



# Approche plurielle des liens musique-parole pour la didactique de la prononciation du français comme langue étrangère/seconde

Gregory Miras

## ► To cite this version:

Gregory Miras. Approche plurielle des liens musique-parole pour la didactique de la prononciation du français comme langue étrangère/seconde. Linguistique. Université de la Sorbonne nouvelle - Paris III, 2014. Français. <NNT : 2014PA030131>. <tel-01159101>

**HAL Id: tel-01159101**

**<https://tel.archives-ouvertes.fr/tel-01159101>**

Submitted on 2 Jun 2015

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

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

**UNIVERSITE SORBONNE NOUVELLE - PARIS 3**

**ECOLE DOCTORALE N°268**

Langage et langues : description, théorisation, transmission

Didactique des langues des textes et des cultures (EA2288 - DILTEC)

Laboratoire de phonétique et phonologie (UMR 7018 - LPP)

Thèse pour l'obtention du grade de  
Docteur de l'Université Sorbonne Nouvelle – Paris 3 en  
Didactique des langues et des cultures

**Grégory MIRAS**

**Approche plurielle des liens musique-parole  
pour la didactique de la prononciation du français  
comme langue étrangère/seconde**

Thèse dirigée par  
Jean-Paul Narcy-Combes  
et co-encadrée par  
Claire Pillot-Loiseau

Soutenue le 2 Décembre 2014

**Jury :**

**Dominique ABRY**, maître de conférences à l'Université Stendhal-Grenoble 3

**Claire KRAMSCH**, professeure à Berkeley University, pré-rapportrice

**Jean-Paul NARCY-COMBES**, professeur à l'Université Sorbonne Nouvelle-Paris 3, co-directeur

**Claire PILLOT-LOISEAU**, maître de conférences à Université Sorbonne Nouvelle-Paris 3, co-encadrante

**Danièle PISTONE**, professeure à l'Université Paris-Sorbonne, pré-rapportrice



# Résumé

## **Approche plurielle des liens musique-parole pour la didactique de la prononciation du français comme langue étrangère/seconde.**

Cette thèse s'intéresse aux utilisations de la musique instrumentale pour la didactique de la prononciation du français comme langue étrangère/seconde (FLES). Elle s'inscrit dans une perspective plurielle et un cadre théorique émergentiste et socioconstructiviste. La thématique est abordée par trois études (1) psychosociale, (2) psychoacoustique et (3) didactique permettant de questionner à plusieurs niveaux les apports potentiellement développementaux de la musique instrumentale pour la prononciation du français en contexte FLES.

Le but de cette thèse est une contribution à la didactique des langues afin d'apporter des réponses empiriques à des pratiques pédagogiques présentes depuis longtemps mais peu questionnées scientifiquement. Elle permet également de prendre du recul sur les récents résultats en neurocognition et notamment leurs applications pour un développement langagier.

L'étude 1 psychosociale, par questionnaire, montre que la didactique de la musique peut nous aider à amener les apprenants à se focaliser sur le son plutôt que sur le sens des unités sonores langagières. L'étude 2 psychoacoustique révèle que les musiciens ont de meilleurs temps de réaction que les non-musiciens chez des apprenants du français pour un test AXB d'oppositions fines des phonèmes vocaliques du français. Grâce à l'étude 3, didactique interventionniste et quasi-expérimentale, nous avons montré qu'une association entre notes de piano et aperture suggère une potentialité développementale des phonèmes vocaliques du français mais que cette dernière n'est pas supérieure à d'autres approches plus conventionnelles. Cependant, elle semble être un apport cognitif pour certains apprenants même si cet apport montre une tendance bimodale. Ces trois études permettent de montrer que la musique instrumentale peut trouver sa place dans la « boîte à outils » de l'enseignant.

***Mots clés : didactique des langues, français langue étrangère, musique instrumentale, développement langagier, didactique de la prononciation, perception et production.***



# Abstract

## **Pluralistic approach of the links between music and speech in the learning & teaching of the pronunciation of French as a foreign/second language.**

This work deals with the use of instrumental music in the teaching and learning of French as a foreign language. It is part of a pluralistic theoretical framework and an integrated approach of emergentism and social cultural theories. The topic is discussed through three studies: (1) a psychosocial study, (2) a psychoacoustic study and (3) an applied linguistics study – questioning at different levels the potentially developmental intake of instrumental music for French pronunciation as a foreign language.

The aim is a contribution to applied linguistics in order to bring empirical data to long-lasting pedagogical practices but also to question recent neurocognitive results in an applied perspective.

