyelles anglaises produites par les locuteurs anglais natifs et les locuteurs français sont présentées dans la Figure 35.

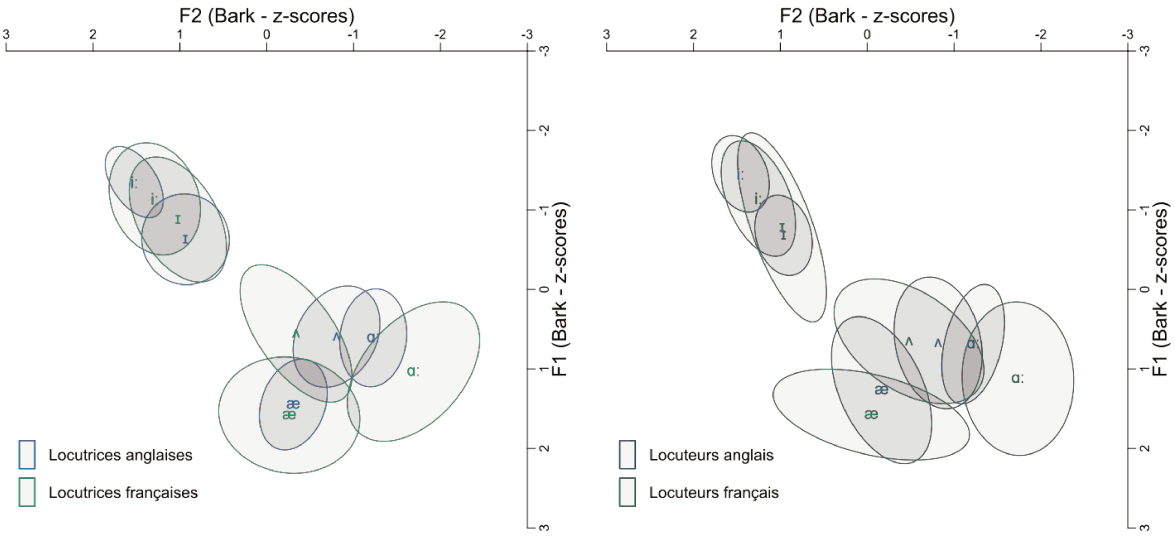

Figure 35 : Voyelles de l'anglais produites par des locuteurs natifs (en bleu) et par les apprenants français (en vert), pour les femmes à gauche, pour les hommes à droite

Une analyse de la variance réalisée sur les aires des ellipses indique ici encore que les productions des français en L2 (en vert) présentent une plus forte variabilité que celles des anglais en L1 (en bleu) toutes les voyelles considérées, $F_{(1, 12)} = 57.84$, $p < .0001$. De nouveau, on n'observe ni d'effet du Sexe, ni d'interaction Sexe × Langue. Si l'on compare la variabilité des productions selon les régions ([a] vs. [i]) et la langue, on observe une interaction Région × Langue significative, $F_{(1, 13)} = 4.82$, $p < .04$. Des comparaisons par paires montrent que les voyelles de la région du [a] sont produites avec une plus grande variabilité que celles du [i] chez les apprenants ce qui n'est pas le cas chez les locuteurs natifs.

Dans la région du [i], on observe que les voyelles produites par les natifs sont plus distinctes que celles produites par les apprenants (proche de 0% de chevauchement chez les natifs versus 49% pour les femmes, 40% pour les hommes chez les apprenants). De plus, on observe une plus grande distance acoustique entre les centres des catégories des natifs (voir Tableau XXIV à gauche en vert) qu'entre ceux des apprenants (voir Tableau XXIV à droite en vert).

Dans la région du [a], on remarque le phénomène inverse : les voyelles des apprenants sont plus distinctes que celles des locuteurs natifs. On observe un chevauchement moins important entre les trois catégories produites par les locuteurs francophones que par les locuteurs natifs, de plus, les distances acoustiques dans F1/F2 entre les centres des catégories des francophones sont plus grandes que celles des locuteurs natifs. Les voyelles /ʌ/ et /ɑː/ sont en effet produites avec un pourcentage de partage d'espace acoustique de 24 et 25% pour les femmes et les hommes natifs alors que ce pourcentage de chevauchement est de 0% et 2% chez les apprenants français. La comparaison des distances acoustiques entre les barycentres de ces deux catégories montre que les locuteurs natifs les produisent avec une moins grande distance acoustique dans le plan des deux premiers formants que les apprenants (voir Tableau XXIV en rouge colonne ʌ). Il semble que dans cette région, les participants français utilisent davantage le timbre des voyelles pour produire des voyelles qui soient à la fois distinctes entre elles mais aussi, pour le /æ/ et le /ɑː/, différentes du /a/ français. Au sujet du /ʌ/, on note que les productions des apprenants s'étendent sur une région bien plus centrale que les productions des natifs, ce qui vient appuyer l'idée que les apprenants utilisent le /œ/ comme cible articulatoire.

Tableau XXIV : Distances acoustiques moyennes entre les voyelles d'une même région acoustique dans les voyelles produites par les locuteurs natifs à gauche et les locuteurs français à droite

Voyelles anglaises (L1)	æ		ʌ		ɪ		Voyelles anglaises (L2)	æ		ʌ		ɪ	
	F	M	F	M	F	M		F	M	F	M	F	M
ʌ	315	275					ʌ	218	216				
ɑː	450	402	168	140			ɑː	545	560	503	384		
iː					456	319	iː					152	139

4.3.1.2 Analyse des durées vocaliques

Les voyelles de l'anglais produites par des locuteurs natifs (Figure 36, en bleu) présentent des différences de durées significatives, $F(4, 148) = 258.57$, $p < .0001$. Les voyelles /æ/ (290 ms), /ɑː/ (340 ms) et /iː/ (287 ms) sont plus longues que les voyelles /ʌ/ (172 ms) et /ɪ/ (159 ms) (Figure 36, en bleu). Les voyelles du français présentent également des différences de durées significatives, $F(3, 141) = 55.80$, $p < .001$, mais cette différence n'est portée que par la voyelle /œ/ produite en contexte bVR, qui a pour effet d'allonger la durée de la voyelle (Léon, 2011) (Figure 36, en orange). Ainsi les voyelles du français ont des durées similaires aux voyelles courtes de l'anglais (sauf le /œ/ qui est plus long) et toutes sont significativement plus courtes que les voyelles longues de l'anglais (y compris /œ/). Lorsque les apprenants français produisent les voyelles anglaises (Figure 36, en vert), les durées des voyelles courtes (/ʌ/ et /ɪ/)

sont similaires aux durées de ces mêmes voyelles produites par des locuteurs natifs (en bleu), $F(1, 84) = 1.66$, $p > .05$. En revanche, pour les voyelles longues, on observe que les durées des voyelles produites par les locuteurs français sont significativement plus courtes que celles produites par les locuteurs natifs, $F(1, 84) = 7.732$, $p < .001$. Cependant, pour ces mêmes voyelles, on observe bien un allongement de la durée par rapport aux voyelles françaises $F(1, 335) = 14.56$, $p < .001$. Ainsi, il semblerait qu'en L2, les apprenants français soient capables de produire un contraste de durée bien que la durée ne soit pas phonologique dans leur L1.

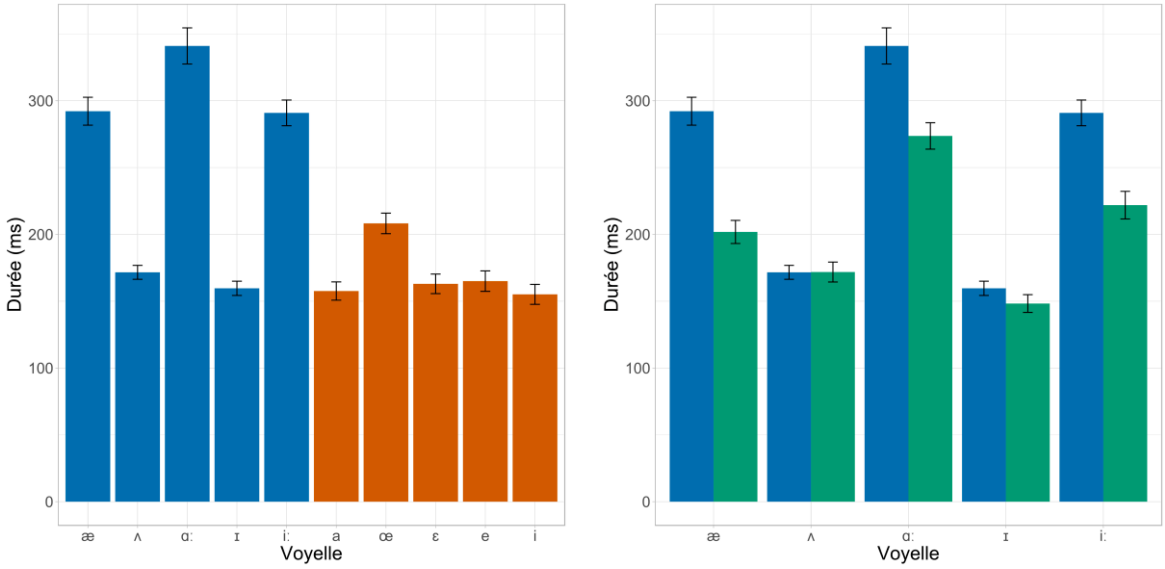

Figure 36 : Durées moyennes des voyelles anglaises (en bleu) et françaises (en orange) produites par des locuteurs L1 à gauche. Durées moyennes des voyelles anglaises produites par des locuteurs de L1 (en bleu) et de L2 (en vert) à droite

4.3.1.3 Détermination automatique des voyelles problématiques du point de vue de la réalisation acoustique

Afin d'évaluer les catégories de voyelles qui posent le plus problème aux apprenants, l'ensemble des voyelles produites par les apprenants ont été classifiées selon un modèle d'analyse linéaire discriminante appris sur les productions des locuteurs natifs. Pour rappels, le nombre de voyelles considéré dans ces analyses est présenté dans le Tableau XXII, p. 102). Mairano et al. (2019) ont montré qu'une analyse discriminante linéaire (*LDA*) des voyelles d'apprenants italophones de l'anglais présentait des résultats corrélés à des évaluations des

évaluations de « *nativelikeness* [10]» et de compréhensibilité telles que mesurées sur des échelles de *Likert*. Cependant, dans cette étude, les classifications étaient entrainées sur les voyelles des participants eux-mêmes et non à partir de cibles natives, ce qui ne permet pas des tester la précision avec laquelle elles sont produites en termes de proximité avec les mêmes voyelles produites par des natifs. Pour cela, nous avons choisi de classifier les voyelles de nos apprenants selon un modèle d'analyse discriminante linéaire appris sur les voyelles produites par des locuteurs natifs. Ainsi, nous pensons que nous obtiendrons une mesure qui soit le reflet de la précision des productions des voyelles de nos apprenants.

Différents modèles d'analyse discriminante linéaire (fonction `lda` du package `MASS` dans R) ont été testés afin de déterminer celui qui permet de classifier le mieux les voyelles des locuteurs natifs. Les différents modèles testés et leur pourcentage de classification correcte sont présentés dans le Tableau XXV.

Tableau XXV : Pourcentages de classification correcte des voyelles produites par les locuteurs natifs ayant enregistré les stimuli et ceux du groupe N selon les différents modèles d'analyse linéaire discriminante testés

Modèle	Classification correcte (%)
lda.m1 = lda(Vowel ~ F1Bkz+F2Bkz+Duration, data = train)	**99.8**
lda.m2 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+DCT2, data = train)	99.8
lda.m3 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+DCT3, data = train)	99.8
lda.m4 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+DCT2+DCT3, data = train)	99.8

Les quatre modèles testés présentent des taux de classification strictement identiques. C'est donc le modèle le plus parcimonieux qui a ensuite été sélectionné pour classifier les voyelles des apprenants. Ce modèle (lda.m1) utilise les valeurs des deux premiers formants en Bark centrés-réduits ainsi que la durée des voyelles. La matrice de confusion de cette classification est présentée dans la Figure 37 à gauche. On observe que selon ces paramètres, les voyelles de la région du [i] sont correctement classifiées à 100% alors que dans la région du [a], le modèle commet en particulier des confusions entre les voyelles ɑː et /ʌ/.

[10] L'évaluation de *nativelikeness* revient à évaluer dans quelle mesure un locuteur non natif s'exprime à la manière d'un locuteur natif.

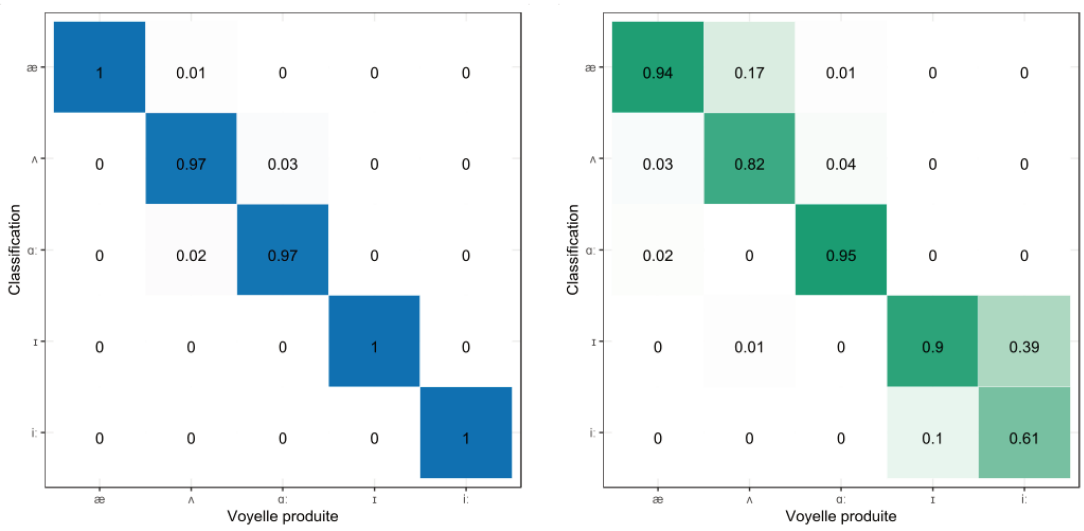

Figure 37 : Matrice de confusion des voyelles des natifs selon le modèle de classification lda.m1 (en bleu à gauche) et des apprenants (en vert à droite)

La classification des voyelles produites par les apprenants selon ce modèle d'analyse linéaire discriminante montre un taux de classification correcte global de 84.6%. Les proportions de classification des voyelles selon les cinq catégories de voyelles sont présentées dans la matrice de confusion de la Figure 37 à droite (en vert). On observe des taux de classification globalement plus bas que ceux des locuteurs natifs.

Cette classification basée sur les deux premiers formants et de la durée des voyelles fait émerger que les voyelles /ʌ/ et /i:/ sont celles qui sont les moins bien classifiées. La voyelle /ʌ/ présente en effet un taux de classification de 17% comme /æ/ et la voyelle /i:/ est classifiée à hauteur de 39% comme /ɪ/ et ce malgré le fait que la classification prend en compte la durée de la voyelle. On peut ainsi penser que, bien que les apprenants soient capables de produire un allongement des voyelles comme vu dans le paragraphe 4.3.1.2, le fait qu'ils produisent les voyelles /i:/ et /ɪ/ avec des centres de catégories trop proches (voir paragraphe 4.3.1.1.3) entraine un fort taux de confusion entre ces deux voyelles.

4.3.1.1 Analyse dynamique des voyelles /i:/ et /i/

L'objectif de cette analyse était de déterminer si la diphtongaison de la voyelle /i:/ était un paramètre pertinent pour évaluer la qualité des productions de cette voyelle par les apprenants francophones. Les comparaisons de différents modèles d'analyse discriminante linéaire (voir Tableau XXVI) montrent que les valeurs des deux premiers formants prises au milieu temporel des voyelles suffisent à distinguer les voyelles /i:/ anglais (97%) et /i/ français (92%) produites

par des natifs lorsque ces seules voyelles sont intégrées dans l'analyse. Un modèle incluant uniquement C1 (deuxième paramètre de la transformée en cosinus discrète) classifie correctement les /i/ français (91%) mais pas les /iː/ anglais (9%), ce qui indique que ce paramètre est plus pertinent pour classifier les /i/ français que les /iː/ anglais, probablement à cause de la non-systématicité de la diphtongaison du /iː/ anglais (Ferragne, 2008). Le meilleur modèle (`lda.m4` dans le Tableau XXVI) pour classifier les voyelles natives a été ensuite utilisé pour classifier les /iː/ des apprenants français. Le modèle ainsi appris sur les /iː/ anglais et /i/ français produits par des locuteurs de L1 porte sur F1, F2, la durée, C1 et C2. Les /iː/ des apprenants ont été soumises à ce modèle qui classe 54% des voyelles de L2 comme des /iː/ anglais et 46% comme des /i/ français. Bien que non significatif, ($\chi^2_{(1)} = 0.8705$, $p = .3508$), ce résultat semble indiquer qu'au moins une partie des /iː/ produits par les apprenants présentent des paramètres acoustiques semblables à ceux des voyelles produites par les natifs, y compris des paramètres dynamiques.

Tableau XXVI : Taux de classification correcte des voyelles [i] natives selon les différents modèles testés

Modèle de classification	Classification correcte (%)
lda.m1 = lda(Vowel ~ F1Bkz+F2Bkz, data = train)	94.3
lda.m2 = lda(Vowel ~ F1Bkz+F2Bkz+Duration, data = train)	95.9
lda.m3 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+C1, data = train)	96.3
lda.m4 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+C1+C2, data = train)	**96.7**
lda.m5 = lda(Vowel ~ DCT2, data = train)	56.6

4.3.2 Évaluation de l'effet des entraînements sur la production

L'effet des entraînements a été évalué de deux manières : 1) à partir des paramètres temporels et spectraux des voyelles étudiées ; 2) à l'aide d'une tâche d'identification des voyelles produites réalisées par des locuteurs natifs.

En ce qui concerne les propriétés acoustiques des voyelles, nous avons utilisé les mêmes mesures que celles utilisées pour décrire les productions des apprenants avant entraînement, c'est-à-dire que nous avons observé les modifications de la distribution des voyelles produites dans les groupes dans le plan des deux premiers formants ainsi que sur les durées vocaliques, enfin nous avons soumis les voyelles produites avant et après entraînement à une classification automatique entrainée sur les données des locuteurs natifs pour déterminer si les productions en posttest étaient mieux classifiées après entraînement. Nos hypothèses pour ces différents points étaient les suivantes : Tout d'abord, nous nous attendions à ce que les participants ayant suivi l'entraînement en production présentent le plus de changement sur le plan des paramètres

acoustiques des voyelles. Une amélioration de la qualité des productions devrait se traduire tout d'abord par une diminution de la variabilité inter-catégorielle observée dans le 4.3.1.1 pour les voyelles de L2. De plus, nous nous attendions à ce que les voyelles /iː/ et /ʌ/ qui posent particulièrement problème aux apprenants présentent des distributions plus proches de celles des natifs en posttest. À savoir, nous pouvons nous attendre à une plus grande distance acoustique entre les voyelles /iː/ et /ɪ/, et à une augmentation du premier formant pour la voyelle /ʌ/. En ce qui concerne les durées des voyelles, nous nous attendons à une augmentation des différences entre voyelles courtes et longues en posttest. Ces changements acoustiques devraient induire une augmentation des proportions de classification correcte des voyelles des apprenants à l'analyse discriminante et de la part des locuteurs natifs dans la tâche d'identification.

En ce qui concerne les différences intergroupes, nous prévoyons que le groupe ayant été soumis à l'entraînement en production soit celui qui présente le plus de changements entre prétest et posttest, suivi du groupe PE puis du groupe C. Nous pensons que le même schéma devrait se produire à niveau de l'évaluation par des locuteurs natifs.

4.3.2.1 Évaluation objective

Dans ce paragraphe, nous décrirons les changements acoustiques des voyelles produites par les apprenants selon les entraînements qui ont été suivi. Nous nous appuierons sur les mesures des deux premiers formants et de durées qui ont été effectuées. Nous présenterons dans un premier temps les résultats sur les durées vocaliques, puis ceux sur la dispersion des voyelles produites dans le plan des deux premiers formants, enfin nous présenterons le résultat d'une classification automatique des voyelles produites selon ces paramètres acoustiques.

4.3.2.1.1 Résumé du corpus utilisé

Afin d'évaluer l'effet des entraînements proposés sur les performances en production, nous avons utilisé un corpus, résumé dans le Tableau XXVII, constitué :

1. Des productions en anglais de mots bVd (3 occurrences par mot) des locuteurs francophones indépendamment pour chacun des trois groupes expérimentaux PE, PR

et C confondus ; n = 16 par groupe) en prétest (T1) et en posttest (T2), soit un total de 48 occurrences par voyelles et par groupe à T1 et à T2.

2. Des productions en anglais de mots bVd (3 occurrences par mot) du groupe N (n = 16), soit un total de 48 occurrences par voyelles.
3. Des productions en anglais de mots bVd (3 occurrences par mot) des locuteurs natifs ayant enregistré le matériel expérimental (n = 18 dont 9 femmes), soit un total de 54 occurrences par voyelles. Cet ensemble a servi à entrainer un modèle d'analyse discriminante linéaire.

Tableau XXVII : Composition du corpus utilisé pour évaluer les performances en production en prétest et en posttest

	Prétest					**Posttest**				
Voyelle	**æ**	**ʌ**	**ɑː**	**ɪ**	**iː**	**æ**	**ʌ**	**ɑː**	**ɪ**	**iː**
Groupe C	48	48	48	48	48	48	48	48	48	48
Groupe PE	48	48	48	48	**43**	48	48	48	48	48
Groupe PR	48	48	48	48	48	**47**	48	48	**47**	48
Groupe N	48	48	48	48	48	-	-	-	-	-
Locuteurs natifs	**53**	54	54	54	54	-	-	-	-	-

Cet ensemble contient au total 1947 voyelles puisque 2 voyelles en posttest (indiquées en gras dans le Tableau XXVII) ont été retirées des analyses, l'une car le fichier son était vide, le locuteur avait prononcé son mot trop tôt, l'autre car les formants n'étaient pas visibles dans le spectrogramme. Cinq productions ont été supprimées en prétest dans le groupe PE car les participants (deux femmes) ont incorrectement prononcé le mot « bead » [bɛd]. De plus dans le corpus issu du matériel expérimental, une voyelle avait été retirée car le spectrogramme n'était pas lisible.

4.3.2.1.2 Effet des entraînements sur les durées vocaliques

Les durées moyennes des voyelles produites par les locuteurs anglais natifs du groupe N et des apprenants des groupes C, PE et PR en prétest (en bleu) et posttest (en rouge) sont présentés dans la Figure 38.