The first study, which is psychosocial and was led by questionnaires, shows that music learning and teaching can help us to have students focus on sound rather than meaning. The second study, which is psychoacoustic, reveals that, among learners of French, musicians have better reaction times than non-musicians to an AXB discrimination test of close French vowels. The third study, which is quasi-experimental and in intervention, suggests that associating piano notes and vowel aperture doesn't lead to bigger developmental intake compared to more conventional approaches. However, it seems to point out to the fact that some learners experience cognitive input even if such impact evokes bimodal tendencies. These three studies prove that instrumental music can be a tool to assist teachers.

***Keywords: applied linguistics, didactics, French as a foreign language, instrumental music, pronunciation, perception & production.***



« *La question de l'éducation est inséparable de celle de l'ordre social et de celle des névroses.* »  
Reich, W. dans Cifali, M & Moll, J. (1985). *Pédagogie et psychanalyse.*  
L'Harmattan, Paris. pp. 27.

« *L'hypothèse sous-jacente était que la circulation globale des informations serait moins vécue  
comme une menace si chacun savait comment se situer.* »  
Narcy-Combes J.-P. (2005). *Didactique des langues et TIC : vers une recherche-action  
responsable.* Ophrys, Paris. pp. 215.





## REMERCIEMENTS

Je tiens à exprimer avant tous mes remerciements à ma codirectrice de thèse, Claire Pillot-Loiseau, pour l'accompagnement et l'aide minutieuse qu'elle a su m'apporter tout au long de ce travail de thèse. Elle a su à la fois me laisser libre de mes choix en tant que chercheur tout en me guidant vers de nouvelles possibilités en tant que médiatrice.

J'aimerais plus particulièrement remercier Jean-Paul Narcy-Combes, mon directeur de thèse, pour m'avoir fait découvrir l'épistémologie et m'avoir ouvert à l'épistémè de l'incertitude. Il a été à la fois un directeur de thèse, un collègue, un médiateur, un mentor voire un psychanalyste. Il est clair que je lui devrai tout au long de ma carrière un cadre théorique qui, je l'espère, ne cessera d'évoluer.

L'ensemble de ce travail n'aurait également pas pu être possible sans Jacqueline Vaissière qui, dès le début de ma thèse, a su me faire confiance et m'apporter des clés qui ont été et qui seront importantes tout au long de ma carrière de chercheur. Je lui dois également l'ensemble des connaissances en phonétique acoustique auxquelles je n'aurais jamais eu accès dans une formation en didactique des langues. J'ai beaucoup appris à ses côtés.

Mes remerciements vont également à l'ensemble des personnes sans qui ce travail de thèse n'aurait pas pu être possible et notamment à l'ensemble des sujets qui ont participé aux trois études, soit au total plus d'un millier d'individus. Il est clair que toute recherche ne peut se faire sans des individus qui donnent de leur temps pour venir remettre en question nos (in)certitudes. Je remercie également à l'Alliance française Paris Ile-de-France pour m'avoir permis de prendre des données.

Mes derniers mots iront à ceux-celles que je nommerai ou oublierai de mentionner : Fabrice Lovarini, Jose Aguilar, Charlotte Koukolia, Yael Uzan-Holveck, Sofia Stratilaki, Marie Rivière, Xingji Rachel Shen, Marine Le Mené, Lin Xue, Sandrine Wachs.

Je remercie également les membres du jury de ma thèse sans lesquels la validation académique de ce travail ne pourrait être assurée : Dominique Abry, Claire Kramsch et Danièle Pistone.

Je souhaite, pour terminer, remercier les membres du DILTEC et du LPP qui ont pu rendre ce travail de thèse possible de sa conception à sa diffusion.



# TABLE DES MATIERES

<b>INTRODUCTION.....</b>	<b>1</b>
DE L'INDIVIDU A LA RECHERCHE .....	1
CE QUE N'EST PAS CETTE THESE .....	2
QUESTIONS DE RECHERCHE.....	3
PLAN ET METHODOLOGIE .....	4
<b>PREMIERE PARTIE : perspectives croisees entre musique et parole .....</b>	<b>6</b>
1. MUSIQUE ET PAROLE : DIACHRONIE ET SYNCHRONIE.....	6
<i>Résumé</i> .....	6
<b>1.1 Une origine ou des origines ?</b> .....	8
1.1.1 Diachronie et inter-espèce en synchronie .....	8
1.1.2 Un inné, des acquis.....	17
1.1.3 Sujets déviants : Amusies, aphasies.....	19
<b>1.2 Parallèles et différences au niveau neurocognitif</b> .....	21
1.2.1 Traitement(s) neurocognitifs.....	21
1.2.2 Pressions environnementales .....	26
1.2.3 Émotions.....	30
<b>1.3 Propriétés acoustiques et perceptives</b> .....	33
1.3.1 Le son.....	33
1.3.2 Instruments : appareil phonatoire et instruments de musique .....	37
1.3.3 Particularités vocaliques du français standard .....	41
2. MUSIQUE ET PAROLE : UN OU DES DOMAINES ? .....	46
<i>Résumé</i> .....	46
<b>2.1 Pluri-inter-transdisciplinarité</b> .....	47
2.1.1 Pluri-inter-transdisciplinarité.....	47
2.1.2 Musique et parole : une science ou des sciences ?.....	50
<b>2.2 L'émergentisme</b> .....	52
2.2.1 Connexionnisme, théories du chaos, complexité et « simplexité ».....	52
2.2.2 Les théories émergentistes .....	57
<b>2.3 Une approche intégrée</b> .....	60
2.3.1 Le socioconstructivisme .....	60
2.3.2 Socioconstructivisme, émergentisme et didactique des langues .....	62
3. DIDACTIQUE DE LA PRONONCIATION EN CONTEXTE FLE/S.....	67
<i>Résumé</i> .....	67
<b>3.1 Articulaire versus perceptif ?</b> .....	68
3.1.1 Panorama de la didactique de la prononciation.....	68
3.1.2 Acquisition phonético-phonologique .....	75
3.1.3 Cibles perceptives et point d'appui .....	78
<b>3.2 De l'articulaire et du perceptif</b> .....	82
3.2.1 Effet McGurk.....	82
3.2.2 Théorie motrice de la perception de la parole.....	83
3.2.3 Les neurones miroirs .....	85
<b>3.3 Didactique de la parole et de la musique</b> .....	88
3.3.1 Les méthodes en marge de la didactique de la prononciation .....	88
3.3.2 La musique en didactique des langues : le cas du FLE/S.....	90
3.3.3 Ce que nous apprend la didactique de la musique .....	93
CONCLUSION GENERALE DE LA PREMIERE PARTIE : .....	96
<i>Synthèse de positionnement</i> .....	97

<b>DEUXIEME PARTIE : les liens musique-parole a l'epreuve de l'experimentation.....</b>	<b>98</b>
4. ETUDE 1 : ENQUETE PSYCHOSOCIALE .....	98
<b>Résumé</b> .....	98
<b>4.1 Méthodologie et questions de recherche</b> .....	101
4.1.1 Matériel et méthodes .....	102
4.1.1.1 Enquête.....	102
4.1.1.2 Recueil des données .....	104
4.1.1.3 Analyse des données.....	104
4.1.2 Profil socio-culturel des sondés .....	104
4.1.2.1 Profil général des sondés .....	105
4.1.2.2 Profil des musiciens.....	108
4.1.2.3 Profil des locuteurs-plurilingues.....	110
4.1.2.4 Profil des locuteurs-signeurs .....	112
<b>4.2 Résultats</b> .....	113
4.2.1 Questionnaires à échelle d'attitudes .....	113
4.2.1.1 Les Musiciens.....	114
4.2.1.2 Les locuteurs-plurilingues .....	118
4.2.1.3 Les locuteurs-signeurs .....	122
4.2.2 Questionnaires à choix multiples .....	124
4.2.2.1 Les Musiciens.....	125
4.2.2.2 Les locuteurs-plurilingues .....	127
4.2.2.3 Locuteurs-signeurs .....	129
<b>4.3 Analyses et discussions</b> .....	130
4.3.1 Musiciens et langage musical .....	130
4.3.2 Locuteurs et langage verbal .....	134
4.3.3 Langages musical et verbal.....	138
4.3.4 Une modalité du signe ? .....	143
<b>4.4 Conclusion générale de l'étude 1</b> .....	146
<b>Synthèse de positionnement</b> .....	148
5. ETUDE 2 : EXPERIMENTATION PSYCHOACOUSTIQUE .....	149
<b>Résumé</b> .....	149
<b>5.1 Méthodologie et questions de recherche</b> .....	152
5.1.1 Matériel et Méthodes .....	153
5.1.1.1 Déroulement général du test.....	154
5.1.1.2 Les auditeurs .....	154
5.1.2 Les test AXB.....	162
5.1.2.1 Choix du test.....	162
5.1.2.2 Déroulement général du test AXB .....	165
5.1.2.3 Les stimuli .....	168
5.1.3 Les tâches .....	172
5.1.3.1 Caractéristiques des tâches .....	173
5.1.3.2 Déroulement de la phase de tâches.....	174
5.1.3.3 Pré-étude 1 : sélection par des musiciens et non-musiciens .....	175
5.1.3.4 Pré-étude 2 : confirmation par des experts en art.....	181
<b>5.2 Résultats</b> .....	183
5.2.1 Questionnaire relatif aux tâches.....	184
5.2.2 Les apprenants du français.....	187
5.2.3 Apprenants vs. locuteurs natifs du français.....	194
<b>5.3 Analyses et discussions</b> .....	198
5.3.1 Taux de discrimination et traitement acoustique.....	199
5.3.2 Effets sur le temps de réaction.....	200
5.3.3 Impact de la tâche .....	201
<b>5.4 Conclusion générale de l'expérimentation 2</b> .....	203
<b>Synthèse de positionnement :</b> .....	204