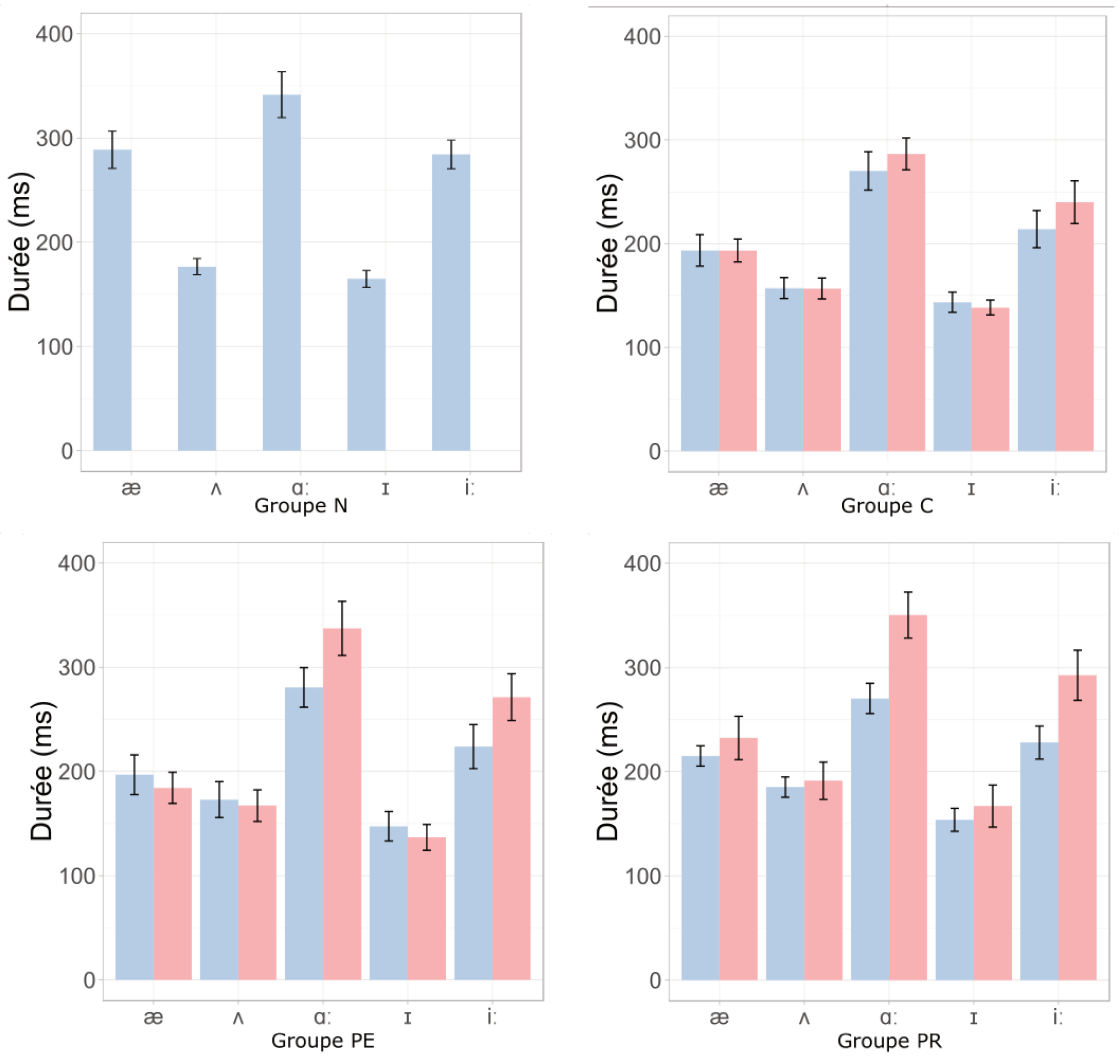

Figure 38 : Durées des voyelles produites par chacun des groupes, en prétest en bleu et en posttest en rouge

Les durées des voyelles ont d'abord été analysées séparément en prétest et en posttest à partir des données des quatre groupes (N, C, PE et PR). Deux modèles mixtes (un en prétest et un en posttest) ont été générés avec les facteurs Groupe et Voyelle en facteur fixes et le facteur Sujet en facteur aléatoire. L'interaction Groupe × Voyelle était spécifiée dans ces modèles. Les analyses post-hoc ont été menées en ne gardant que les contrastes de durée pertinents : /iː/ vs. /ɪ/, /ɑː/ vs. /ʌ/ et /æ/. Une ANOVA **en prétest** sur ce modèle montre des effets principaux significatifs du Groupe ($\eta^2_p = 0.07$) et de la Voyelle ($\eta^2_p = 0.75$) ainsi qu'une interaction Groupe × Voyelle significative ($\eta^2_p = 0.21$) (voir Tableau XXVIII). Des analyses post-hoc montrent que les apprenants des trois groupes ne produisent pas les voyelles courtes /ʌ/ et /ɪ/ avec des durées différentes des locuteurs natifs. En revanche, les participants des groupes PE, et C produisent la voyelle /ɑː/ statistiquement plus courte que le groupe N, $M_{Nɑː\text{-}PEɑː} = 64.38$,

$t(134) = 3.80$, $p = .027$ et , $M_{Nɑː\text{-}Cɑː} = 70.21$, $t(134) = 4.15$, $p = .009$, et les groupes PR et C produisent la voyelle /iː/ significativement plus courte que les natifs, $M_{Niː\text{-}PRiː} = 71.35$, $t(134) = 4.22$, $p = .007$ et $M_{Niː\text{-}Ciː} = 71.35$, $t(134) = 4.22$, $p = .007$. On peut également noter que les trois groupes produisent des /æ/ plus court que les natifs. On peut toutefois noter que dans chacun des groupes, les contrastes de durée sont réalisés de manière significative : en effet, dans tous les groupes, la durée de la voyelle /ɑː/ est plus élevée que celle des voyelles /æ/ et /ʌ/ et celle de la voyelle /iː/ est plus élevée que celle de la voyelle /ɪ/ (tous les $p < .05$).

Tableau XXVIII : Table d'analyse des durées vocaliques en fonction du Groupe et de la Voyelle en prétest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	F	p	η^2_p
Groupe	20196	6732	3	60	6.619	0.001***	0,074
Voyelle	764122	191030	4	240	187.818	0.000***	0,752
Groupe:Voyelle	65627	5469	12	240	5.377	0.000***	0,206

En posttest, les effets principaux du Groupe ($\eta^2_p = 0.05$) et de la Voyelle ($\eta^2_p = 0.81$) sont toujours significatifs et l'interaction Groupe × Voyelle l'est également ($\eta^2_p = 0.19$) (voir Tableau XXIX). Les analyses post-hoc indiquent que seule la voyelle /æ/ produite par les groupes PE et C est produite différemment du groupe natif en terme de durée. On observe également, que les contrastes de durée /iː - ɪ/ et /ɑː - ʌ/ sont produits de manière significative dans les trois groupes d'apprenants (tous les $p < .05$), les voyelles longues étant bien produites avec des durées plus élevées que les voyelles courtes.

Tableau XXIX : Table d'analyse des durées vocaliques en fonction du Groupe et de la Voyelle en posttest

Source	SC	CM	ddl_{effet}	ddl_{erreur}	F	p	η^2_p
Groupe	15050	5017	3	60	3.997	0.012***	0.046
Voyelle	1343000	335750	4	240	267.498	0.000*	0.812
Groupe:Voyelle	71179	5932	12	240	4.726	0.000**	0.187

Afin d'évaluer l'effet des entraînements, nous avons généré un modèle linéaire mixte avec les données des trois groupes d'apprenants (sans celles du groupe N) en prétest et en posttest avec les facteurs Test, Groupe et Voyelle comme facteurs explicatifs et le facteur Sujet comme facteur aléatoire. Toutes les interactions étaient spécifiées dans ce modèle. Le modèle, dont les résultats sont présentés dans le Tableau XXX, montre un effet significatif du Test ($\eta^2_p = 0.08$) et de la Voyelle ($\eta^2_p = 0.73$). L'interaction Test × Groupe est significative ($\eta^2_p = 0.03$), elle reflète, selon des analyses post-hoc, le fait que le groupe PR, $M_{T1\text{-}T2} = -36.22$, $t(405) = -6.39$, $p < .001$, allonge ses durées vocaliques significativement entre prétest et posttest. Le Groupe PE, $M_{T1\text{-}T2} = -15.78$, $t(405) = -2.78$, $p < .07$, les allongent également mais de manière tendancielle alors que le groupe C ne produit qu'un allongement non significatif, $M_{T1\text{-}T2} = -$

7.33, $t(405) = -1.28$, $p > .10$. L'interaction Test × Voyelle est également significative ($\eta^2_p = 0.11$), et des analyses post-hoc indiquent que les modifications de durées se produisent uniquement sur les voyelles /iː/, $M_{T1\text{-}T2} = -47.26$, $t(405) = -6.45$, $p < .001$, et /ɑː/, $M_{T1\text{-}T2} = -51.01$, $t(405) = -6.97$, $p < .001$. Les interactions Groupe × Voyelle et Test × Groupe × Voyelle ne sont pas significatives.

Tableau XXX : Table d'analyse des durées vocaliques en fonction du Test (prétest vs. posttest), du Groupe et de la Voyelle.

Source	SC	CM	ddl_{effet}	ddl_{erreur}	F	p	η^2_p
Test	46923	46923	1	405	36.473	0.000***	0.082
Groupe	5167	2583	2	45	2.008	0.146	0.010
Voyelle	1389842	347460	4	405	270.084	0.000***	0.726
Test:Groupe	17649	8825	2	405	6.859	0.001**	0.033
Test:Voyelle	69236	17309	4	405	13.454	0.000***	0.116
Groupe:Voyelle	15008	1876	8	405	1.458	0.171	0.028
Test:Groupe:Voyelle	11818	1477	8	405	1.148	0.330	0.022

4.3.2.1.3 Description de l'effet des entraînements sur les deux premiers formants des voyelles

Dans cette section, nous décrirons l'évolution de la dispersion des voyelles produites par les apprenants des différents groupes dans le plan des deux premiers formants entre prétest et posttest au moyen de représentations graphiques des ellipses de dispersion des catégories vocaliques.

Les voyelles produites des locuteurs natifs sont ainsi représentées dans la Figure 39.

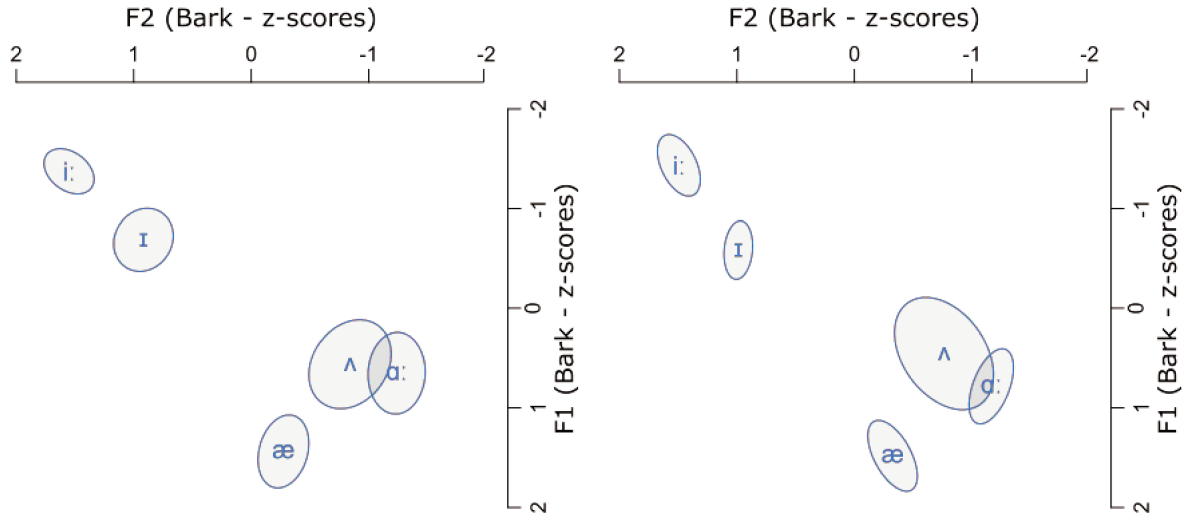

Figure 39 : Représentation des voyelles produites par les 16 locuteurs natifs (Groupe N) dans l'espace des deux premiers formants en Bark centrés-réduits. Les voyelles produites par les femmes (n = 15) sont présentées à gauche, celles des hommes (n = 5) à droite. Les symboles phonétiques représentent les barycentres des catégories de voyelles dans le groupe, les ellipses représentent un écart-type

Pour les femmes et les hommes du groupe N, nous observons des ellipses de dispersion plus petites que celles des apprenants représentées dans la Figure 40, avec aucun chevauchement inter-catégoriel dans la région du [i], la voyelle /æ/ qui est réalisée dans une région acoustique des voyelles /ɑ:/ et /ʌ/ qui elles sont produites.

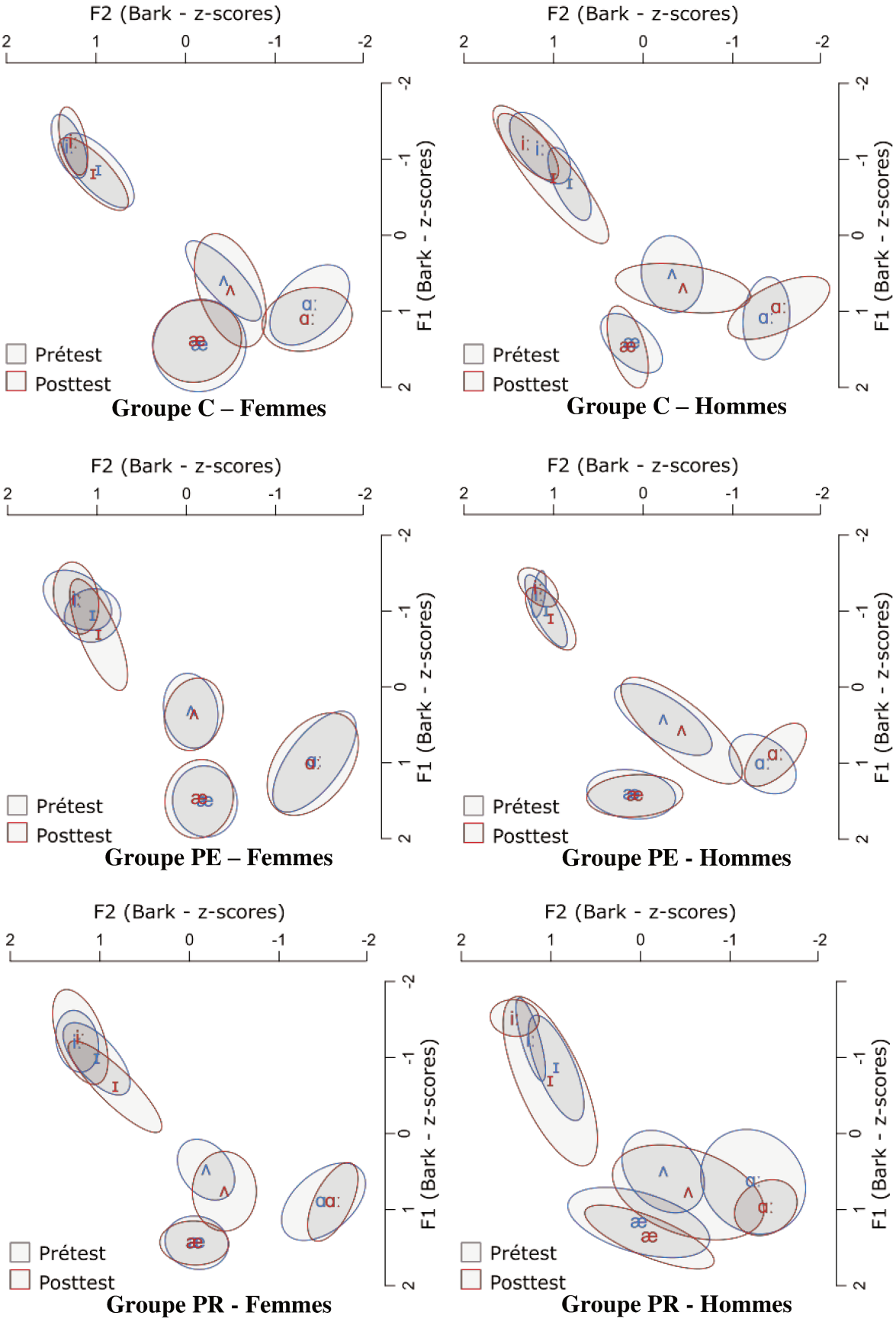

Figure 40 : Ellipses de dispersion des voyelles produites en prétest (en bleu) et en posttest (en rouge) pour chacun des trois groupes d'apprenants

En ce qui concerne le **groupe C**, on observe que les femmes montrent très peu de changement entre prétest et posttest, les ellipses et les barycentres de catégories sont quasi superposés. Chez les hommes, on observe de grandes ellipses de dispersion tant en prétest, qu'en posttest ainsi que des chevauchements inter-catégoriels importants signe que la production de ces voyelles est très variable au sein du groupe.

On observe peu de changement entre prétest et posttest dans le **groupe PE**. Le seul changement remarquable est pour la voyelle /ɪ/ chez les femmes dont l'ellipse tend à s'étendre vers le bas. L'observation des profils individuels pour les voyelles de la région du [i] montre que l'ensemble des participants de ce groupe produisent les deux voyelles dans une région similaire entre prétest et posttest avec peu de changement sauf deux participantes qui produisent le /ɪ/ avec des valeurs de F1 plus élevées et de F2 moins élevées en posttest. La distribution des voyelles produites par l'une d'elle, la participante PE05 est présentée dans la Figure 41. On peut voir pour cette participante qu'en prétest, les deux voyelles sont produites sur un même espace alors qu'en posttest, elle produit le /ɪ/ avec des valeurs de F1 très bas, à la hauteur de son /ʌ/. À l'écoute ses voyelles ressemblent beaucoup à des /ɛ/.

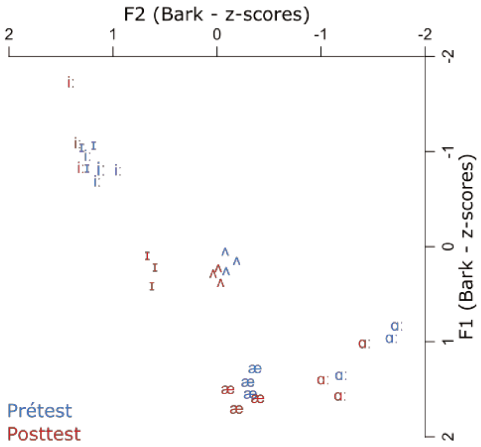

Figure 41 : Voyelles produites par la participante **PE05** dans l'espace des deux premiers formants en Barks centrés-réduits. Les symboles phonétiques bleus indiquent les voyelles produites en prétest, ceux en rouge indiquent celles produites en posttest

En ce qui concerne les hommes de ce groupe, les voyelles de la région du [i] ne semblent pas avoir été affectées par l'entraînement. En effet, alors qu'elles aient été initialement produites avec une forte confusion dans l'espace des deux premiers formants en prétest, c'est toujours le cas en posttest. Le /æ/ est produit avec plus de dispersion que dans le groupe natif mais n'occupe pas l'espace d'une autre voyelle de la région. En revanche, on observe des changements pour les voyelles /ɑː/ et /ʌ/. D'une part, les ellipses s'étendent davantage

postérieurement pour le /ɑː/ et vers le bas pour le /ʌ/, ce qui entraine les barycentres de ces deux catégories. Cela indique qu'au moins une partie de locuteurs ont modifié leurs productions entre prétest et posttest. De la même manière que chez les femmes, cette variation entre prétest et posttest est due à un seul participant qui postériorise son /ɑː/ et abaisse un peu son /ʌ/ entre prétest et posttest. Les autres hommes tendent à ne pas modifier leurs productions entre prétest et posttest pour ce contraste.

Dans le groupe PR, chez les femmes, bien que le barycentre soit resté stable, on observe une diminution de l'ellipse de dispersion pour du ɑː qui s'est affinée. En revanche, la variabilité des productions des voyelles /ʌ/ a augmenté et son barycentre s'est abaissé indiquant que certaines femmes produisent cette voyelle de manière plus ouverte après entraînement. Pour les voyelles de la région du [i], l'ellipse du /ɪ/ s'est allongée vers le bas et son barycentre s'est abaissé, indiquant ici encore que certaines femmes produisent cette voyelle plus distincte du /iː/ en posttest qu'en prétest. Concernant le /iː/, on observe que le barycentre n'a pas bougé mais que l'ellipse de dispersion s'est réduite indiquant que les productions ont convergé vers une cible plus stable. Chez les hommes de ce groupe, comme chez les femmes, la variabilité des /iː/ a nettement diminué dans l'espace des deux premiers formants entre prétest et posttest. On observe que les barycentres des deux [i] se sont éloignés et l'ellipse du /ɪ/ s'est étendue signe d'une plus grande variabilité en posttest dans le groupe. Pour les voyelles de la région du [a], on observe que les 3 barycentres se sont abaissés et que les ellipses de /æ/ et ɑː ont diminué contrairement au /ʌ /dont l'ellipse s'est étendue entre prétest et posttest.

4.3.2.1.4 Choix du modèle de classification automatique pour évaluer les productions des apprenants en prétest et posttest

Nous avons souhaité évaluer les productions des apprenants de manière automatique à partir des paramètres acoustiques des voyelles enregistrées. Dans le même temps, nous avons souhaité comparer les apprenants au groupe N. Pour cela, nous avons utilisé un modèle d'analyse discriminante linéaire appris sur les voyelles produites par les locuteurs ayant produit le matériel expérimental en contexte bVd. Nous avons choisi d'utiliser un modèle d'apprentissage différent de celui nous ayant permis de déterminer les voyelles posant problèmes aux francophones décrit dans le paragraphe 4.3.1 car ce dernier avait été entrainé sur un ensemble de données contenant les données du groupe N or nous souhaitons aussi classifier les données de ce groupe afin de pouvoir les comparer à nos apprenants. L'ensemble

de données que nous avons utilisé pour entrainer le modèle contenait 269 voyelles (voir Tableau XXVII, p116). Différents modèles ont été comparés afin de déterminer le meilleur modèle permettant de classifier les voyelles des locuteurs natifs ayant enregistré le matériel expérimental. Le modèle le plus parcimonieux permettant la meilleure classification est celui portant sur F1, F2 (en Bark centrés réduits) et la durée. Ce modèle classifie correctement 268 voyelles sur 269 (un /ɑː/ est classifié comme /ʌ/ selon ce modèle) soit 99,7% de classification correcte.

Tableau XXXI : Différents modèles de classification utilisés pour classifier les voyelles produites par les locuteurs natifs ayant enregistré le matériel expérimental

Modèle de classification	Classification correcte (%)
`lda.m1 = lda(Vowel ~ F1Bkz+F2Bkz+Duration, data = train)`	99.7
`lda.m2 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+C2, data = train)`	99.7
`lda.m3 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+C2, data = train)`	98.9
`lda.m4 = lda(Vowel ~ F1Bkz+F2Bkz+Duration+C1+C3, data = train)`	99.2

4.3.2.1.5 Classification automatique des données des apprenants en prétest et posttest

Les productions des apprenants indépendamment à T1 et T2 ont ensuite été classifiées pour chacun des groupes à partir de ce modèle (`lda.m1`). La Figure 42 présente les taux de classification correcte des voyelles produites par les apprenants à T1 en bleu et T2 en rouge selon les groupes.

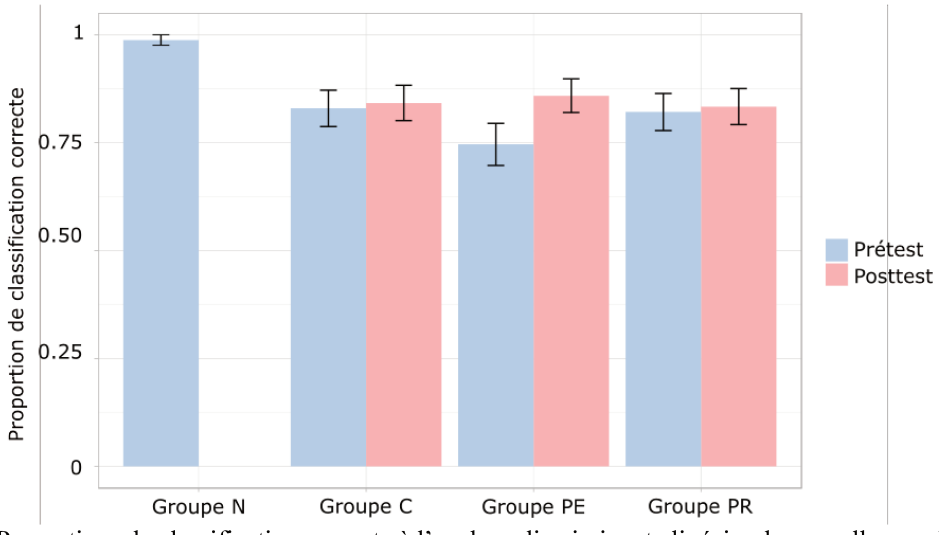

Figure 42 : Proportions de classification correcte à l'analyse discriminante linéaire des voyelles produites selon les groupes en prétest en bleu et en posttest en rouge

Les proportions de classification correcte des apprenants ont d'abord été comparées à celles des participants du groupe N indépendamment pour les productions issues du prétest et du

posttest. Des régressions logistiques à effets mixtes avec la fonction `glmer` du package `lme4` ont été utilisées. Les valeurs de *p* ont été obtenues avec la fonction `Anova` du package `car` et des tests post-hoc ont été réalisés avec la fonction `lsmeans` du package `lsmeans`. Le facteur Groupe a été inclus comme facteur fixe et le facteur Sujet comme facteur aléatoire dans chacun des deux modèles (prétest et posttest). Le facteur Voyelle n'a pas été inclus dans l'analyse car les modèles ne parvenaient à converger.

En prétest, les résultats montrent un effet du Groupe significatif, $\chi^2(3) = 24.93$, $p < .0001$, et les post-hoc indiquent que les trois groupes d'apprenants ont des scores inférieurs au groupe de locuteurs natifs (tous les $p < .001$) et que les trois groupes C, PE et PR ne diffèrent pas significativement entre eux. **En posttest**, l'effet du groupe est toujours significatif, $\chi^2(3) = 18.78$, $p < .001$, avec le groupe N présentant des scores plus élevés que les trois autres groupes (tous les $p < .001$) et que les scores des groupes C, PE et PR ne sont pas significativement différents.

L'effet des entraînements sur ces scores de classification correcte a été évalué à l'aide d'une régression logistique à effets mixtes en ne prenant que les scores des groupes C, PE et PR. Les facteurs Test et Groupe ont été inclus comme facteurs fixes et le facteur Sujet comme facteur aléatoire. Les résultats montrent un effet du Test significatif, $\chi^2(1) = 4.410$, $p < .05$, et les tests post-hoc indiquent que les scores sont meilleurs en posttest, log odd ratio $_{\text{prétest – posttest}} = -0$, 6022, SE = 0,248, z = -2,432, $p = .015$, $p_{\text{prétest}} = 0{,}80$ and $p_{\text{posttest}} = 0{,}84$. L'effet du Groupe et l'interaction Groupe × Test ne sont pas significatifs indiquant que les trois groupes s'améliorent entre prétest et posttest.

La matrice de confusion issue de la classification des voyelles produites par les participants du groupe N est présentée dans la Figure 43.

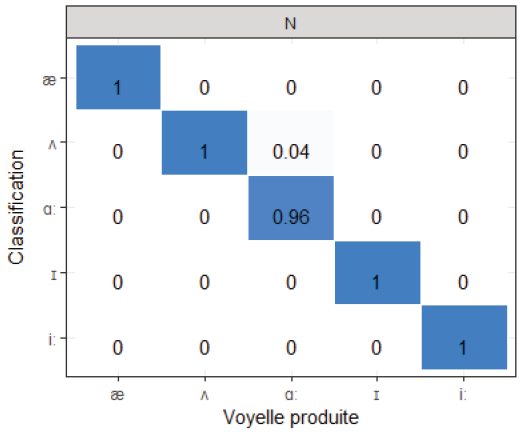

Figure 43 : Matrice de confusion issue de la classification des voyelles du Groupe N selon le modèle lda.m1 appris sur les données des locuteurs ayant enregistré le matériel expérimental

Les données des locuteurs natifs sont correctement classifiées à 98.75% sur la base des deux premiers formants et de la durée des voyelles. Seulement deux ɑː sont faussement classifiés comme /ʌ/. Ce taux de classification élevé confirme que cette classification est adaptée.

Des matrices de confusions entre voyelles ont été calculées pour chacun des groupes en prétest et posttest. Ces résultats sont présentés dans la Figure 44.

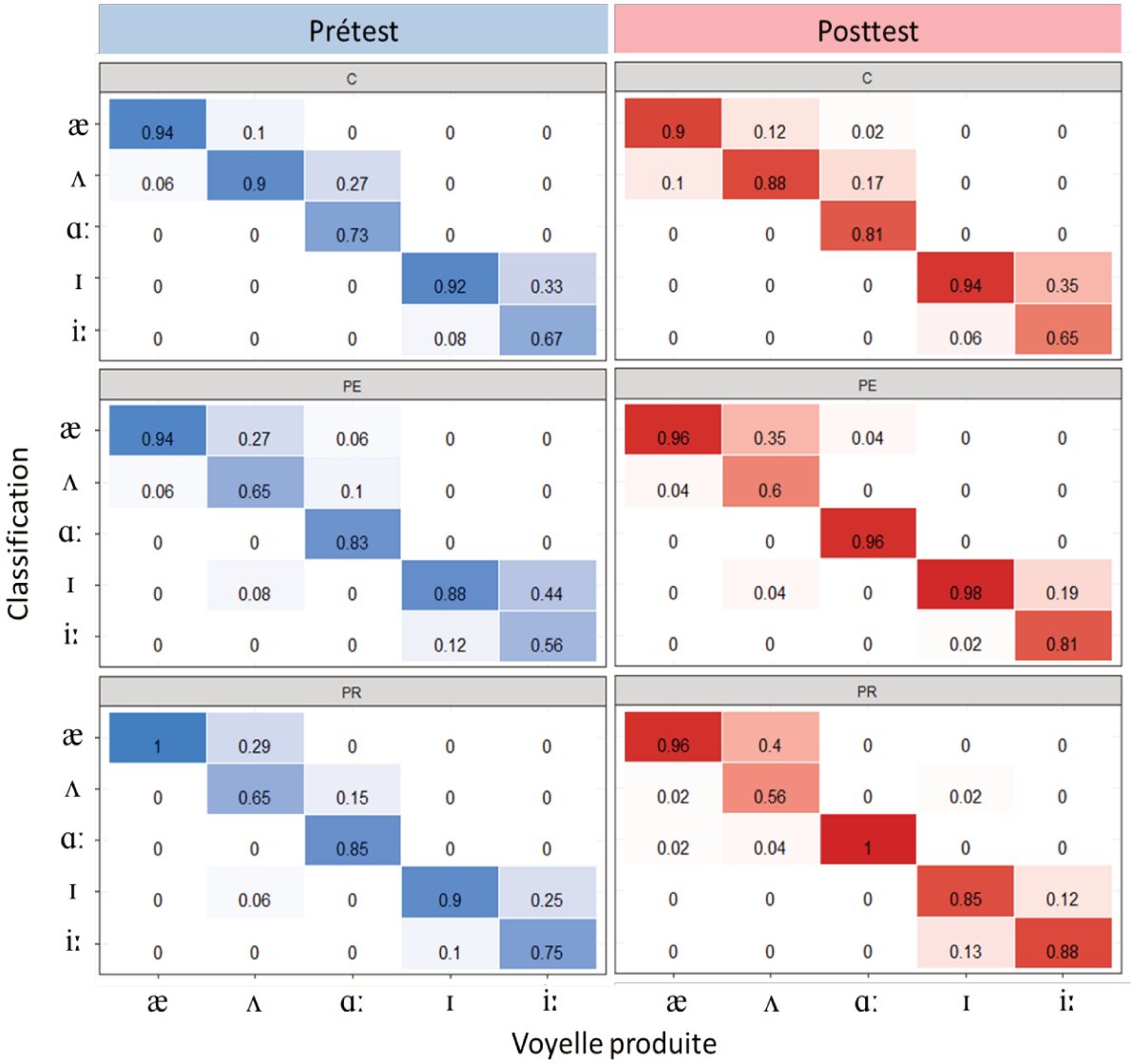

Figure 44 : Matrices de confusion de la classification des voyelles produites par les apprenants en prétest à gauche et en posttest à droite. L'intensité de la couleur code la proportion de classification : une proportion proche de 1 avec une couleur intense, une proportion proche de 0 avec une intensité de couleur faible

De manière générale, le modèle ne fait pas de classification inattendue, sauf des /ʌ/ classifiés comme /ɪ/, et inversement, mais nous avons vu plus haut que les réalisations de ces voyelles dans l'espace des deux premiers formants pouvaient parfois conduire à un rapprochement des espaces occupés par ces voyelles.

Les classifications des voyelles des apprenants avant entraînement indépendamment pour chacun des groupes montrent que les *patterns* observés dans les groupes sont similaires à ceux

observés précédemment tous groupes confondus. En effet, les voyelles qui conduisent le plus à des erreurs de classification sont les voyelles /ʌ/ et /iː/, sauf pour le groupe C pour lequel le ɑː conduit à plus d'erreur de classification que le /ʌ/.

On observe pour les deux groupes PE et PR, que le / ʌ/ est fortement confondu avec le /æ/, à 27 et 29% respectivement, ce qui n'est pas du tout le cas dans le groupe C. Pour les 3 groupes, le ɑː est assez fortement confondu avec le /ʌ/, 27% des cas pour le groupe, C, 10 et 15% pour les groupes PE et PR respectivement.

En ce qui concerne les différences entre prétest et posttest, on observe que les participants du groupe C voient leurs scores de classification se dégrader pour les voyelles /ʌ/ et /iː/, alors qu'ils obtiennent de meilleurs scores pour les voyelles /æ/, /ɑː/ et /ɪ/.

Dans le groupe PE, on constate que toutes les voyelles à l'exception du /ʌ/ montrent une augmentation du pourcentage de classification correcte. Ce constat est particulièrement vrai pour les voyelles de la région [i] puisque l'amélioration se traduit à la fois par des /iː/ moins confondus avec des /ɪ/ et ces derniers moins confondus avec les /iː/ en posttest par rapport au prétest. Pour les voyelles de la région du [a], on constate une dégradation pour la voyelle /ʌ/ qui est confondue avec le /ɑː/ dans 35% des cas. Enfin, pour ce groupe, on peut observer que la voyelle /ɑː/ n'est plus confondue en posttest avec la voyelle /ʌ/ comme c'était le cas en prétest. Au regard des ellipses de dispersion en prétest et posttest présentées dans la Figure 40 qui ne montraient quasiment pas de changement dans le plan des deux premiers formants, on peut penser que l'allongement des durées vocaliques a joué un rôle dans l'amélioration des classifications des voyelles pour ce groupe.

Enfin pour le groupe ayant reçu l'entraînement en production (Groupe PR), on observe une amélioration de la classification pour les voyelles longues /ɑː/ et /iː/ uniquement alors que les autres voyelles voient leurs scores se dégrader. 11% supplémentaires des /ʌ/ sont classifiés comme /æ/ en posttest par rapport au prétest. Pour ce groupe, l'analyse des durées vocaliques a montré une augmentation significative des durées vocaliques des voyelles longues entre prétest et posttest, ce qui a pu entrainer une meilleure classification de ces dernières, la durée étant prise en compte dans l'analyse. Dans ce groupe, les ellipses de dispersion avaient montré, certes un abaissement des voyelles /ɪ/ et /ʌ/ les rendant plus distinctes de leurs voisines respectives mais également un étalement des ellipses de dispersion, indiquant qu'une plus grande proportion des voyelles produites en posttest l'étaient dans un espace acoustique

partagé avec celui des voyelles voisines ce qui a pu conduire aux erreurs de classification accrues.

4.3.2.2 Évaluation subjective par des locuteurs natifs

Les évaluations des productions avant et après entraînements des trois C, PE, PR ainsi que celles du groupe N ont été évaluées par 6 juges natifs (voir section 4.2.3).

Des accords inter-juges (Kappa de Fleiss[11]) selon les groupes de juges ayant évalué les mêmes items (voir Tableau XXI, p101) ont été calculés avec la fonction `kappam.fleiss` du package `irr` de R. Ils étaient de 67.5% pour le Groupe 1, 73.1% pour le Groupe 2, 72.1% pour le Groupe 3 et 67,7% pour le Groupe 4. Les accords inter-juges pour chaque voyelle dans les groupes sont présentés dans la Figure 45. De manière descriptive, on n'observe pas de nette différence entre les groupes, sauf pour la voyelle /ʌ/ qui présente un accord inter-juges plus élevé dans le groupe 2 par rapport aux autres groupes mais surtout globalement assez bas. Cela rappelle les scores d'identification par les locuteurs natifs qui étaient assez bas pour cette voyelle.

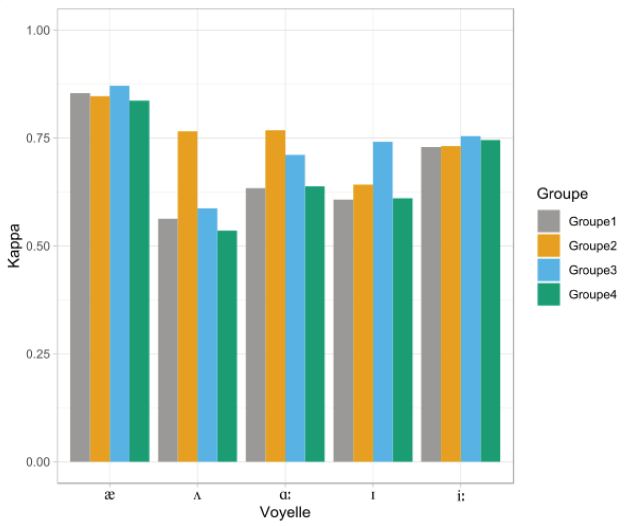

Figure 45 : Accords inter-juges (Kappa de Fleiss) selon les voyelles dans les groupes de juges

[11] Le coefficient Kappa de Fleiss est compris dans l'intervalle]-∞ ; 1], plus sa valeur tend vers 1, plus l'accord entre les juges est élevé. Un kappa entre 0.41 et 0.60 indique un accord modéré, alors qu'un kappa entre 0.61 et 0.80 indique un accord fort (Fleiss, 1971).

Les proportions d'évaluation correctes toutes voyelles confondues pour chacun des groupes en prétest et en posttest sont présentés dans la Figure 46. Les scores obtenus par les enregistrements des apprenants ont d'abord été comparés aux scores des enregistrements natifs indépendamment pour les productions issues du prétest et du posttest. Des régressions logistiques à effets mixtes avec le facteur Groupe en facteur fixe et le facteur Sujet en facteur aléatoire ont été modélisées.

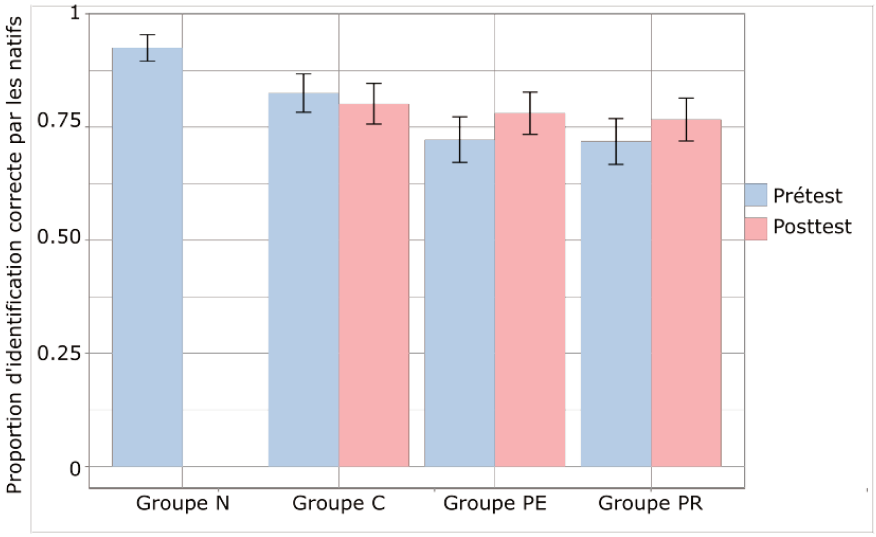

Figure 46 : Proportions d'identification correcte des voyelles produites par les apprenants en prétest en bleu, en posttest en rouge et par les locuteurs natifs dans la tâche de jugement par des locuteurs natifs

En prétest, les résultats montrent un effet du Groupe significatif, $\chi^2(3) = 59.50$, $p < .0001$, et les post-hoc indiquent que les trois groupes d'apprenants ont des scores inférieurs au groupe de locuteurs natifs (tous les $p < .001$) et que le groupe C a de meilleurs scores que les groupes PE, log odd ratio $_{C-PE} = 0.63$, SE = 0,213, z = 2.973, $p = .016$ ($P_C = 0.83$ et $P_{PE} = 0.72$) et PR, log odd ratio $_{C-PR} = -0.66$, SE = 0.213, z = 3.089, $p = .011$ ($P_{PR} = 0.72$). **En posttest**, l'effet du groupe est toujours significatif, $\chi^2(3) = 34.11$, $p < .001$, avec le groupe N présentant des scores plus élevés que les trois autres groupes (tous les $p < .001$) mais les différences ne sont plus significatives entre les groupes d'apprenants.

L'effet des entraînements sur les proportions d'identification correctes des productions des apprenants par les juges natifs a été évalué au moyen d'une régression logistique à effets mixtes avec les facteurs Test et Groupe en facteur fixe et le facteur aléatoire Sujet. L'effet principal du Test est significatif, $\chi^2(1) = 4.98$, $p < .05$ et l'effet du Groupe est tendanciel, $\chi^2(2) = 5.38$, $p < .07$. L'interaction Test × Groupe est significative, $\chi^2(2) = 8.08$, $p < .05$. Les analyses post-hoc montrent que les deux groupes PE, log odd ratio $_{T2-T1} = 0.33$, SE = 0.125, z = 2.624, $p = .009$, et PR, log odd ratio $_{T2-T1} = 0.27$, SE = 0.123, z = 2.185, $p = .029$, améliorent

significativement leurs performances contrairement au groupe C, log odd ratio $_{T2-T1}$ = -0.16, SE = 0.138, z = -1.180, *p* = .24.

Des matrices de confusion issues de la tâche d'identification réalisées par les 6 juges natifs sur les voyelles produites par les participants du groupe N (voir Figure 47) et des groupes C, PE et PR en prétest et posttest (voir Figure 48) ont été calculées.

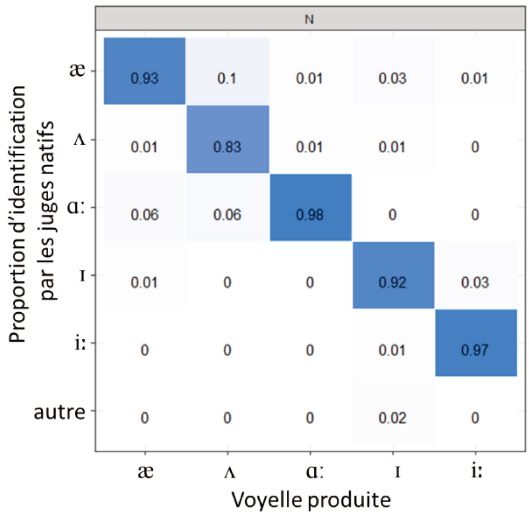

Figure 47 : Matrice de confusion issue de la tâche d'identification réalisée par les juges natifs des voyelles produites par les locuteurs du groupe N

Les voyelles produites par les locuteurs du groupe N ne sont pas identifiées avec des taux d'identification maximaux. En effet, on remarque que chacune des 5 voyelles est parfois confondue avec ses voisines dans l'espace acoustique. Le /ʌ/ est celle qui présente le plus fort taux de confusion, notamment avec la voyelle /ɑː/.

De manière générale, on observe que les locuteurs produisent des catégorisations très inattendues, par exemple des /iː/ catégorisés comme /æ/ et ce dans des proportions qui ne sont pas négligeables (e.g., 6% dans le groupe PE en prétest). L'écoute des fichiers sons concernés par ces classifications montrent que pour la plupart il s'agit du mot « bead » prononcé « bed » comme nous l'avions déjà souligné plus haut. Bien que nous ayons pris soin d'ajouter le label « autres » permettant de parer à ce genre de cas, certains juges semblent ne pas l'avoir utilisé chaque fois que nécessaire. Pour d'autres items ainsi étrangement catégorisés, nous n'avons pas d'autres explications que l'erreur de boutons car certaines voyelles du groupe N ont été

également catégorisées ainsi alors qu'elles ne semblaient à l'écoute ne présenter aucune ambiguïté.

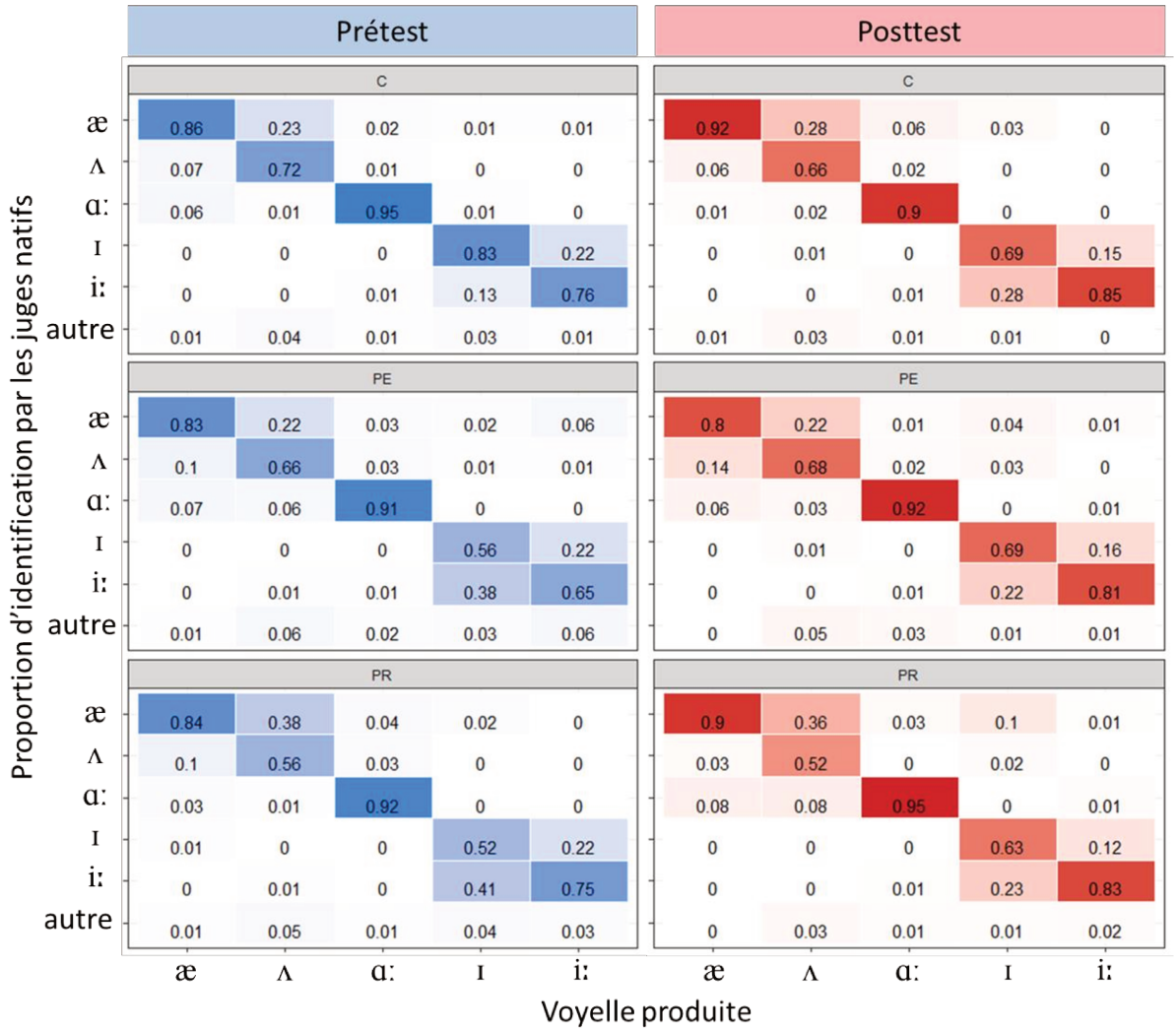

Figure 48 : Matrices de confusion issues de la tâche d'identification réalisée par les juges natifs sur les données apprenants en prétest et en posttest

Concernant les voyelles produites par les apprenants, la matrice de confusion issue de la tâche d'identification montre que les voyelles /ʌ/ et /ɪ/, suivies du /i:/ sont celles qui sont les moins bien reconnues par les locuteurs natifs en prétest dans nos 3 groupes d'apprenants. De manière intéressante, on observe dans les 3 groupes que la voyelle /ʌ/ est fortement confondue avec le /æ/.

Dans le groupe C, deux catégories sont /æ/ et /i:/ sont mieux classifiées en posttest par rapport au prétest alors que les autres voyelles présentent le schéma inverse. On observe que le groupe PE améliore ses performances pour toutes les voyelles sauf le /æ/ et que cette amélioration est particulièrement importante pour les voyelles /ɪ/ (13%) et /i:/ (16%). En revanche, le /ʌ/ ne

prend que 2% d'identification correcte supplémentaire, il est moins confondu avec le ɑː mais en grande proportion confondu avec le /æ/. Dans le groupe PR, la plus grande amélioration s'observe pour la voyelle /iː/ qui est moins confondue avec la voyelle /ɪ/ et l'inverse est également vrai pour ce groupe. Le /ʌ/ semble perdre en précision de réalisation puisqu'il est davantage confondu avec le /ɑː/ en posttest.

4.3.2.3 Comparaison des deux méthodes d'évaluation

Afin de tester la comparabilité des deux méthodes d'évaluation que nous avons utilisées, nous avons testé les corrélations entre les scores moyens des participants avec la *LDA* et ceux obtenus par identification par des juges natifs dans chacun des groupes en prétest et posttest. Les tests de corrélation ne montrent pas de lien entre ces deux mesures, ni en prétest, $r(46) = 1.80$, $p > .05$, ni en posttest, $r(46) = -0.92$, $p > .3$. Cela incite à considérer ces mesures comme complémentaires plutôt que redondantes. De plus, nous avons réalisé un test de corrélation sur les différences entre prétest et posttest des scores obtenus par participant par chacune des deux méthodes d'évaluation. Nous nous attendions à une corrélation positive entre ces deux mesures. Cette corrélation n'est pas significative, $r(46) = 1.65$, $p > .10$, indiquant ainsi qu'une augmentation du score de classification correcte à l'analyse discriminante linéaire n'implique pas une meilleure classification des voyelles par des locuteurs natifs. Les densités de distribution en fonction de la proportion de classification correcte par individu selon le groupe et le test (prétest et posttest) sont présentées dans la Figure 49 ; celles des proportions

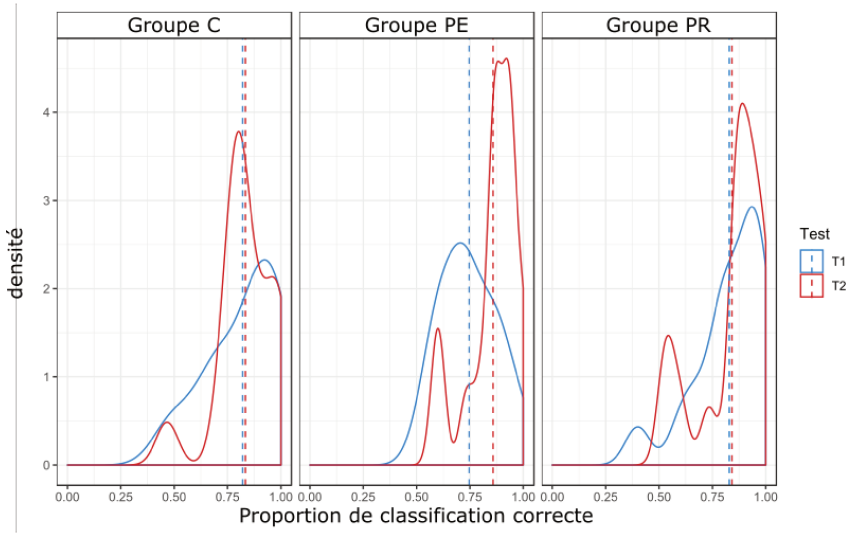

Figure 49 : Distribution des scores individuels moyens de classification correcte à l'analyse discriminante linéaire selon le groupe et le test (prétest vs. posttest). Les barres verticales pointillées indiquent les moyennes des distributions

d'identification correcte issues de la tâche d'identification par les juges natifs sont présentées dans la Figure 50.

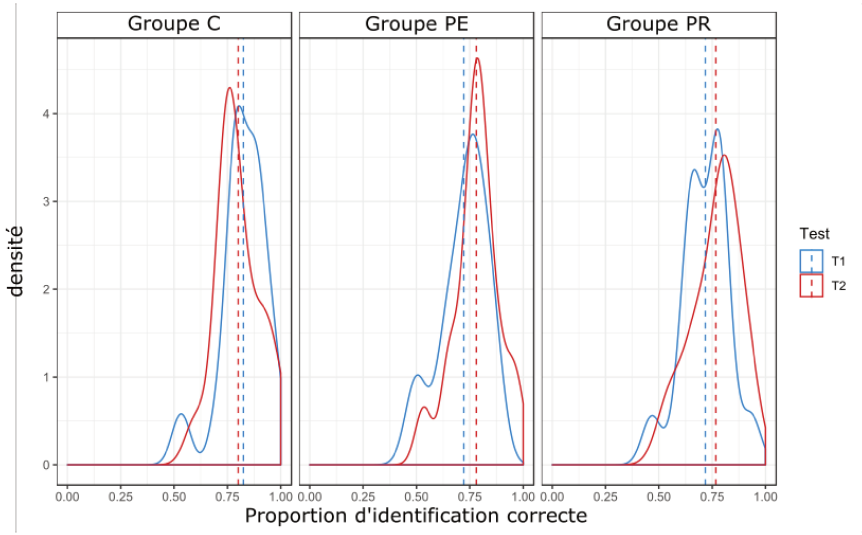

Figure 50 : Distribution des scores individuels moyens d'identification correcte par les juges natifs selon le groupe et le test (prétest vs. posttest). Les barres verticales pointillées indiquent les moyennes des distributions

De manière surprenante, alors qu'aucune différence entre groupes n'est apparue sur les scores de classification correcte à l'analyse discriminante contrairement aux scores d'identification correcte par des juges natifs, la comparaison de ces distributions semble montrer que l'effet des entraînements est plus visible sur les paramètres mesurés par l'analyse discriminante. En effet, les formes des distributions issues de l'analyse discriminante sont plus différentes entre prétest et posttest que ne le sont celles issues de l'identification par des natifs. On remarque également que les formes des distributions issues de la classification par *LDA* après entraînement sont plus proches de celles issues de l'identification par des locuteurs natifs. En particulier, on observe que la distribution des scores de classification sur la *LDA* devient bimodale en posttest, ce qui n'était pas le cas en prétest. On peut penser que les groupes étaient plus homogènes en prétest et que, en posttest, ils le sont moins. Ceci est visible surtout dans les groupes ayant suivi un entraînement, et pourrait indiquer qu'au sein des groupes, certains participants ont bénéficié des entraînements, avec une amélioration des performances à la clé, alors que pour d'autres les éventuels modifications ne seraient pas conformes aux attentes en L2. De plus, l'augmentation du nombre de participants ayant des scores élevés dans ces groupes est très marquée par rapport au groupe C. On observe que cet effet est moins évident sur les jugements par les natifs mais cette mesure tient probablement compte de changements de plus grande ampleur (changements phonémiques) dans un espace très multidimensionnel

alors que la mesure faite par *LDA* tient compte de changements plus fins (changements phonétiques) dans un espace limité à peu de dimensions.

4.4 Discussion des résultats en production

Les participants francophones (groupes C, PE et PR) ont été enregistrés dans une tâche de lecture de mots bVd isolés afin d'évaluer leur manière de produire les voyelles anglaises /æ - ʌ - ɑː - ɪ - iː/. Les participants francophones ont également enregistré les voyelles françaises /a - œ - e - ɛ - i/. Les voyelles en contexte /bVd/ extraites des enregistrements réalisés par les locuteurs ayant constitué le matériel expérimental ont également été considérées ainsi que celles du groupe de locuteurs natifs (Groupe N). Les objectifs de cette partie étaient 1) de caractériser les réalisations vocaliques des apprenants francophones avant entraînement, 2) de déterminer dans quelle mesure les entraînements ont eu un impact sur ces réalisations.

4.4.1 Évaluation avant entraînement

Nous avons dans un premier temps cherché à caractériser la manière dont les apprenants de l'anglais francophones produisaient les voyelles de l'anglais de notre étude par rapport à des locuteurs de l'anglais britanniques natifs. Pour cela, nous avons analysé les paramètres acoustiques des voyelles produites par l'ensemble des 48 participants français de notre étude et les avons comparés aux paramètres acoustiques des voyelles produites par les 34 locuteurs anglais britanniques natifs que nous avons enregistrés dans cette étude que ce soit pour la constitution du matériel expérimental ou dans le groupe expérimental constituant le groupe natif.

Nous avons tout d'abord analysé les voyelles produites dans les deux langues du point de vue spectral. Nous avons pu noter que les voyelles des deux langues (anglais et français) étaient produites avec des variabilités inter-catégorielles similaires par des locuteurs natifs. En revanche, nous avons observé, que les voyelles de L2 étaient produites avec une plus grande variabilité intra-catégorielle que celles de L1, qu'elles soient comparées aux voyelles anglaises, ou aux voyelles françaises. De manière générale, ces analyses ont montré que les voyelles de L2 sont produites avec une plus grande variabilité que celles de L1. Ceci peut s'expliquer par le fait que les cibles articulatoires ne sont pas clairement établies chez les apprenants (Gick, Bernhardt, Bacsfalvi et Wilson, 2008).

À la lumière des deux distributions observées pour les voyelles des deux langues et en supposant que les espaces vocaliques couverts par leurs réalisations soient le reflet, au moins en partie de leurs espaces perceptifs, on peut relire les observations faites sur leurs performances dans les tâches de perception au regard notamment du modèle PAM de Best que nous avons présentés dans le 0. Une interprétation selon ce modèle nous inviterait à penser que le contraste /æ - ʌ/, dont les réalisations acoustiques observées dans nos données montrent que le premier membre tombe en dehors de toutes catégories existantes en français et que le second peut au moins en partie être assimilé à une catégorie du français (le /a/ ou le /œ/), serait assimilé selon le type « Non Catégorie – Catégorie », menant à une discrimination bonne à très bonne de celui-ci. Nos résultats en perception montraient en effet que ce contraste était le mieux discriminé parmi les trois contrastes testés dans nos trois groupes. Selon les distributions observées sur les productions, le contraste /iː - ɪ/ devrait être traité de la même manière, voire assimilé selon deux catégories prédisant ainsi une bonne discrimination des deux voyelles.

Pour les contrastes de l'anglais /iː - ɪ/ et /ʌ - ɑː/, qui portent à la fois sur le timbre et la durée, nous avons observé des phénomènes différents dans les productions des locuteurs natifs. En effet, les catégories /iː/ et /ɪ/ sont produites par les natifs avec une plus grande distance acoustique entre les centres de catégories que ne le sont les voyelles /ʌ/ et /ɑː/. Le plus grand chevauchement des aires des catégories /ʌ/ et /ɑː/ résultant est contrebalancé par un contraste de durée plus marqué pour ce contraste. Chez les apprenants, ce schéma est différent : ils produisent le contraste /iː - ɪ/ avec une faible distance acoustique entre les catégories et donc un fort chevauchement entre leurs aires, alors qu'ils produisent le contraste /ʌ - ɑː/ avec une grande distance entre les catégories et un faible chevauchement entre leurs aires, et ce, malgré une grande variabilité intra-catégorielle. Il semblerait que, pour les voyelles de la région du [a] (/æ/ compris), les apprenants cherchent à produire des voyelles qui se distinguent du point de vue du timbre, à la fois entre elles, mais aussi du /a/ français. Cette observation renvoie au phénomène de dissimilation décrit dans le *Speech Language Model* de Flege (1995). Selon ce modèle, la probabilité de formation d'une catégorie phonémique augmente avec la dissimilarité entre un son de L2 et le plus proche voisin en L1. En l'absence de formation de nouvelle catégorie, les propriétés d'un son L2 et d'un son L1 voisin peuvent fusionner pour former une catégorie « composite ». Les éléments phonétiques des deux inventaires coexistent dans un espace phonologique commun et interagissent les uns avec les autres conduisant parfois à des déviations (assimilation ou dissimilation) des catégories voisines de L1, L2 et L1/L2 pour maintenir les contrastes phonétiques.

Pour le contraste /i: - ɪ/, les apprenants semblent au contraire présenter des difficultés à produire des voyelles distinctes du point de vue du timbre, ne produisant quasiment qu'un contraste de durée. Bien que des précautions aient été prises lors de la présentation des consignes (le mot était accompagné du symbole phonétique de la voyelle), on peut imaginer pour ce contraste que les participants francophones aient été influencés par le système orthographique de leur L1 selon lequel la graphie du mot « *bit* » induit la production d'un [i] (Nimz, 2013). Selon cette interprétation, la réalisation du /ɪ/ tendrait à se rapprocher de celle du /i/ pour des raisons orthographiques, et ce malgré le fait que le /e/ soit acoustiquement un meilleur support articulatoire pour la réalisation du /ɪ/. Au sujet de la voyelle /ʌ/ plus particulièrement, on a pu remarquer que la distribution des voyelles produites par les apprenants s'étendait de manière plus importante que chez les natifs en direction de la région du /œ/, ce qui confirme que cette voyelle pourrait en effet servir de support articulatoire aux apprenants francophones.

En ce qui concerne la durée des voyelles, nos données montrent que, bien que la durée ne soit pas phonologique en français, les apprenants francophones produisent des voyelles longues de l'anglais plus longues que les voyelles françaises même si celles-ci restent moins longues que celles des locuteurs natifs.

Afin de déterminer de manière automatique les voyelles qui posent problème aux apprenants francophones, nous avons mis en place une classification des voyelles produites par les francophones selon un modèle d'analyse discriminante linéaire appris sur les données des locuteurs anglais natifs. Les paramètres pris en compte dans cette analyse étaient les deux premiers formants et la durée. Le modèle avait permis de classifier 99.7% des voyelles natifs, commettant des erreurs entre les voyelles /ʌ/ et ɑː alors que les voyelles de la région du [i] était classifiées à 100% selon ces paramètres. Ces résultats sur les voyelles natives rappellent ceux que nous avons obtenus dans le groupe N en perception, puisque nous avions observé que les voyelles de la région du [a], notamment les voyelles /ʌ/ et /ɑː/ présentaient des taux d'identification significativement plus bas que celles de la région du [i]. Certes, la tâche de perception reposait sur un ensemble restreint (i.e., elle ne comportait pas les voyelles produites par le groupe N) mais le fait que la classification automatique suive les résultats obtenus en perception par le groupe de locuteurs natifs nous laisse penser que cette classification est un bon outil d'évaluation des productions.

Enfin, l'analyse dynamique pour les voyelles /iː/ de l'anglais et /i/ du français montre tout d'abord que les indices statiques des deux premiers formants seuls suffisent à catégoriser ces

voyelles alors qu'un indice dynamique seul n'est pertinent que pour les voyelles françaises, ce qui pourrait indiquer une certaine variabilité dans les productions des /i:/ par les natifs (Ferragne, 2008). Enfin, une partie des productions des apprenants présentent des paramètres similaires aux voyelles natives.

4.4.2 Évaluation des entraînements

Pour évaluer l'effet des entraînements sur les productions, nous avons utilisé deux méthodes : une première méthode que nous qualifions d'« *objective* » (V. Delvaux, Huet, Piccaluga et Harmegnies, 2014) qui consiste en une analyse des paramètres acoustiques des voyelles produites avant et après entraînement ; une seconde méthode basée sur les jugements subjectifs des voyelles produites par les apprenants avant et après entraînement par des locuteurs natifs dans une tâche d'identification.

La classification automatique utilisée reposait sur les deux premiers formants des voyelles et sur leur durée. Le modèle de classification ayant été appris sur des valeurs provenant de voyelles produites par des locuteurs anglais natifs, une meilleure classification devrait indiquer que les paramètres pris en compte se rapprochent de ceux des locuteurs natifs. Dans la littérature sur les études ayant évalué l'effet d'entraînements sur la production de voyelles, différentes méthodes « objectives » ou tout du moins automatiques ont été utilisées : certaines ont considéré qu'une augmentation des distances euclidiennes entre les voyelles produites en L2 après entraînement était signe d'une plus grande distinguabilité entre elles et donc d'une « efficacité » de l'entraînement (e.g., Rato & Rauber, 2015). D'autres ont pris comme indice de l'efficacité de l'entraînement la diminution des distances de Mahalanobis entre les voyelles produites et des cibles natives (e.g., Kartushina et al., 2015). Nous avons souhaité utiliser une autre méthode car une augmentation de la distance acoustique entre les voyelles de L2 ou une plus grande proximité avec des cibles natives après entraînement peut certes, indiquer que les entraînements ont eu un effet, mais ces mesures ne préviennent pas des cas où le changement n'empêche pas que la voyelle produite soit toujours confondue avec une voyelle voisine, même si la distance les séparant a augmenté. En prenant en compte non seulement les valeurs formantiques mais aussi les durées des voyelles dans une classification dont le modèle a été appris sur des valeurs issues de locuteurs natifs, nous disposons d'une sorte de double validation du modèle en utilisant un groupe de locuteurs natifs (groupe N) dont les productions sont évaluées par ce modèle également. Bien que cette classification se soit montrée efficace

pour classifier les voyelles des locuteurs natifs du groupe N, elle ne nous a pas permis de distinguer les groupes selon le type d'entraînements qu'ils ont reçu. Les résultats issus de cette méthodologie montrent que les trois groupes (C, PE et PR) ont des scores de classification correcte de leurs voyelles à l'analyse discriminante significativement plus faibles que ceux des locuteurs anglais natifs du Groupe N avant comme après entraînement. Pour aucune des voyelles, les entraînements, de type *HVPT* ou celui orienté sur la production, n'ont permis d'atteindre un niveau natif. De plus, les résultats selon cette méthode ne permettent pas de mettre en évidence de différences significatives entre les groupes. On observe en effet que les trois groupes ne se distinguent les uns des autres ni en prétest, ni en posttest et que l'interaction Test × Groupe n'est pas significative, c'est-à-dire qu'aucun des groupes ne présente un changement prétest-posttest différent des autres groupes. Cependant, l'analyse des paramètres acoustiques ayant contribué à la classification que nous avons utilisée montre que les groupes ayant suivi des entraînements produisent des durées vocaliques différentes entre prétest et posttest avec en particulier un allongement des voyelles longues. Cela n'est pas le cas du groupe contrôle (groupe C). En revanche, l'analyse des deux premiers formants que nous avons réalisée montre une très grande hétérogénéité dans les groupes et suivant les voyelles, et ne nous permet pas de tirer des conclusions quant à l'influence des entraînements proposés sur ces paramètres.

En ce qui concerne l'identification des voyelles produites dans les groupes par des juges natifs, les analyses montrent qu'en prétest et en posttest, les participants des 3 groupes C, PE et PR ont des scores inférieurs aux locuteurs natifs du groupe N. On observe que le groupe C présente en prétest de meilleurs taux d'identification correcte que les autres groupes en prétest. La comparaison de l'évolution des performances selon l'entraînement suivi montre que seuls les participants des groupes PE et PR progressent significativement après entraînement mais peut-être seulement parce qu'ils avaient une plus grande marge de progression que le groupe C. Pour tester cette hypothèse, nous avons vérifié la corrélation entre les performances en prétest et les différences entre prétest et posttest avec l'hypothèse de l'existence d'une corrélation négative entre les deux. C'est-à-dire qu'une grande amélioration des performances entre prétest et posttest serait liée à un plus faible niveau de performances en prétest. Le résultat de ce test ne met en évidence de corrélation significative, $r(46) = -0.475$, $p > .6$.

Nous avons souhaité utiliser ces deux méthodes que nous jugions complémentaires puisque la première tend davantage à mettre en exergue les modifications au niveau acoustique efficaces

qui sont induites par les entraînements, elle s'apparente plutôt à la détection d'erreurs phonétiques alors que la seconde permet d'évaluer l'efficacité des entraînements dans le sens où ils améliorent l'intelligibilité des apprenants par des locuteurs natifs. Cette seconde méthode repose davantage sur la détection d'erreurs phonémiques. La méthode par classification automatique donne une mesure de l'accent étranger, alors l'identification par des locuteurs natifs est plutôt une mesure de compréhensibilité. En ce qui concerne la comparabilité des deux méthodes, un test de corrélation réalisé sur les différences entre prétest et posttest des scores de classification correcte à l'analyse discriminante linéaire et des scores d'identification correcte par les juges natifs montrent que ces mesures ne sont pas corrélées. Ainsi, une amélioration de la classification automatique des voyelles produites n'implique pas nécessairement que celles-ci soient mieux identifiées par des locuteurs natifs.

5 Évaluation en EEG

5.1 Objectifs et hypothèses

Dans cette partie, nous avons souhaité mettre en évidence des corrélats électroencéphalographiques liés à l'acquisition des voyelles consécutivement à nos entraînements. Pour cela, nous avons mis en place une expérience en EEG permettant d'observer la MMN puisque cette onde s'est avérée être sensible à un entraînement sur un contraste vocalique dans l'étude menée par Ylinen et collaborateurs (2010). Nous nous sommes intéressés à l'acquisition du contraste /æ-ʌ/ par nos apprenants francophones. Afin de comparer la MMN observée pour cette paire de voyelles à celle observée pour une paire de voyelles dont le contraste est présent dans le système phonologique du français, nous avons également enregistré les réponses évoquées pour la paire /æ-ɛ/. Nous avons choisi cette seconde paire car, la MMN étant sensible aux propriétés acoustiques des stimuli, en ce qu'elle varie avec la distance acoustique entre les stimuli présentés (Näätänen, Paavilainen, Rinne et Alho, 2007), la paire /æ-ɛ/ est la paire de voyelles qui présente la distance euclidienne entre voyelles la plus comparable à celle de la paire /æ-ʌ/. De plus, afin d'observer dans quelle mesure les réponses évoquées chez les apprenants se rapprochaient de celles de locuteurs natifs, nous avons également testé notre groupe de locuteurs natifs pour ces deux contrastes. Les enregistrements en EEG avaient lieu consécutivement à la passation des épreuves comportementales lors de la première séance en prétest et lors de la dernière séance en posttest

pour les participants des groupes C, PE et PR et lors de l'unique séance de test pour les participants du groupe N.

Nos hypothèses concernant la MMN étaient les suivantes :

H1 : En prétest, chez le groupe N, nous pensions observer des MMN équivalentes pour les contrastes /æ-ʌ/ et /æ-ɛ/. Pour les trois groupes d'apprenants, (C, PE et PR), nous nous attendions à ce que la MMN générée par le contraste /æ-ɛ/ soit plus grande que celle générée par le contraste /æ-ʌ/ puisque ce dernier n'appartient pas au système phonologique des apprenants et que ce contraste n'implique a priori pas de franchissement de frontière phonologique en français.

H2 : Après entraînement, nous n'attendons pas de changement sur la MMN pour le groupe C comparativement au prétest. Nous nous attendons à ce que la MMN pour les deux contrastes soit toujours plus faible dans les groupes d'apprenants PE et PR que dans le groupe natif mais dans une moindre mesure qu'en prétest.

H3 : Nous nous attendons à ce que la MMN pour les groupes PE et PR soit de plus grande amplitude en posttest qu'en prétest pour le contraste /æ-ʌ/ qui a été entrainé, ce qui serait le signe que les entraînements ont provoqué des changements au niveau pré-attentionnel.

5.2 Méthodologie

5.2.1 Matériel expérimental

Dans la majorité des études sur la MMN, les stimuli standards et déviants sont des exemplaires uniques, acoustiquement identiques, des catégories phonologiques ce qui impliquent que la MMN obtenue est, au moins en partie, le reflet de la perception de différences acoustiques (Näätänen et al., 1997). Afin d'exclure la possibilité que la MMN observée soit uniquement le reflet des propriétés acoustiques de nos stimuli, nous avons choisi d'adapter le paradigme *Oddball* classique en présentant une certaine variabilité dans les stimuli (Ghislaine Dehaene-Lambertz & Pena, 2001 ; Miglietta et al., 2013 ; Shestakova et al., 2002). Ainsi, nous avons décidé d'utiliser cinq occurrences différentes de chacun des stimuli. Un locuteur masculin de langue maternelle anglaise (dont les productions ont été utilisées dans les phases test) a enregistré 5 occurrences différentes des mots anglais *bat*, *but* et *bet* spécifiquement pour la

tâche en EEG. Les stimuli enregistrés ont été segmentés et normalisés en amplitude du pic. Les paramètres acoustiques des voyelles (durée et valeurs centrales des deux premiers formants des voyelles) ont été extraits. Les propriétés temporelles et spectrales des stimuli sont présentées dans le Tableau XXXII. Une représentation dans l'espace des deux premiers formants des voyelles est présentée dans la Figure 51. Les distances euclidiennes séparant les barycentres des catégories sont de 1.90 Bark centrés-réduits pour la paire /æ-ʌ/ et de 2.59 Bark centrés-réduits pour /æ-ɛ/ ce qui représente respectivement 334 Hz et 474 Hz dans l'espace en Hertz.

Tableau XXXII : Caractéristiques temporelles (en ms) et spectrales (en Hz) des stimuli utilisés dans la tâche en EEG. Début et fin indique le début et la fin de la voyelle par rapport au début du fichier son.

Voyelle	Occurrence	Début	Fin	Durée	Durée du fichier son	F1	F2
/æ/	1	43	181	138	384	725	1522
	2	32	185	153	399	687	1512
	3	55	217	162	478	701	1495
	4	25	151	127	376	694	1593
	5	44	179	135	384	703	1482
/ɛ/	1	46	154	108	364	608	1664
	2	32	147	115	387	584	1700
	3	40	162	122	368	562	1704
	4	28	165	137	381	600	1635
	5	35	166	131	438	606	1601
/ʌ/	1	49	152	103	416	687	1190
	2	69	199	129	431	646	1171
	3	34	158	124	368	643	1236
	4	40	150	110	359	663	1176
	5	37	183	146	368	630	1180

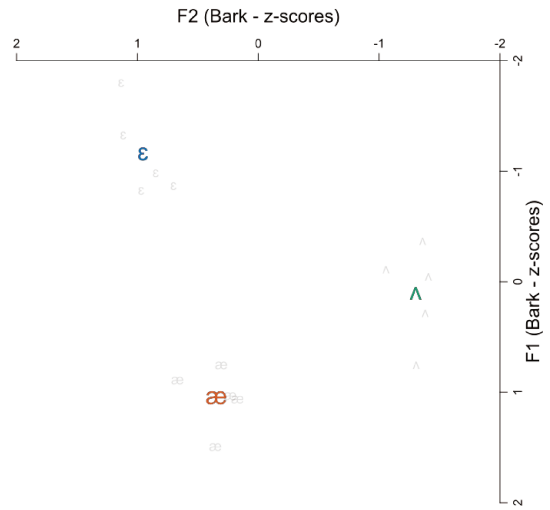

Figure 51 : Représentation des voyelles utilisées dans la tâche en EEG en Bark centrés réduits. Les petits symboles phonétiques représentent les voyelles, les gros symboles phonétiques représentent les barycentres des catégories.

5.2.2 Protocole expérimental

Les stimuli étaient présentés dans un paradigme *Oddball*, qui consiste à présenter une séquence de stimuli identiques (standards) entrecoupés de stimuli rares (déviants). Le nombre de stimuli standards consécutifs séparant les déviants variait de 3 à 6 afin de ne pas provoquer d'attente de la part du participant. Les stimuli étaient présentés avec un intervalle fixe de 1000 ms entre chaque début de stimulus (SOA). L'expérience était organisée en quatre blocs (deux blocs pour la paire *bat-bet*, deux pour la paire *bat-but*) de telle sorte que chaque mot soit standard dans un bloc et déviant dans l'autre. Chaque bloc comportait 500 stimuli standards et 100 stimuli déviants soit une proportion de 0.2 déviants. L'ordre des blocs était présenté dans un ordre aléatoire avec contraintes entre les participants et entre les séances (prétest/posttest) pour un même participant de telle sorte que les blocs contenant les mêmes stimuli soient toujours consécutifs. Chaque bloc commençait par une série de 10 standards non pris en compte dans les analyses. De courtes pauses étaient proposées entre chaque bloc.

Tableau XXXIII : Organisation de l'enregistrement EEG en 4 blocs

Bloc	Standards (de 3 à 6)	Déviants
1	bat	bet
2	bet	bat
3	bat	but
4	but	bat

Les enregistrements avaient lieu dans la salle anéchoïque de la Maison des Sciences de l'Homme à Lyon. Les sujets étaient placés devant un écran d'ordinateur diffusant un film sans son et sous-titré de leur choix. Il leur était demandé de ne pas prêter attention aux stimuli présentés auditivement à un niveau d'écoute confortable. Il a été vérifié que les participants ne portaient pas d'objet contenant du métal. Ils étaient équipés d'un bonnet comportant 32 électrodes actives (Electro-Cap International, INC. Ohio USA) selon le système international 10-20, ainsi que huit électrodes actives externes, placées sur le visage : 2 VEOG, 2 HEOG (électro-oculogrammes verticaux et horizontaux permettant d'enregistrer les clignements et mouvements des yeux), une sur le nez, une sur la joue et deux sur les mastoïdes droit et gauche. Il leur était demandé de bouger le moins possible et de minimiser les clignements des yeux pendant l'enregistrement afin d'éviter toute interférence avec le signal EEG ; les lentilles de contact augmentant le clignement des yeux, les sujets ont été encouragés à porter leurs lunettes. L'électrode *offset* était maintenu dans l'intervalle +/-20 mV. Un filtrage a été effectué de telle sorte que la bande passante du signal numérique enregistré soit de 0.1 à 100 Hz. La fréquence

d'échantillonnage était de 512 Hz avec une résolution d'amplitude de 24 bits. Le logiciel Presentation®[12] a servi à programmer l'expérience ; l'acquisition des données EEG a été effectuée avec le logiciel d'acquisition ActiView de Biosemi.

5.2.3 Analyses des données

Les données EEG ont été analysées avec MATLAB version R2019a (The MathWorks, 2019) et les *toolbox* EEGLAB[13] (Delorme et Makeig, 2004) et ERPLAB[14] (Lopez-Calderon et Luck, 2014). Des périodes de -150 à 800 ms autour du début des stimuli ont été extraites des signaux continus. Pour les standards, nous n'avons conservé que les stimuli précédant un déviant. Ainsi nous avions la même proportion de standards et déviants dans les analyses. L'ensemble des déviants étaient conservés à ce stade. Après l'application d'une correction selon la ligne de base (-150 à 0 ms) et d'un filtre passe bas de 30 Hz (Pakarinen, Takegata, Rinne, Huotilainen et Näätänen, 2007), une décomposition en composantes indépendantes (ICA) a été réalisée dans EEGLAB avec l'algorithme `runica`. Les artefacts oculaires ont été identifiés à l'aide de la *toolbox* ADJUST (Mognon, Jovicich, Bruzzone et Buiatti, 2011)[15] parmi ces composantes et rejetés après inspection visuelle. Une détection des artefacts restants avec une fenêtre mobile allant de pic en pic (`Moving Peak to Peak Window`) d'une largeur de 400 ms avec un pas de 50 ms a été appliquée sur les 32 électrodes ainsi que les électrodes contenant les canaux oculaires avec un seuil de 75 µV. Les époques contenant des artefacts ont été retirées des analyses avant qu'elles soient moyennées par condition. Tous les participants avaient un nombre d'items acceptés supérieur à 50 dans toutes les conditions. Enfin, nous avons re-référencé le signal à la moyenne des mastoïdes pour maximiser le ratio signal sur bruit (Kujala, Tervaniemi et Schroger, 2007 ; Miglietta et al., 2013 ; Ylinen et al., 2006, 2010).

Comme tous nos stimuli étaient présentés en tant que standard dans un bloc et en tant que déviant dans un autre (Ylinen et al., 2006, 2010 ; Zevin et al., 2010), nous avons étudié la iMMN (MMN *identity*) pour les deux voyelles /ɛ/ et /ʌ/ : celle-ci est de le résultat de la

[12] Version 17.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com

[13] Version 14.1.2b

[14] Version 7.0.0

[15] Version 1.1.1

différence entre l'onde générée par la voyelle présentée en tant que standard soustrait à la celle de l'onde générée par la voyelle présentée en tant que déviant. Par exemple, la iMMN pour la voyelle /ɛ/ était calculée en faisant la différence de la moyenne des essais où *bet* était présenté déviant avec la moyenne des essais où *bet* était présenté comme standard. Afin de déterminer la présence d'une MMN pour chaque contraste, nous avons utilisé la *toolbox* MASS univariate ERP (Groppe, Urbach et Kutas, 2011). L'utilisation de cette *toolbox* présente les avantages de ne pas nécessiter de définir de fenêtre temporelle ou régions d'intérêt à priori. Elle peut mettre en évidence des effets non attendus quand des fenêtres temporelles ou région d'intérêt sont définies à priori. Elle tire pleinement parti de la résolution spatiale et temporelle de l'EEG et est utile pour fournir des limites inférieures aux débuts temporels des effets (par exemple, le moment à partir duquel une variable ou une manipulation affecte le traitement du stimulus).

Pour chaque groupe (N, C, PE et PR) et chaque test (prétest et posttest, pour les apprenants seulement), les ERP de ces conditions ont été soumis à des tests de permutation bilatéraux à mesures répétées basés sur la statistiques *tmax* (Blair et Karniski, 1993) en utilisant un niveau alpha de .05. Tous les points temporels entre 0 et 800 ms sur les électrodes fronto-centrales (F3, Fz, F4, FC1, FC2, C3, Cz et C4) ont été inclus dans les tests (3280 comparaisons au total). Des tests de Student pour mesures répétées ont été réalisés pour chaque comparaison en utilisant les données originales et 2500 permutations aléatoires intra-participant. Le score *t* le plus extrême pour chaque série de tests (i.e., le « *tmax* » de chaque série de test) a été enregistré et utilisé pour estimer le *tmax* de la distribution de l'hypothèse nulle (pas de différence entre les conditions). Sur la base de cette estimation, des t-critiques ont été dérivés pour chaque voyelle et toute différence dans les données d'origine excédant cette valeur a été jugée comme fiable. Ce test de permutation a été utilisé à la place d'une ANOVA conventionnelle sur l'amplitude moyenne car il permet une meilleure résolution spatiale et temporelle en maintenant un niveau alpha souhaitable (i.e., correct pour les grands nombres de comparaisons). De plus, la statistique *tmax* a été choisie pour les tests de permutation car il a été montré qu'elle présentait un niveau de puissance intéressant pour les données dont les dimensions sont fortement corrélées (comme les *ERP*) (Hemmelmann et al., 2004). 2500 permutations ont été utilisées pour estimer la distribution de l'hypothèse soit un nombre deux fois supérieur au nombre recommandé par Manly (2018) pour un niveau alpha de 0.05. Les tracés des potentiels évoqués seront présentés sur l'électrode Fz dans les figures qui suivent.

5.3 Résultats

Nous avons d'abord analysé les données des locuteurs natifs pour nous assurer qu'une MMN est bien présente pour les deux contrastes proposés. Les tracés moyens des potentiels évoqués obtenus pour les seize locuteurs natifs pour chacune des voyelles /ɛ/ et /ʌ/ sont présentés dans les figures Figure 52 :

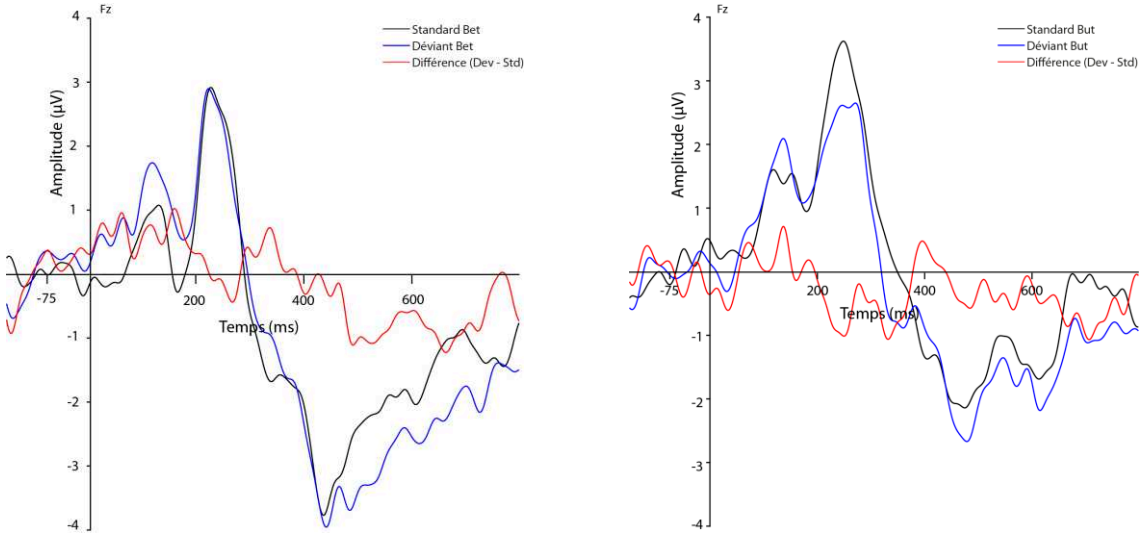

Figure 52 : (gauche) Grandes moyennes des signaux générés par les stimuli présentés en tant que standards (noir), déviants (bleu) pour le mot *bet* à gauche et *but* à droite chez les 16 locuteurs du groupe natif (Groupe N). Le tracé rouge représente la différence entre les réponses à un même stimulus présenté en tant que déviant dans un bloc et standard dans un autre bloc. Ce tracé est censé faire apparaitre la iMMN

Pour les deux contrastes, de manière inattendue, nous n'observons pas de MMN. En effet, visuellement, sur le tracé moyenné sur l'ensemble de sujets du groupe, nous nous attendions à une négativité plus importante de l'onde générée par les déviants (en bleu) entre 150 et 200 ms. Et cette négativité devrait se traduire sur la MMN (en rouge) par une négativité également comme sur la Figure 53 reproduite de l'étude de Zevin et collègues (2010) pour le contraste /r/-/l/ chez des locuteurs anglais natifs.

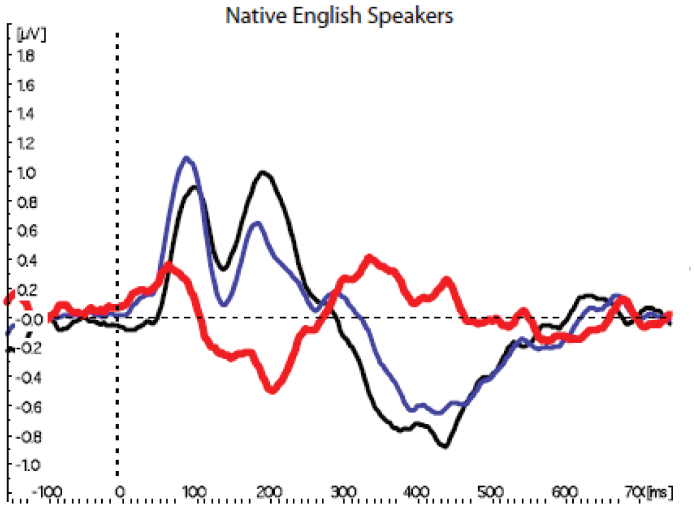

Figure 53 : Onde iMMN pour le contraste /r/-/l/ sur l'électrode Fz dans l'étude de Zevin et al. (2010) pour les apprenants japonais. La condition standard est représentée en noir, les déviants en bleu et la différence entre déviants et standards en rouge

Des *t*-critiques de +/- 4.48 (tous les $p > .25$) pour la voyelle /ɛ/ et de +/- 4.53 (tous les $p > .95$) (*ddl* = 15) pour la voyelle /ʌ/ ont été dérivés. Les tests de permutation réalisés pour ce groupe pour les deux voyelles confirment l'absence de différences significatives entre les ERP des stimuli standards et déviants. Les mêmes analyses menées sur les données des autres groupes (C, PE et PR) en prétest et posttest montrent la même absence de MMN pour les deux contrastes étudiés.

Le groupe N constituant en quelque sorte notre « groupe contrôle », en tout cas celui pour lequel une MMN doit être présente pour les deux contrastes, nous avons tenté plusieurs autres types d'analyses sur les données de ce groupe.

1. La iMMN étant calculée sur moins d'items que la MMN classique, qui elle est calculée en faisant la différence entre la moyenne des déviants et standards des deux blocs par conditions (e.g., MMN/æ-ʌ/ = moyenne(*but* et *bat* comme déviants) – moyenne(*but* et *bat* comme standards), nous avons également calculé cette dernière pour le groupe Natif. Les tracés moyens des potentiels évoqués et de la différence sont présentés dans la Figure 54. De nouveau, la MMN n'est pas significativement différente de 0 sur toutes les électrodes testées et tous les points temporels et ce pour les deux contrastes testés (/æ-ɛ/ et /æ-ʌ/) dans tous les groupes. Les *t* critiques sont de +/- 4.51 (tous les $p > .59$) pour /æ-ɛ/ et +/-4.89 pour /æ-ʌ/ (tous les $p > .17$).

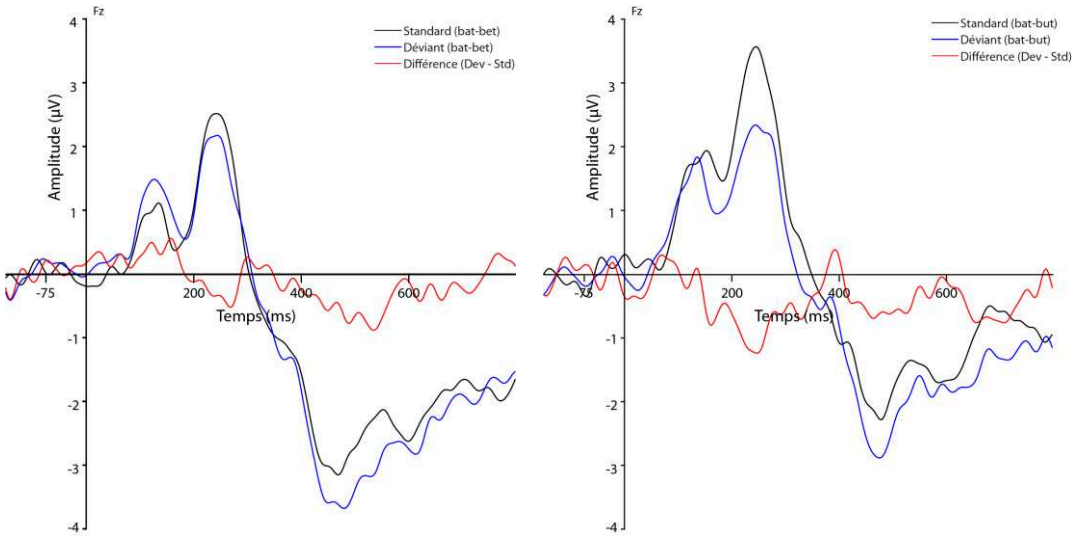

Figure 54 : Signaux moyennés sur les 16 locuteurs du groupe N. Le tracé noir représente la moyenne des stimuli standards d'une condition, le tracé bleu la moyenne des stimuli déviants. Le tracé rouge est calculé comme la différence entre la moyenne des stimuli déviants d'une condition (/ɛ/ ou /ʌ/) et celle des stimuli standards

2. Nous avons tenté de refaire les analyses en gardant davantage de stimuli standards (voir Figure 55). Ainsi nous avons refait les analyses en n'enlevant que les standards immédiatement précédés d'un déviant (e.g., Miglietta et al., 2013 ; Tamminen et al., 2015). Cette nouvelle analyse comptait donc 300 standards et 100 déviants par condition et par participant, contrairement à 100 standards et 100 déviants dans l'analyse initiale.

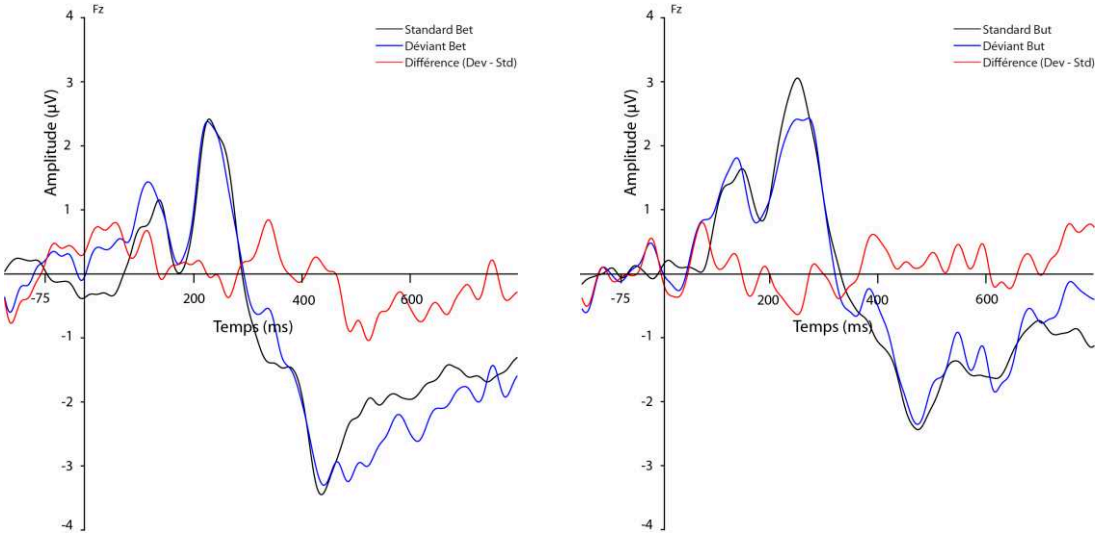

Figure 55 : Signaux standards (noir) et déviants (bleu) moyennés sur les 16 locuteurs natifs et la différence censée faire émerger la iMMN (en rouge) pour les analyses contenant 100 déviants et 300 standards par blocs

Cette analyse ne met pas non plus en évidence de MMN significativement différente de 0 sur les électrodes fronto-centrales pour le groupe de locuteurs natifs pour les deux

contrastes. Les t critiques sont de +/- 4.42 pour /ɛ/ (tous les p >.70) et de +/- 4.56 (p > .54) pour /ʌ/.

3. À la relecture des scripts ayant été utilisés pour ces enregistrements, il s'avère que plusieurs erreurs ont été commises. Lors des enregistrements EEG, les stimuli sont envoyé automatiquement par un logiciel de stimulation et envoie en parallèle un marqueur (appelé *trigger*) afin d'indiquer sur le signal le moment où le stimulus a été présenté. La précision temporelle de ce *trigger* est très importante pour l'analyse des potentiels évoqués. Par définition, la MMN est généré lorsqu'une déviation est perçue dans une séquence de stimuli répétés (standards) du fait de la présentation d'un stimulus déviant. Dans nos stimuli, la déviation entre standard (e.g., *bat*) et déviant (e.g., *but*) intervient au moment de la présentation de la voyelle. Il est donc nécessaire que les *triggers* soient placés au début des voyelles. Notre première erreur de codage tient au fait que les *triggers* placés sur le signal EEG n'ont pas été placés de manière adéquate. En effet, ils ont été placés au début des fichiers sons, or les stimuli présentés étant constitués de cinq occurrences différentes par voyelle, et la voyelle n'apparait pas exactement au même moment dans le fichier son comme le montre le Tableau XXXII. En effet, le délai d'apparition de la voyelle peut varier de 30 ms d'un item à l'autre. De plus, les différentes occurrences d'une même voyelle n'ont pas été distinguées par des *triggers* différents mais rassemblées sous un même code, ce qui nous empêche de réaligner le signal à postériori. Enfin, la randomisation des stimuli a été faite de manière automatique et différente pour chaque sujet et les traces des ordres de passage résultants (fichiers *log*) n'ont pas été conservées, ce qui rend là encore impossible un réalignement des signaux à postériori.

   Dans l'hypothèse où l'absence de MMN serait liée à la variabilité des stimuli en terme de délai d'apparition de la voyelle, nous avons tenté de re-synchroniser les signaux EEG sur le début des voyelles. Pour cela, à partir des signaux segmentés avant moyennage, pour chaque bloc (un bloc contenant 5 versions différentes du standard et 5 versions différentes du déviant), sur les électrodes fronto-centrales (F3, F4, Fz, FC1, FC2, C3, C4 et Cz), nous avons essayé de prédire pour chaque essai le stimulus utilisé le plus probable parmi les 5 sur la base d'hypothèses faites à partir des délais d'apparition des voyelles dans les différents stimuli (voir Tableau XXXII). Plus précisément, nous avons retenu comme hypothèse la combinaison de stimuli qui

minimisait la distance entre signaux EEG d'une même condition dans un même bloc. Cette minimisation des distances peut être vue comme une hypothèse optimale susceptible de faire apparaître la MMN attendue. Nous avons ainsi recalé le temps zéro de chaque stimulus et refait la fin des analyses comme décrit précédemment[16]. Les potentiels évoqués issus de cette analyse sont présentés dans la Figure 56.

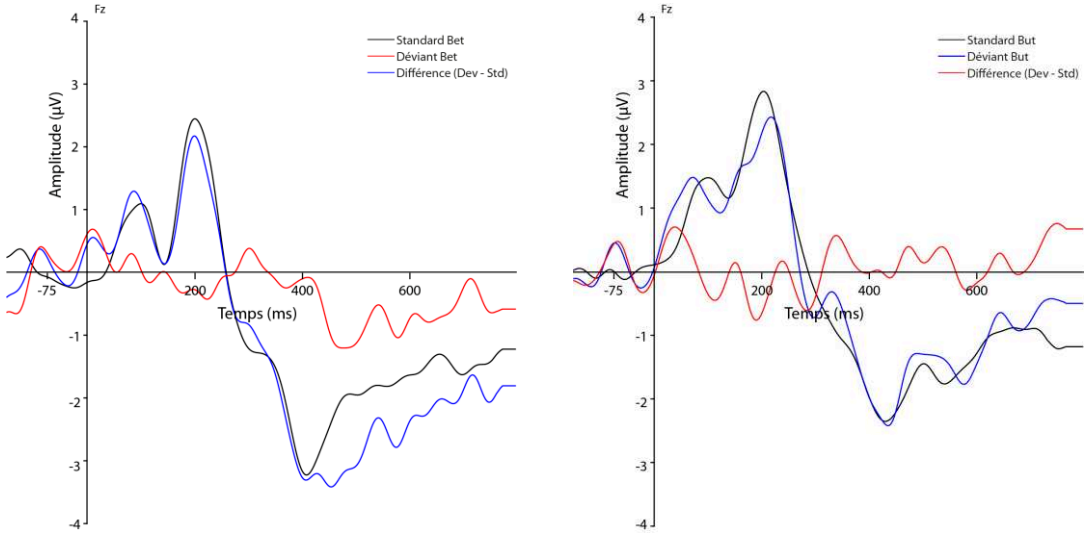

Figure 56 : (gauche) Grandes moyennes des signaux générés par les stimuli présentés en tant que standards (noir) et déviants (bleu) pour le mot *bet* à gauche et *but* à droite chez les 16 locuteurs du groupe natif (Groupe N) après recalage du signal. Le tracé rouge représente la différence entre les réponses à un même stimulus présenté en tant que déviant dans un bloc et standard dans un autre bloc

Cette analyse n'a pas permis non plus de mettre en évidence de MMN pour le groupe de participants natifs pour les deux contrastes sur toutes les électrodes testées. Les *t* critiques sont de +/- 4.19 pour /ɛ/ (tous les $p > .58$) et +/-5.53 pour /ʌ/ (tous les $p > .47$).

5.4 Discussion des résultats en EEG

L'objectif de l'approche présentée dans cette partie était de mettre en évidence des corrélats EEG de l'acquisition de contrastes phonologiques consécutifs à la participation à nos entraînements.

[16] Ce type de recalage temporel s'apparente à une méthode connue sous le nom de *Woody filter* (Luck, 2014) qui est implémentée dans l'algorithme RIDE (Ouyang, Sommer et Zhou, 2016)

Nous avons proposé une tâche EEG dont le but était de tester la sensibilité de nos participants francophones aux contrastes anglais /æ-ɛ/ et /æ-ʌ/. Nous pensions que les apprenants francophones censés présenter plus de difficultés à discriminer le contraste /æ-ʌ/, absent en français, devraient, d'une part présenter une MMN pour ce contraste de plus petite amplitude que pour le contraste /æ-ɛ/ qui est présent en français, ou qui en tout cas est proche du contraste français /a-ɛ/, d'autre part, présenter des MMN pour les deux contrastes plus petites que celles observées dans le groupe de locuteurs anglais natifs (Groupe N). Suite aux entraînements, nous nous attendions à observer un changement dans les profils des MMN.

Les résultats obtenus suite aux enregistrements EEG ne sont pas ceux attendus puisque dans aucun des groupes et pour aucun des deux contrastes étudiés nous n'avons observé de MMN significativement différente de 0, et ce que l'on étudie la MMN classique ou la iMMN. Nous avions choisi l'approche de la iMMN car selon (Pulvermüller et Shtyrov, 2006) celle-ci présente l'avantage d'éliminer les composantes liés aux différentes acoustiques pures dans la morphologie des ERP. Suite à l'étude des scripts utilisés pour programmer l'expérience, nous nous sommes rendu-compte que les marqueurs sur le signal EEG n'avaient pas été placés comme ils doivent l'être pour ce type de protocole. Du fait que nous avions inséré de la variabilité dans nos stimuli et n'avions pas conservé les traces de cette variabilité, il nous a été impossible, malgré une tentative d'optimisation, de re-synchroniser les signaux. Nous supposons également que la trop grande proportion de stimuli déviants (0.2) et le trop faible nombre de stimuli par blocs (500) est venu atténuer davantage le ratio signal sur bruit et affaiblissant ainsi encore l'effet attendu. Nous avions choisi de réduire le plus possible la durée des enregistrements EEG par soucis de confort pour nos participants qui réalisaient les tâches comportementales avant l'enregistrement. Cela dit, à titre de comparaison, dans l'étude en MMN menée par Zevin et collaborateurs (2010), les auteurs avaient utilisés 1080 stimuli par blocs avec une probabilité d'apparitions des déviants de 0.17 relativement similaire à celle de notre étude. Par contre, leur protocole n'incluait qu'une seule occurrence par stimulus réduisant ainsi la variabilité liée aux stimuli. Quant aux nombre de participants dans nos groupes, bien qu'il ne soit pas aussi élevé que dans les études comportementales, il se situe dans la fourchette haute des études en potentiels évoqués et a fortiori dans celles testant l'effet d'entraînements jusqu'à récemment (e.g., Kraus et al., 1995 ; Lu et al., 2015 ; Reinke et al., 2003 ; Ylinen et al., 2010 ; Zevin et al., 2010). La plupart des études comptent entre 12 et 15 participants dans leurs groupes d'apprenants. Si le nombre de participants ne semblent pas en cause, trois sources de difficultés ont probablement contribué à cette absence de MMN : la variabilité acoustique

inter-stimuli, le nombre peut-être insuffisant d'essais et les sources d'incertitudes introduites dans la méthodologie même.

Si nous devions refaire cette partie de l'étude, il serait d'abord utile de tester l'expérience avec seulement deux groupes de participants : un groupe de participants anglophones et un groupe de participants francophones. Ainsi, nous commencerions par tester nos hypothèses de départ selon lesquels nous devrions observer une différence entre les profils des ERP des natifs et non-natifs selon le type de contraste. Nous testerions toujours la perception des deux contrastes anglais /æ-ʌ/ et /æ-ɛ/ pour nos deux groupes de participants. Il serait tentant de supprimer la variabilité que nous avions insérée dans les stimuli mais nous souhaiterions plutôt la conserver en nous assurant de bien coder le début des voyelles. Il semble cependant que la question de la présence d'une MMN robuste à la variabilité acoustique inter-stimuli est encore ouverte et que l'interprétation des études basées sur des stimuli de synthèse (e.g., Díaz, Mitterer, Broersma, Escera et Sebastian-Galles, 2016) ou sur des stimuli naturels sans variabilité doit rester prudente. Nous effectuerons en revanche des modifications au niveau du nombre de stimuli par condition et de la fréquence d'apparition des stimuli déviants dans les blocs afin de nous rapprocher des quantités et proportions utilisées dans Ylinen et collaborateurs (2010). Dans cette étude, ils testaient un entraînement incitant des locuteurs finnois à utiliser le timbre plutôt que la durée pour discriminer les voyelles anglaises /ɪ-iː/, les auteurs ont proposé un paradigme *oddball* dans lequel chaque voyelle était présenté plus de 1100 fois comme standard ($p = .85$) et presque 200 fois comme déviant ($p = .15$). Ainsi, nous devrions doubler le nombre de nos stimuli et il est certain que cela augmenterait le rapport signal sur bruit. Cependant, cela augmenterait considérablement le temps d'enregistrement d'un peu plus d'une demi-heure à plus d'une heure. Le protocole comprenant en plus les tâches comportementales, nous devrions alors plutôt nous orienter vers un protocole de type *roving standard* dans lequel des séries de stimuli répétés s'enchainent, de telle sorte que dans un même bloc, chaque premier stimulus d'une série constitue le déviant de la série précédente dont les stimuli constituait des standards (Garrido et al., 2008; Stefanics, Kremláček et Czigler, 2014). Ainsi, nous pourrions réduire un peu la durée de l'enregistrement.

Partie 3. Conclusion générale

1 Résumé des résultats

L'objectif de ce travail de thèse était de tester l'effet d'un entraînement permettant à des apprenants francophones d'améliorer à la fois la perception et la production de monophtongues de l'anglais, les voyelles /æ, ʌ, ɑː, ɪ et iː/. À notre connaissance, une seule étude (Iverson et al., 2012) a mis en place un protocole de ce type avec des apprenants francophones, et l'entraînement proposé, de type *High Variability Phonetic Training*, bien qu'il ait déjà prouvé son efficacité dans la littérature, n'a pas permis d'améliorer les voyelles ciblées dans notre étude, notamment en production. Ce type d'entraînement repose sur l'existence d'une relation étroite entre perception et production de la parole telle que la décrivent les modèles d'acquisition de la L2 les plus connus (Best & Tyler, 2007 ; Flege, 1995). Selon ces modèles, les difficultés de production de contrastes étrangers sont le fait de difficultés de perception. Ainsi, ces entraînements reposent essentiellement sur des tâches de perception avec l'idée selon laquelle améliorer la perception de contrastes étrangers améliorerait la précision avec laquelle ceux-ci sont produits. Nous avons fait l'hypothèse qu'un entraînement davantage orienté vers la production mais impliquant également la perception pourraient être bénéfique pour aider des apprenants sur ces contrastes vocaliques. Quelques travaux ont montré des effets bénéfiques d'entraînements basés sur de simples tâches d'écoute et répétition pour l'acquisition de contrastes étrangers (Saloranta et al., 2020 ; Tamminen et al., 2015). Et d'autres ont montré l'efficacité d'un *feedback* s'appuyant sur l'analyse des paramètres acoustiques des voyelles produites (Carey, 2015 ; Kartushina et al., 2015).

L'entraînement que nous avons développé (reçu par le groupe PR) est composé d'une tâche de répétition de mots dont nous avons fait l'hypothèse qu'elle pourrait mettre en jeu le *feedback auditif* décrit dans le modèle de l'acquisition de la parole de Guenther (2011), permettant ainsi d'affiner les cibles articulatoires des voyelles par l'ajustement des commandes d'anticipation motrices suite aux retours sensorimoteurs. De plus, cet entraînement proposait un *feedback* visuel permettant à l'apprenant d'évaluer sa production et ainsi de l'ajuster pour se rapprocher d'une cible native. Parallèlement à cet entraînement, nous avons mis au point un entraînement classique de type *HVPT* (groupe PE) afin d'en comparer les effets. Et un groupe contrôle

(Groupe C) ne suivant pas d'entraînement a également été testé afin de nous assurer que les changements observés dans les groupes entrainés étaient bien le fait des entraînements proposés. Les participants des trois groupes ont été évalués dans des tâches de perception (identification et discrimination) et également en production par l'intermédiaire d'une tâche de lecture de mots isolés. Dans chaque tâche, les performances des apprenants francophones étaient comparées à celles de locuteurs anglais natifs (groupe N) et les résultats ont montré qu'aucun des groupes n'atteignaient des performances similaires aux natifs après entraînement.

L'évaluation des productions a été réalisée de deux manières : d'une part par l'analyse des paramètres acoustiques des voyelles que nous avons comparés à ceux des voyelles produites par des locuteurs natifs, en particulier par le biais d'une analyse discriminante linéaire (*LDA*) ; d'autre part, par l'évaluation des voyelles dans une tâche d'identification réalisée par des juges natifs. Une synthèse générale des résultats obtenus quant à l'effet des entraînements en fonction des différentes tâches d'évaluation est proposée dans le Tableau XXXIV.

Tableau XXXIV : Tableau récapitulatif des effets des entraînements dans les différentes tâches. Le signe égal (=) indique qu'aucune différence significative n'est observée entre prétest et posttest. Le signe (+) indique qu'une amélioration significative des performances est observées entre prétest et posttest.

	Groupe C	**Groupe PE**	**Groupe PR**
Identification	=	+	+
Discrimination	=	+	=
Production (évaluation par *LDA*)	+	+	+
Production (identification par juges natifs)	=	+	+

De manière générale, l'entraînement de type *HVPT* reçu par le groupe PE a été efficace pour améliorer les performances des participants sur les deux tâches de perception (identification et discrimination) mais également selon les deux modalités d'évaluation de leurs productions.

Dans les tâches de perception, les résultats ont montré que l'entraînement perceptif a été efficace pour améliorer à la fois les performances en identification et en discrimination. C'est-à-dire qu'après avoir suivi cet entraînement, les participants ont, d'une part, établi des catégories perceptives plus stables (ils sont moins sujets à des confusions phonémiques, comme l'atteste les scores en identification) et d'autre part, ils ont su extraire et intégrer les indices acoustiques pertinents pour la discrimination des contrastes vocaliques qui leur posaient davantage de problèmes avant entraînement. Nous n'observons pas de différences significatives entre voyelles en identification, ni entre contrastes en discrimination, ce qui suggère que l'entraînement a bénéficié à l'ensemble des voyelles/contrastes entrainés.

En ce qui concerne les résultats du groupe PE en production, les deux méthodes utilisées montrent une amélioration des performances entre prétest et posttest. Cependant, cette amélioration n'est significative par rapport au groupe contrôle que lorsque les productions sont évaluées par les locuteurs natifs, ce qui suggère que les participants de ce groupe font significativement moins d'erreurs phonémiques après entraînement comparativement aux participants du groupe C. Ce constat rappelle tout d'abord l'importance d'avoir un groupe contrôle dans ce type de recherche, ce qui n'est pas toujours le cas dans la littérature (voir Sakai & Moorman, 2018 ; Thomson & Derwing, 2015 pour des revues sur les entraînements perceptifs et les entraînements en production). Ensuite, cette distinction nous révèle la complémentarité des mesures utilisées. Selon Thomson et Derwing (2015), ces deux mesures distinguent deux types de performances : celles suivant le *principe d'intelligibilité* dont l'efficacité de l'entraînement est évaluée à partir de jugements réalisés par des locuteurs natifs ; celles suivant le *principe de « nativeness »*. Ainsi, l'entraînement proposé à ce groupe a permis d'augmenter de manière significative l'intelligibilité des participants mais n'a pas permis d'améliorer leur accent étranger. Le fait que cet entraînement perceptif a permis également une amélioration des performances en production est en accord avec les études ayant proposé des entraînements perceptifs pour des voyelles de L2 et ayant testé le transfert de l'apprentissage en production (Iverson et al., 2012 ; Lambacher et al., 2005 ; Rato & Rauber, 2015 ; Thomson, 2011, 2018 ; Thomson & Derwing, 2016). Cependant, contrairement à l'étude d'Iverson et collaborateurs avec des participants francophones, dont l'effet était restreint à certaines voyelles entrainées, dans notre étude, l'ensemble des voyelles et contrastes s'améliore de manière équivalente suite à l'entraînement. À la différence de cette étude, qui proposait un entraînement pour les 14 monophtongues et diphtongues de l'anglais, notre étude ne proposait d'entrainer que 5 voyelles. Le fait que nous ayons ciblé un ensemble restreint de voyelles a pu favoriser l'effet de l'entraînement sur celles-ci. Nishi & Kewley-Port (2007) avaient montré dans une étude comparant un entraînement portant sur l'ensemble des monophtongues de l'anglais et un autre ne portant que sur les voyelles difficiles pour des apprenants japonais que l'entraînement était plus efficace sur le sous-ensemble de voyelles mais ne permettait pas de montrer d'amélioration pour les voyelles non entrainées.

Les résultats sont plus contrastés chez les participants du groupe PR ayant reçu l'entraînement en production. En effet, en perception nous avons observé que ce groupe améliore ses performances significativement entre prétest et posttest comparativement au groupe contrôle seulement en identification et ce pour toutes les voyelles indifféremment les unes des autres.

Ces résultats laissent penser que l'entraînement a permis aux participants de ce groupe d'améliorer leurs représentations de catégories phonémiques. En revanche, dans la tâche de discrimination, ce groupe ne s'améliore pas plus que le groupe contrôle. Ce résultat indique que l'entraînement en production n'a pas permis aux participants de ce groupe de traiter avec plus d'efficacité les traits phonétiques pertinents pour la discrimination des voyelles adjacentes. Ce résultat est en contradiction avec ceux de Karthushina et al., (2015) qui avaient montré une amélioration des performances en perception des voyelles du danois par des apprenants français ayant suivi un entraînement similaire au nôtre par rapport à leur groupe contrôle. Ce résultat est d'autant plus étonnant que dans cette étude, le groupe contrôle réalisaient la même tâche que le groupe expérimental avec la différence qu'il ne recevait pas le *feedback*, ce qui montrait l'effet du *feedback* par rapport à la simple tâche d'écoute-répétition Ainsi, il était exposé pendant l'entraînement à la même quantité de matériel auditif que le groupe expérimental. Dans notre étude, le groupe contrôle n'était pas soumis au matériel des entraînements. Par ailleurs, une série d'études comparant des entraînements en perception seule versus des entraînements impliquant perception et production a mis en évidence que la production réalisée pendant l'entraînement pouvait altérer la formation de catégories phonologiques. Par exemple, Baese-Berk et Samuel (2016) ont entrainé des participants espagnols à un contraste fricative/affriquée du basque. Les participants étaient entrainés soit simplement par une tâche de discrimination avec *feedback* (perception), soit par une tâche d'écoute-répétition (perception plus production). Leurs résultats, répliquant ceux d'une précédente étude ayant entrainé des locuteurs anglais à un contraste de voisement de l'hindi (Baese-Berk, 2010), montrent que le groupe entrainé en perception et production n'améliore pas ses performances en discrimination. Les auteurs ont également montré qu'une expérience préalable avec le contraste à apprendre pouvait atténuer cet effet délétère de la production pendant l'entraînement. Ils ont également montré que le désavantage de l'entraînement en production était présent même si le son à produire n'était pas relié au contraste acquis (prononcer une lettre isolée). Ils ont ainsi conclu que cette altération de l'apprentissage perceptif était induite par l'implication du système de production de manière générale pendant l'entraînement perceptif, générant un conflit avec le système qui développe les nouvelles catégories perceptives. Cependant, Thorin et al., (2018) relativisent ces résultats en montrant que des entraînements en perception et production sont bénéfiques à la perception et la production de voyelles non natives. Mais cette dernière étude n'avait pas de groupe contrôle. Enfin en production, comme le groupe PE, ce groupe ne se distingue du groupe contrôle que lorsque les performances sont évaluées par des juges natifs. Ainsi cet entraînement permet

d'améliorer l'intelligibilité mais ne corrige pas l'accent étranger de manière significative. La littérature sur les effets des entraînements en production est assez contradictoire ; notre étude apporte ainsi de nouveaux éléments sur les voyelles, d'autant plus qu'elle comporte un groupe contrôle et un groupe de comparaison entrainé en perception.

2 Perspectives

Sans parler du protocole en EEG, qui n'a pas permis d'étudier les phénomènes que nous souhaitions, nous avons relevé des améliorations à apporter au protocole que nous avions suivi de manière générale.

Bien que les entraînements que nous avons proposés se soient montrés au moins en partie efficaces pour améliorer les performances en perception et en production, en tout cas telles que nous les avons mesurées, nous pouvons noter d'une part, comme relevé précédemment, qu'ils ne permettent pas aux participants d'atteindre un niveau natif. D'autre part, il est peu probable, vu les faibles améliorations observées, qu'un transfert vers des tâches plus écologiques soit constaté et a fortiori que les apprenants généralisent leur apprentissage au lexique qu'ils ont d'ores et déjà appris. Voici un exemple, un peu éloigné certes mais qui illustre qu'il est possible d'apprendre un contraste « non natif » sans le généraliser au lexique déjà appris. Dans une étude mené par Dufour et collaborateurs (2010), suite à un entraînement au contraste français /e/-/ɛ/ par des locuteurs du français méridional dont le contraste n'est pas présent dans l'inventaire phonologique, les participants étaient capables de discriminer des paires minimales de nouveaux mots (/ʒurde/ vs. /ʒurdɛ/) mais ils n'étaient pas capables d'utiliser cet apprentissage pour distinguer des mots qu'ils connaissaient déjà (/epe/ vs. /epɛ/). De plus, les mesures de production que nous avons faites sont artificielles puisque nous utilisons une tâche de lecture de mots isolés. Il serait nécessaire de tester également la production de parole « spontanée » ou « quasi-spontanée » afin d'évaluer le bénéfice de l'entraînement en condition naturelle. Dans une étude récente, Lengeris (2018) a utilisé un paradigme d'élicitation très intéressant afin d'évaluer l'amélioration des performances en production d'apprenants grecs suite à un entraînement de type *HVPT* pour des voyelles de l'anglais britannique. Il a utilisé une sorte de jeu des différences dans lequel les participants devaient décrire les différences entre deux images. Les différences présentes entre les images permettaient d'éliciter les voyelles d'intérêt dans l'étude. Les participants enregistraient également des mots /bVt/ contenus dans des phrases porteuses. Les productions étaient ensuite évaluées par une tâche

d'identification réalisée par des locuteurs natifs de l'anglais britannique, et par des mesures de distances euclidiennes entre voyelles adjacentes. Leurs résultats montrent que l'amélioration des performances en production des voyelles est non seulement visible en lecture de phrases mais également en parole spontanée, et ce pour les deux types de mesures effectuées, avec cependant une amélioration plus importante sur les phrases lues.

Notre protocole présente l'avantage de comporter un groupe contrôle ; nous sommes cependant conscients que la méthode d'inclusion des participants dans les groupes mérite d'être améliorée. Comme dans l'étude de Kartushina et al., (2015), nous avons été confrontés au fait que pour l'une des tâches dans lesquelles ils étaient évalués (en production), les participants du groupe contrôle présentaient des performances plus élevées que les groupes expérimentaux. Ainsi, nous ne sommes pas sûrs que le fait que ce groupe contrôle n'améliore pas ses performances entre prétest et posttest comparativement aux autres groupes soit dû au fait qu'il n'a pas suivi d'entraînement ou plutôt d'une sorte d'effet plafond. Nous voyons deux explications possibles à cette différence : la première est le malheureux hasard de l'échantillonnage. La deuxième tient à la méthode de recrutement. Pour des raisons de contenus des entraînements différents et du fait que la participation au groupe contrôle ne nécessitait de venir au laboratoire que 2 fois au lieu de 7 dans les autres groupes, les participants du groupe contrôle étaient recrutés par l'intermédiaire d'une annonce différente de celle des autres groupes, qui précisait la durée de la participation ainsi que le défraiement reçu. Les participants de ce groupe recevaient 40€ plutôt que 100€ dans les autres groupes. Même s'il s'agit déjà d'une somme conséquente pour des étudiants, il est possible que les participants de ce groupe aient davantage été motivés par le sujet de l'étude que par l'attrait de l'argent. Des études ont en effet montré que différents dédommagements, en termes de valeur ou de type (argent ou crédits de cours), peuvent attirer différents types de personnes à participer à des études (Hsieh & Kocielnik, 2016 ; Sharp et al., 2006). Ainsi, il a pu s'en suivre, soit que ces participants s'appliquent davantage dans les tâches réalisées, soit qu'ils présentent un intérêt plus important pour l'apprentissage de l'anglais, expliquant leurs meilleures compétences de départ. Quoi qu'il en soit, bien que la mise en place d'un tel protocole soit compliquée, en particulier parce qu'il comporte un grand risque que les participants ne le poursuivent pas jusqu'au bout, il serait préférable de constituer les groupes sur la base d'une première évaluation des participants afin de les répartir de manière équilibrée. De plus, comme nous l'avons expliqué dans la description de la méthodologie, les participants de notre groupe contrôle devaient écouter des *audiobooks* entre le prétest et le posttest car nous n'avions pas pu mettre en place pour eux une tâche

alternative non reliée aux entraînements donnés dans les autres groupes. Ainsi, il est possible que cette tâche ait impacté les performances évaluées.

L'entraînement de type *HVPT* que nous avons proposé s'est montré efficace ; cependant nous pensons qu'il pourrait être amélioré, notamment en sélectionnant des stimuli qui ne présentent aucune ambiguïté pour des locuteurs natifs. Nous avons en effet observé, sur les stimuli que nous avons employés dans les tests, que les participants du groupe N présentaient un taux d'identification et de discrimination qui ne plafonnaient pas, contrairement à nos attentes, en particulier pour le contraste /ʌ - ɑː/. Il est probable, mais nous ne l'avons pas testé, que les stimuli de nos entraînements conduisent également à des confusions y compris chez des natifs pour ce contraste. Nous pensons qu'il est préférable de s'assurer que le matériel constituant le support d'entraînement ne présente pas ce type d'ambiguïté tout en conservant le niveau de variabilité nécessaire à la création de représentations phonémiques adaptées. Ainsi, il serait nécessaire de soumettre les stimuli expérimentaux à un groupe de locuteurs natifs pour une pré-évaluation et une pré-sélection des stimuli les plus à même de favoriser l'établissement de catégories phonétiques bien définies.

Bien que l'entraînement en production se soit montré relativement efficace pour améliorer les productions des participants l'ayant suivi, nous pensons que cet entraînement souffre de plusieurs défauts. Tout d'abord, il serait souhaitable que, à la manière de celui développé par Kartushina et al., (2015), une partie didactique expliquant explicitement les implications articulatoires derrières les représentations graphiques proposées en *feedback* aux apprenants soit ajoutée. En effet, plusieurs études ayant testé les effets d'entraînements informatisés explicites et implicites pour l'acquisition de certains traits syntaxiques ont montré que l'apprentissage explicite conduisait à des effets plus importants que les apprentissages implicites (Spada & Tomita, 2010 pour une revue de questions). Cependant, des études en potentiels évoqués ont montré qu'un apprentissage implicite conduisait à des traitements cognitifs davantage similaires à ceux réalisés par des locuteurs natifs (Morgan-Short, Sanz, Steinhauer et Ullman, 2010; Morgan-Short, Steinhauer, Sanz et Ullman, 2012). En termes de *feedback*, trois indices étaient donnés ; l'un reposait sur le timbre des voyelles, l'autre sur la durée et le troisième donnait le résultat d'une classification automatique de la voyelle produite selon un modèle natif. Concernant l'indice sur le timbre de la voyelle, il consistait en la représentation de la voyelle produite et la cible native dans le plan des deux premiers formants. Il était simplement expliqué au participant que l'axe vertical était lié au degré d'aperture des

voyelles, les voyelles plus ouvertes comme par exemple le /a/ étant en bas, et que l'axe horizontal était lié à la position de la langue dans la cavité buccale sur l'axe antéro-postérieur, l'avant étant à gauche. Nous les laissions ensuite expérimenter par eux-mêmes le comportement du programme avec l'appui du résultat de la classification automatique. Dans une revue de question sur les *feedback* articulatoires donnés dans ce type d'entraînement, Bliss et al., (2018) précisent que pour qu'un *feedback* soit efficace, il doit présenter les caractéristiques suivantes : 1) être naturel et logique, 2) compréhensible, 3) immédiat, 4) permettre d'effectuer des comparaisons, 5) flexible, 6) enrichissant, c'est-à-dire qu'il doit proposer quelque chose de plus qu'un humain ne peut pas apporter et enfin 7) abordable. Nous pensons que le *feedback* que nous avons proposé remplissait au moins en partie ces critères. Le critère de la compréhensibilité pourrait être amélioré comme précisé plus haut en ajoutant un entraînement à l'utilisation de l'outil sur des voyelles du français par exemple. De plus, une manière d'évaluer ce critère serait d'analyser les journaux d'événements (fichiers .log) de chaque session. Ceux-ci contiennent les valeurs médianes des trois premiers formants et les durées des voyelles produites au cours des sessions d'entraînement. Nous n'avons pas conservé les enregistrements, mais ces seules valeurs permettraient de suivre l'évolution des voyelles produites d'une part au cours d'une session, d'autre part entre les sessions, notamment entre les sessions 4 et 5 qui portaient sur l'ensemble des contrastes.

Le critère de la possibilité d'effectuer des comparaisons est en partie rempli car nous proposions des cibles natives en comparaison des voyelles produites. Cependant, il aurait été intéressant de rajouter la possibilité de réécouter cible et voyelle produites au regard des indices visuels présentés afin d'expliciter davantage les différences produites par rapport à la cible native. Aussi, il pourrait être utile lors des essais qui sont rejoués de garder la trace de l'essai précédent afin que le participant se rende compte visuellement des effets engendrés par la variation qu'il a produite entre les deux tentatives. Ce type de *feedback* respecte en partie le critère de flexibilité puisqu'il peut très simplement être appliqué à d'autres voyelles. Enfin, le critère « d'abordabilité » est plutôt respecté (en tout cas financièrement) car il est programmable sur un logiciel gratuit tel que Praat. Néanmoins, comme il propose une comparaison des voyelles produites avec celles de locuteurs natifs, sa mise en place nécessite l'enregistrement et l'extraction des paramètres acoustiques d'un corpus de voyelles natives.

En ce qui concerne les éléments donnés en *feedback*, nous avions proposé de manière plutôt arbitraire une limite de 200 Hertz que la voyelle produite ne devait pas dépasser par rapport à

la cible en plus de la catégorisation à l'analyse discriminante pour que la voyelle soit jugée comme correcte. Il serait probablement mieux de préciser cette limite à partir de la dispersion réelle des voyelles des locuteurs natifs afin qu'elle suive de manière plus ajustée les contours des régions acoustiques occupées par chacune des voyelles. Toujours au sujet du *feedback* que nous avons proposé aux participants recevant l'entraînement en production et comparativement à l'étude menée par Kartushina et al., (2015), nous pensons que la différence en termes d'efficacité des entraînements peut trouver une autre explication. Les contrastes vocaliques appris par les apprenants francophones dans cette étude ne s'opposaient pas sur la durée. Ainsi le *feedback* qu'ils recevaient ne portait que sur les deux premiers formants. De plus, l'information de type Correct/Incorrect n'était pas communiquée, contrairement à ce que nous avons fait à partir de la classification *LDA* réalisée en temps réel. Pour tout cela, nous pensons que les participants de notre étude, qui recevaient trois informations en *feedback*, n'ont peut-être pas utilisé l'ensemble de ces informations pendant les entraînements. Par exemple, s'ils n'ont pas réussi à comprendre le *feedback* donné sur le timbre de la voyelle, on peut imaginer qu'ils n'aient utilisé que le *feedback* de type correct/incorrect rendant ainsi l'entraînement moins efficace.

En ce qui concerne l'évaluation des performances, nous n'avons utilisé qu'un sous-ensemble des systèmes vocaliques de l'anglais et du français et n'avons ainsi qu'une représentation partielle de ceux-ci. Afin de comprendre de quelle manière les systèmes phonologiques cohabitent en fonction de l'amélioration de la précision de catégories acquises en L2, il serait nécessaire de pouvoir avoir un état de l'ensemble des voyelles de chacun des deux systèmes, avant et après entraînement, même si l'entraînement ne porte que sur certaines voyelles. Nous pensons également qu'il serait intéressant, ne serait-ce qu'avec les données dont nous disposons, i.e., les voyelles anglaises /æ, ʌ, ɑː, ɪ et iː/ produites par des locuteurs de L1 et L2 et les voyelles françaises /a, œ, ɛ, e, i/, de classifier les voyelles produites en L2, sur un modèle appris sur l'ensemble des voyelles de L1, un peu comme nous l'avons fait dans l'étude du /iː/. Nous pensons qu'une telle classification nous permettrait de voir dans quelle mesure les voyelles produites le sont dans un espace phonologique spécifique à la L2 ou si le système de L1 est préférentiellement utilisé. De la même manière, il serait intéressant de pouvoir prendre en compte l'inventaire phonologique de L1 des participants dans les *feedbacks* donnés aux participants. Detey et al. (2018) mentionnent à ce sujet la difficulté d'établir une référence à laquelle comparer les productions des apprenants mais également la nécessité de prendre en compte le système phonologique de l'apprenant dans les méthodes automatiques

d'apprentissage de la prononciation L2. Selon les auteurs, deux types d'approches sont ainsi possibles pour l'élaboration d'un *feedback* dans les systèmes actuellement disponibles : certains reposent sur la détection d'erreurs à partir de corpus d'apprenants annotés par des experts, d'autres (comme le nôtre) reposent sur la comparaison des voyelles des apprenants avec des cibles natives. Pour cela, nous pensons que la recherche dans le domaine gagnerait à davantage s'appuyer non seulement sur des corpus de locuteurs natifs mais également de locuteurs L2. Ce type de corpus, comme le corpus ANGLISH (Tortel, 2008), s'ils étaient suffisamment riches en termes de contenu et d'annotations devraient présenter l'avantage de permettre l'étude de phénomènes segmentaux mais également supra-segmentaux. Herment (2018) a par exemple montré que les outils d'annotation et de visualisation du signal pouvaient être utilisés dans l'enseignement L2 afin de permettre aux apprenants de comparer leurs productions à celles de locuteurs de L1. Ce type de corpus pourraient être utilisé à la fois pour l'étude à grande échelle des particularités de la production en langue seconde, en combinant approche de corpus et approche expérimentale comme le suggère Meunier et Littre (2013), mais également servir à la mise au point de systèmes d'assistance par ordinateur à l'entraînement de la prononciation.

Le fait de constituer un groupe de locuteurs natifs auxquels sont comparés les apprenants suivant des entraînements n'est pas fréquent dans les études de ce type. Bien qu'il ne soit pas forcément facile à mettre en place, en particulier lorsqu'un accent spécifique est ciblé, nous pensons qu'il est essentiel à plusieurs titres : le premier est que sans cette référence, il est compliqué de conclure fermement qu'un entraînement quel qu'il soit ait permis d'améliorer les performances dans une direction se rapprochant de la norme native. En effet, nous avons pu remarquer dans notre étude que nous serions passés à côté de faits importants sans la présence de ce groupe Natif. Par exemple, alors que nous nous attendions à des scores de perception plafonnant pour le groupe Natif, nous nous sommes aperçus que ce n'était pas le cas pour tous les contrastes. Ainsi, nous ne pouvions pas nous attendre à ce que les apprenants atteignent un tel seuil. De plus, nous avons pu nous rendre compte que, malgré les précautions prises dans la constitution du matériel expérimental, certains des stimuli portaient à confusion pour des locuteurs natifs également.

Nous pensons que le fait que l'entraînement en perception ait comporté beaucoup de variabilité a été bénéfique, conformément à la littérature sur le sujet ; cependant, ce n'est peut-être pas le cas pour l'apprentissage en production. En effet, si l'élaboration d'un système perceptif

performant et précis nécessite d'être soumis à beaucoup de variabilité, comme décrit dans la théorie des modèles à exemplaires de Pierrehumbert (2003; 2001), il n'en est peut-être pas de même pour le système de production de la parole. Certains travaux de Al-Tamimi et Barkat-Defradas (2003) et Al-Tamimi et al., (2002) ont montré que les espaces perceptifs des voyelles d'individus étaient plus grands que leurs espaces de production des voyelles. Cette observation est cohérente avec l'*hyperspace effect* décrit par Whalen et Magen (2001) : des participants choisissent des prototypes vocaliques présentant des valeurs de formants plus extrêmes dans le plan des deux premiers formants que les voyelles qu'ils produisent. Cette différence dans l'étendue des espaces de perception et de production peut s'expliquer par le fait que la production subit des contraintes mécaniques inhérentes notamment à la physiologie de l'individu (e.g., taille du conduit vocal, longueur des cordes vocales). Du fait de ces contraintes, nous pensons que l'acquisition des schémas articulatoires serait plutôt facilitée par une certaine invariance que par de la variabilité. Si l'on considère l'acquisition du langage chez l'enfant dans la perspective du modèle DIVA de Tourville et Gunther (2011), l'enfant apprend les gestes articulatoires par l'intermédiaire de l'imitation de l'adulte. Or, le langage adressé à l'enfant est un langage adapté. Andruski & Kuhl (1997), par exemple, ont montré dans les discours de mères adressés à leur enfant que les voyelles produites étaient différentes que de celles produites dans des énoncés adressés à des adultes. Les auteurs proposent que ces modifications augmentent la discriminabilité entre voyelles, ce qui favoriserait l'apprentissage des catégories phonémiques par des mécanismes d'apprentissage statistique. Bien que d'autres auteurs (McMurray, Kovack-Lesh, Goodwin et McEchron, 2013) aient remarqué que cela n'était pas vrai pour les voyelles centrales, si l'on se place du point de vue de la production, on peut penser que certaines adaptations réalisées dans le discours adressé à l'enfant le sont dans le but que l'enfant puisse produire par imitation les sons produits. C'est-à-dire que la mère produirait des sons qui sont physiologiquement reproductible par son enfant. Dans le cadre de notre étude et de l'entraînement en production que nous avons proposé, nous pensons qu'il aurait fallu, à minima, restreindre les stimuli à répéter à ceux qui ont été produits par un locuteur du même sexe que le participant, comme l'ont fait Karthusina et al., (2015). En effet, il est possible que dans l'exercice que nous leur avons proposé, les participants se soient trouvés en difficulté de reproduire des voyelles dont les caractéristiques acoustiques n'étaient pas atteignables physiologiquement. Ceci pourrait être une des raisons pour lesquelles l'étude de Karthusina et al., (2015) montrent une amélioration des performances des apprenants là où ils ne progressent pas significativement suite à l'entraînement que nous leur avons proposé.

Cette étude montre qu'un entraînement perceptif peut se montrer efficace pour améliorer les compétences de francophones en perception et production de voyelles de l'anglais, bien que les bénéfices sur la production soient limités à l'intelligibilité. En revanche un entraînement en production tel que nous l'avons proposé nécessite que des améliorations soient apportées afin qu'il se montre efficace à la fois sur l'identification et sur la discrimination de phonèmes difficiles pour les apprenants et pour qu'il permette non seulement l'amélioration de l'intelligibilité mais qu'il corrige aussi l'accent étranger.

Référence bibliographiques

Aaltonen, O., Niemi, P., Nyrke, T. et Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*(3), 197‑207.

Abdi, H. (2007). Signal detection theory (SDT). Dans N. J. Salkind (dir.), *Encyclopedia of Measurement and Statistics* (p. 886‑889). Thousand Oaks : Sage.

Akahane-Yamada, R., Strange, W., Downs-Pruitt, J. et Masuda, Y. (1998). Modification of L2 vowel production by perception training as evaluated by acoustic analysis and native speakers. *Journal of the Acoustical Society of America*, *103*(5), 3089-3089.

Aliaga-García, C. et Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. Dans A. M. A. Watkins, A. S. Rauber et B. O. Baptista (dir.), *Recent research in second language phonetics/phonology: Perception and production* (p. 2‑31). Newcastle upon Tyne, UK : Cambridge Scholars Publishing.

Alshangiti, W. et Evans, B. G. (2014). *Investigating the domain-specificity of phonetic training for second language learning: Comparing the effects of production and perception training on the acquisition of English vowels by Arabic learners of English*. International Seminar for Speech Production, Cologne, Allemagne.

Al-Tamimi, J. et Barkat-Defradas, M. (2003). *Inter-dialectal and inter-individual variability in production and perception : A preliminary study in Jordanian and Moroccan Arabic*. I. Ferrando et J. J. Sanchez Sandoval (dir.), 5th AIDA (Association Internationale de Dialectologie Arabe), Cádiz, Espagne.

Al-Tamimi, Jalal-eddin, Girard, M. et Marsico, E. (2002). *Variabilité inter-langue et inter-individuelle en production et en perception : étude préliminaire en arabe dialectal et en français*. Journées d'Étude sur la Parole, Nancy, France.

Andruski, J. E. et Kuhl, P. K. (1997). *The acoustic structure of vowels in infant-and adult-directed speech*. Biannual Meeting of the Society for Research in Child Development, Washington, DC.

Asher, J. J. et García, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, *53*(5), 334‑341.

Baese-Berk, M. M. (2010). Learning novel phonetic categories in perception and production. *The Journal of the Acoustical Society of America*, *128*(4), 2489‑2489.

Baese-Berk, M. M. et Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23‑36.

Barriuso, T. A. et Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, *30*(1), 177‑194.

Bassetti, B. (2008). Orthographic input and second language phonology. Dans T. Piske et M. Young-Sholten (dir.), *Input matters in SLA* (p. 191–206). Clevedon, UK : Multilingual Matters.

Bassetti, B., Escudero, P. et Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, *36*(1), 1‑6.

Bates, D., Maechler, M., Walker, S. et Matrix, L. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1‑48.

Best, C. T. (1995). A direct-realist view of cross-language speech perception. Dans W. Strange (dir.), *Speech perception and linguistic experience: Issues in cross-language research* (p. 171‑204). Timonium, MD : York Press.

Best, C. T. et Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. Dans O. S. Bohn et M. Munro (dir.), *Language experience in second language speech learning: In honor of James Emil Flege* (p. 13‑34). Amsterdam : John Benjamins.

Blair, R. C. et Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*(5), 518-524.

Bliss, H., Abel, J. et Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, *4*(1), 129-153.

Boersma, P. et Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. (version 6.0.26).

Bongaerts, T., Planken, B. et Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. Dans D. Singleton et Z. Lengyel (dir.), *The age factor in second language acquisition* (p. 30-50). Clevedon, U.K : Multilingual Matters.

Boulenger, V., Ferragne, E., Bedoin, N. et Pellegrino, F. (2011). *Derived contrasts in Scottish English: an EEG study*. 17th International Congress of Phonetic Sciences (ICPhS), Hong-Kong.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B. et Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. et Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Burgos, P. (2018). *Non-native pronunciation: Patterns of learner variation in Spanish-accented Dutch.* (Thèse de doctorat). Landelijke Onderzoekschool Taalwetenschap, Utrecht.

Callan, D., Callan, A., Gamez, M., Sato, M. et Kawato, M. (2010). Premotor cortex mediates perceptual performance. *NeuroImage*, *51*(2), 844-858.

Capliez, M. (2011). Typologie des erreurs de production d'anglais des francophones : segments vs. suprasegments. *Cahiers de l'Apliut*, *XXX*(3), 44-60.

Capliez, M. (2016). *Acquisition and learning of English phonology by French speakers* (Thèse de doctorat). Université Lille III - Charles de Gaulle, Lille.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B. et Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421-428.

Carey, M. (2004). CALL visual feedback for pronunciation of vowels: Kay Sona-Match. *CALICO Journal*, *21*(3), 571-601.

Carlet, A. et Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *International Symposium on the Acquisition of Second Language Speech.*

Carlet, A. et Cebrian, J. (2015). *Identification vs. discrimination training: Learning effects for trained and untrained sounds.* 18th International Congress of Phonetic Sciences (ICPhS), Glasgow, Écosse.

Chan, Y. Y. (2004). *EyeSpeak v2.0 software specifications development* (Mémoire de Master). University of Canterbury, Christchurch.

Cheour, M., Korpilahti, P., Martynova, O. et Lang, A. H. (2001). Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. *Audiology & Neuro-Otology*, *6*(1), 2‑11.

Christensen, R. H. B. et Brockhoff, P. B. (2010). sensR – an r-package for sensory discrimination (version R Package version 1.5-1).

Clements, G. N. et Ridouane, R. (2006). *Quantal Phonetics and Distinctive Features: a Review*. ISCA Tutorial and Research Workshop on Experimental Linguistics, Athène, Grèce.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*(4), 919‑924.

Dehaene-Lambertz, G. et Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*(6487), 292‑295.

Dehaene-Lambertz, Ghislaine et Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, *12*(14), 3155-3158.

Delorme, A. et Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9‑21.

Delvaux, V., Huet, K., Piccaluga, M. et Harmegnies, B. (2014). *Production training in second language acquisition: A comparison between objective measures and subjective judgments*. 15th Annual Conference of the International Speech Communication Association (Interspeech), Singapour.

Detey, S. (2016). Enseignement de la prononciation et correction phonétique: principes essentiels. Dans S. Detey, I. Racine, Y. Kawaguchi et J. Eychenne (dir.), *La prononciation du français dans le monde : du natif à l'apprenant* (p. 226‑236). Paris : CLE International.

Detey, S., Fontan, L. et Pellegrini, T. (2018). Traitement de la prononciation en langue étrangère : approches didactiques, méthodes automatiques et enjeux pour l'apprentissage. *Traitement Automatique des Langues*, *57*(3), 15‑39.

Díaz, B., Mitterer, H., Broersma, M., Escera, C. et Sebastian-Galles, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition*, *19*(5), 955-970.

Dufour, S., Nguyen, N. et Frauenfelder, U. H. (2007). The perception of phonemic contrasts in a non-native dialect. *The Journal of the Acoustical Society of America*, *121*(4), EL131‑EL136.

Dufour, S., Nguyen, N. et Frauenfelder, U. H. (2010). Does training on a phonemic contrast absent in the listener's dialect influence word recognition? *The Journal of the Acoustical Society of America*, *128*(1), EL43‑EL48.

Elliott, A. R. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, *79*(3), 356‑371.

Escudero Neyra, P. R. (2005). *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization* (Thèse de doctorat). Landelijke Onderzoekschool Taalwetenschap, Utrecht.

Escudero, P. (2009). Linguistic perception of "similar" L2 sounds. Dans P. Boersma et S. Hamann (dir.), *Phonology in perception* (p. 151‑190). Berlin : Mouton de Gruyter.

Escudero, P. et Boersma, P. (2003). *Modelling the perceptual development of phonological contrasts with optimality theory and the gradual learning algorithm*. S. Arunachalam, E. Kaiser et A. Williams (dir.), 25th annual Penn Linguistics Colloquium.

Ettlinger, M. et Johnson, K. (2009). Vowel discrimination by English, French and Turkish speakers: Evidence for an exemplar-based approach to speech perception. *Phonetica*, *66*(4), 222‑242.

Ferragne, E. (2008). *Étude phonétique des dialectes modernes de l'anglais des Îles Britanniques: Vers l'identification automatique du dialecte* (Thèse de doctorat). Université Lyon 2, Lyon.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70‑79.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233–277.

Flege, J. E., Bohn, O.-S. et Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437‑470.

Flege, J. E. et Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(1), 370‑389.

Flege, J. E. et Liu, S. (2001). The effect of experience on adult's acquisition of a seconde language. *Studies in Second Language Acquisition*, *23*(4), 527‑552.

Flege, J. E, Schirru, C. et MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467‑491.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378.

Fonteneau, E., Frauenfelder, U. H. et Rizzi, L. (1998). On the contribution of ERPs to the study of language comprehension. *Bulletin Suisse de Linguistique Appliquée*, (68), 111-124.

Fougeron, C. et Smith, C. L. (1993). French. *Journal of the International Phonetic Association*, *23*(2), 73‑76.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3‑28.

Frenck-Mestre, C., Peri, P., Meunier, C. et Espesser, R. (2010). *Perceiving non-native vowel contrasts: ERP evidence of the effect of experience*. K. Dziubalska-Kołaczyk, M. Wrembel et M. Kul (dir.), Achievements and perspectives in SLA of speech: News sounds, Poznan, Pologne.

Frumkin, L. (2007). Influences of accent and ethnic background on perceptions of eyewitness testimony. *Psychology, Crime & Law*, *13*(3), 317‑331.

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T. et Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, *42*(2), 936‑944.

Gess, R., Lyche, C. et Meisenburg, T. (2012). Introduction to phonological variation in French. Dans R. Gess, C. Lyche et T. Meisenburg (dir.), *Phonological variation in French: Illustrations from three continents* (p. 1-19). Philadelphia, PA : John Benjamins.

Gibson, J. J. et Gibson, E. J. (1955). What is learned in perceptual learning? A reply to Professor Postman. *Psychological Review*, *62*(6), 447-450.

Gick, B., Bernhardt, B., Bacsfalvi, P. et Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. Dans J. G. Hansen Edwards et M. L. Zampini (dir.), *Phonology and Second Language Acquisition* (p. 309‑322). Amsterdam : John Benjamins.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251.

Grenon, I., Kubota, M. et Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17‑34.

Groppe, D. M., Urbach, T. P. et Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711‑1725.

Guenther, F. H. (1994). Skill acquisition, coarticulation, and rate effects in a neural network model of speech production. *Biological Cybernétics*, *72*(1), 43‑53.

Hemmelmann, C., Horn, M., Reiterer, S., Schack, B., Süsse, T. et Weiss, S. (2004). Multivariate tests for the evaluation of high-dimensional EEG data. *Journal of Neuroscience Methods*, *139*(1), 111-120.

Herment, S. (2018). Apprentissage et enseignement de la prosodie: l'importance de la visualisation. *Revue Française de Linguistique Appliquée*, *23*(1), 73-88.

Hsieh, G. et Kocielnik, R. (2016). *You get who you pay for: The impact of incentives on participation bias*. 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16, San Francisco, CA.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175‑1199.

Isel, F. (2017). Apport de l'électroencéphalographie à l'étude des atypies du traitement du langage. Dans C. Bigliotti, F. Isel et A. Lacheret-Dujour (dir.), *Atypies langagières de l'enfance à l'âge adulte. Apports de la psycholinguistique et des neurosciences cognitives*. Louvain-la-Neuve : De Boeck.

Iverson, P. et Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America*, *122*(5), 2842‑2854.

Iverson, P. et Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866‑877.

Iverson, P et Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, *97*(1), 553‑562.

Iverson, P., Pinet, M. et Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145‑160.

Jakoby, H., Goldstein, A. et Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology*, *48*(11), 1517‑1531.

Jamieson, D. G. et Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205-215.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. Dans K. Johnson et J. W. Mullenix (dir.), *Talker variability in speech processing* (p. 145‑165). San Diego, CA : Academic Press.

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, *1*(6), 571‑591.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H. et Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817‑832.

Kinzler, K. D., Dupoux, E. et Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577‑12580.

Kinzler, K. D., Shutts, K., DeJesus, J. et Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*(4), 623‑634.

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K. et Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, *7*(1), 25-32.

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

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, *21*, 125‑135.

Kuhl, P. K. et Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. Dans W. Strange (dir.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (p. 121‑154). Timonium, MD : York Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. et Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kujala, T., Aho, E., Lepisto, T., Jansson-Verkasalo, E., Nieminen-von Wendt, T., von Wendt, L. et Näätänen, R. (2007). Atypical pattern of discriminating sound features in adults with Asperger syndrome as reflected by the mismatch negativity. *Biological Psychology*, *75*(1), 109‑114.

Kujala, T., Tervaniemi, M. et Schroger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological Psychology*, *74*(1), 1‑19.

Kutas, M. et Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203‑205.

Lacabex, E. G. et Lecumberri, M. (2010). *Investigating training effects in the production of English weak forms by Spanish learners*. 6th International Symposium on the Acquisition of Second Language Speech, Poznan, Pologne.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A. et Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227‑247.

Lee, S., Potamianos, A. et Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455-1468.

Lengeris, A. (2018). Computer-based auditory training improves second-language vowel production in spontaneous speech. *The Journal of the Acoustical Society of America*, *144*(3), EL165-EL171.

Lengeris, A. et Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, *128*(6), 3757‑68.

Lenneberg, E. H. (1960). Review of Speech and Brain Mechanisms by Wilder Penfield, Lamar Roberts. *Language*, *36*(1), 97‑112.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59‑67.

Léon, P. R. (2011). *Phonétisme et prononciations du français* (6e éd.). Paris : Armand Colin.

Liberman, A. M. et Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1‑36.

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of sociolinguistics*, *7*(3), 348-364.

Liu, X. et Tian, X. (2018). The functional relations among motor-based prediction, sensory goals and feedback in learning non-native speech sounds: Evidence from adult Mandarin Chinese speakers with an auditory feedback masking paradigm. *Scientific Reports*, *8*(1), 1−13.

Lively, S. E., Logan, J. S. et Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242−1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. et Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076‑2087.

Logan, J. S., Lively, S. E. et Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874‑886.

Lopez-Calderon, J. et Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

Lu, S., Wayland, R. et Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research*, *1624*, 28‑44.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA : MIT Press.

Macmillan, N. A. et Creelman, C. D. (2005). *Detection theory: a user's guide* (2e éd.). New York, NY : Lawrence Erlbaum Associates.

Maddieson, I., Flavier, S., Marsico, E. et Pellegrino, F. (2014, 2019). LAPSyD - Lyon-Albuquerque Phonological Systems Database, Version 1.1. *http://www.lapsyd.ddl.cnrs.fr/lapsyd/*

Mairano, P., Bouzon, C., Capliez, M. et De Iacovo, V. (2019). *Acoustic distance, Pillai scores, and LDA classification scores as metrics of L2 comprehensibility and nativelikeness*. 19th International Congress of Phonetic Sciences (ICPhS), Melbourne.

Manly, B. F. (2018). *Randomization, Bootstrap and Monte Carlo methods in biology*. New York, NY : Chapman and Hall.

Martin, B. A., Sigal, A., Kurtzberg, D. et Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *The Journal of the Acoustical Society of America*, *101*(3), 1585‑1599.

McClaskey, C. L., Pisoni, D. B. et Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, *34*(4), 323‑330.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D. et McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, *129*(2), 362‑378.

Menning, H. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning & Memory*, *9*(5), 253‑267.

Meunier, F. et Littre, D. (2013). Tracking learners' progress: Adopting a dual 'corpus cum experimental data' approach. *The Modern Language Journal*, *97*(S1), 61‑76.

Miglietta, S., Grimaldi, M. et Calabrese, A. (2013). Conditioned allophony in speech perception: An ERP study. *Brain and Language*, *126*(3), 285‑290.

Mognon, A., Jovicich, J., Bruzzone, L. et Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229‑240.

Morgan-Short, K., Sanz, C., Steinhauer, K. et Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154‑193.

Morgan-Short, K., Steinhauer, K., Sanz, C. et Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933‑947.

Munro, M. J., Derwing, T. M. et Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect*, *21*(1), 67‑79.

Myers, J. L., Well, A. et Lorch, R. F. (2010). *Research design and statistical analysis* (3rd éd.). New York, NY : Routledge.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen A., Vainio M., Alku P., Ilmoniemi R.J., Luuk A., Allik J., Sinkkonen J. et Alho K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432‑434.

Näätänen, R., Paavilainen, P., Rinne, T. et Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544‑2590.

Näätänen, R. et Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375‑425.

Nenonen, S., Shestakova, A., Huotilainen, M. et Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research*, *16*(3), 492‑495.

Nimz, K. (2013). *Sound perception and production in a foreign language: Die Rolle der Orthographie in der fremdsprachlichen Lautperzeption und -produktion* (Thèse de doctorat). Universitätsverlag Postdam, Postdam.

Nishi, K. et Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496‑1509.

Nobre-Oliveira, D. (2007). *Effects of perceptual training on the learning of English vowels in non-native settings*. 5th International Symposium on the Acquisition of Second Language Speech.

Oh, Y. M., Coupé, C., Marsico, E. et Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of functional load. *Journal of phonetics*, *53*, 153-176.

Oh, Y. M., Pellegrino, F., Coupé, C. et Marsico, E. (2013). *Cross-language comparison of functional load for vowels, consonants, and tones.* 14th Annual Conference of the International Speech Communication Association (Interspeech), Lyon.

Ouyang, G., Sommer, W. et Zhou, C. (2016). Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: A solution based on a novel application of residue iteration decomposition. *International Journal of Psychophysiology*, *109*, 9‑20.

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M. et Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*, *118*(1), 177‑185.

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O. et Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*(1), 25‑28.

Picton, T. (1992). The P300 Wave of the Human Event-Related Potential. *Journal of Clinical Neurophysiology*, *9*, 456‑479.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech*, *46*(2‑3), 115‑154.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. Dans J. L. Bybee et P. J. Hopper (dir.), *Typological Studies in Language* (p. 137‑157). Amsterdam : John Benjamins.

Pillot-Loiseau, C., Antolík, T. K. et Kamiyama, T. (2013). Contribution of ultrasound visualisation to improving the production of the French/y/-/u/contrast by four Japanese learners. *Phonetics, phonology, languages in contact Contact (PPLC13): varieties, multilingualism, second language learning*. Paris.

Piske, T., Flege, J. E., MacKay, I. R. et Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*(1), 49-71.

Piske, T., MacKay, I. R. et Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191‑215.

Pisoni, D. B., Aslin, R. N., Percy, A. J. et Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology. Human Perception and Performance*, *8*(2), 297‑314.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O. et Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865‑7870.

Pulvermüller, F. et Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*(1), 49‑71.

Purcell, E. T. et Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*(2), 271‑287.

Rato, A. et Rauber, A. S. (2015). *The effects of perceptual training on the production of English vowel contrasts by Portuguese learners.* 18th International Congress of Phonetic Sciences (ICPhS), Glasgow, Écosse.

Reinke, K. S., He, Y., Wang, C. et Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research*, *17*(3), 781–791.

Sakai, M. et Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224.

Saloranta, A., Alku, P. et Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, *147*, 72‑82.

Sharma, A. et Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, *106*(2), 1078‑1083.

Sharma, A. et Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, *107*, 2697‑2703.

Sharp, E. C., Pelletier, L. G. et Lévesque, C. (2006). The double-edged sword of rewards for participation in psychology experiments. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *38*(3), 269‑277.

Sheehan, K. A., McArthur, G. M. et Bishop, D. V. M. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research*, *25*(2), 547‑553.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemen, R. J. et Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Neuroreport*, *13*(14), 1813‑1816.

Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 269–285.

Spada, N. et Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263‑308.

Stefanics, G., Kremláček, J. et Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in Human Neuroscience*, *8*, 666.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1‑2), 3‑45.

Stevens, K. N. et Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10‑19.

Strange, W. et Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131‑145.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, *26*(2), 233‑253.

Tamminen, H., Peltola, M. S., Kujala, T. et Näätänen, R. (2015). Phonetic training and non-native speech perception - New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, *97*(1), 23‑29.

The MathWorks. (2019). MATLAB and Statistics Toolbox (version Release 2019a). Natick, MA.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*(2), 177‑204.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744‑765.

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT). *Journal of Second Language Pronunciation*, *4*(2), 208‑231.

Thomson, R. I. et Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326‑344.

Thomson, R. I. et Derwing, T. M. (2016). *Is phonemic training using nonsense or real words more effective*. 7th Pronunciation in Second Language Learning and Teaching Conference, Iowa State University Ames, IA.

Thorin, J., Sadakata, M., Desain, P. et McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America*, *144*(1), 92‑103.

Tortel, A. (2008). ANGLISH. Une base de données comparatives de l'anglais lu, répété et parlé en L1 & L2. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, (27), 111‑122.

Tourville, J. A. et Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952‑981.

Tranel, B. (2003). Les sons du français. Dans M. Yaguello (dir.), *Le grand livre de la langue française* (p. 259‑316). Paris : Seuil.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97‑100.

Tremblay, K., Kraus, N., Carrell, T. D. et McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *Journal of the Acoustic Society of America*, *102*(6), 3762‑73.

Tremblay, K., Kraus, N., McGee, T., Ponton, C. et Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79‑90.

Vaissière, J. (2009). Articulatory modeling and the definition of acoustic-perceptual targets for reference vowels. *The Chinese Phonetics Journal*, *2*, 22‑33.

Vaissière, J. (2015). *La phonétique*. Paris : Presse Universitaire de France.

Valdman, A. (2000). Comment gérer la variation dans l'enseignement du français langue étrangère aux États-Unis. *French Review*, 648‑666.

Wang, X. et Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539‑552.

Wee, D. T., Grenon, I., Sheppard, C. et Archibald, J. (2019). *Identification and discrimination training yield comparable results for contrasting vowels*. 19th International Congress of Phonetic Sciences (ICPhS), Melbourne, Australie.

Whalen, D. H. et Magen, H. (2001). Exploring the ‘‘hyperspace’’ effect in vowel perception. *The Journal of the Acoustical Society of America*, *109*(5), 2291‑2291.

Williams, D. et Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the acoustical society of America*, *136*(5), 2751‑2761.

Wilson, S. M., Saygin, A. P., Sereno, M. I. et Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701‑702.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., … Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), 638‑642.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P. et Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Research*, *1072*(1), 175‑185.

Ylinen, S., Uther, M., Latvala, A., Vepsalainen, S., Iverson, P., Akahane-Yamada, R. et Näätänen, R. (2010). Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, *22*(6), 1319‑1332.

Zevin, J. D., Datta, H., Maurer, U., Rosania, K. A. et McCandliss, B. D. (2010). Native language experience influences the topography of the mismatch negativity to speech. *Frontiers in Human Neuroscience*, *4*, 212.