6. ETUDE 3 : DIDACTIQUE QUASI-EXPERIMENTALE ET INTERVENTIONNISTE .....	205
<i>Résumé</i> .....	205
<b>6.1 Méthodologie et questions de recherche</b> .....	207
6.1.1 Matériel et Méthodes .....	210
6.1.1.1 Le dispositif général .....	210
6.1.1.2 Les apprenants .....	212
6.1.1.3 Les enregistrements .....	213
6.1.2 Médiation de la perception et production des sons de la parole .....	217
6.1.2.1 La réflexion théorique .....	217
6.1.2.2 Le trapèze vocalique .....	219
6.1.2.3 Les exercices d'entraînement .....	220
<b>6.2 Résultats</b> .....	221
6.2.1 L'enquête par questionnaire .....	222
6.2.1.1 Tendances entre le pré-test t-1 et le post-test t+1 .....	223
6.2.1.2 Tendances au post-test t+2 .....	226
6.2.1.3 Tendances spécifiques au travail en groupe .....	232
6.2.2 Analyses préliminaires .....	238
6.2.2.1 Mise en place .....	239
6.2.2.2 Durée .....	239
6.2.2.3 Fréquence fondamentale .....	242
6.2.3 Analyse perceptive .....	246
6.2.3.1 Mise en place .....	246
6.2.3.2 Degré de confiance .....	247
6.2.3.3 Analyse perceptive du pré-test .....	249
6.2.3.4 Analyse perceptive du post-test .....	257
<b>6.3 Analyses et discussions</b> .....	260
6.3.1 Représentations et progression générale .....	260
6.3.2 Impact du groupe et de la modalité de production .....	262
6.3.3 Confusions phonémiques et dénativisation/accommodation phonologique .....	264
<b>6.4 Conclusion générale de l'étude 3</b> .....	267
<i>Synthèse de positionnement</i> : .....	269
<b>DISCUSSION GENERALE</b> .....	<b>270</b>
BILAN GLOBAL .....	270
LIMITES ET IMPLICATIONS DIDACTIQUES .....	274
<b>CONCLUSION GENERALE</b> .....	<b>278</b>
<b>BIBLIOGRAPHIE</b> .....	<b>281</b>
<b>INDEX DES AUTEURS ET DES CONCEPTS</b> .....	<b>301</b>
<b>LISTE DES FIGURES</b> .....	<b>307</b>
<b>LISTE DES TABLEAUX</b> .....	<b>312</b>

## Liste des sigles et abréviations

→ Une feuille volante des abréviations utiles est disponible.

F0 : fréquence du fondamental

F1, F2, F3, F4 : premier, deuxième, troisième et quatrième formants

Hz : Hertz

ms : milliseconde(s)

API : Alphabet Phonétique International

FLE/S : français langue étrangère/seconde

LE : langue étrangère

LS : langue des signes ou locuteur-signeur (précisé si ambigu)

Fr : français

MSS : musicien **sans** solfège

MAS : musicien **avec** solfège

LPR : locuteur-plurilingue **avec** pratique régulière d'au moins une langue étrangère en contexte naturel

LPNR : locuteur-plurilingue **avec** pratique régulière d'au moins une langue étrangère en contexte naturel

DV : document visuel

DAP : document de parole continue en français

DAM : document de musique instrumentale

## INTRODUCTION

### De l'individu à la recherche

« Parle-moi de tes recherches, je te dirai qui tu es ». Cet adage me semble vrai particulièrement en sciences humaines et sociales et d'autant plus vrai pour un sujet de recherche qui a été choisi de manière personnelle. L'origine de ce projet de recherche doctorale ne se situe pas initialement dans des besoins socioéducatifs ni sur la résolution de tensions sociales liées à l'éducation, la place des langues et leur enseignement dans un contexte particulier. Ce travail est né d'une curiosité et d'une quête permanente de remise en question de représentations et de catégories figées même s'il répond également à des besoins liés à la pratique orale du français à travers sa dimension psychoaffective. Il est d'autant plus important pour moi que j'ai pu constater qu'un certain nombre d'apprenants s'intéressent à leur prononciation et qu'un certain nombre d'enseignants ne se sentent pas légitimes dans ce domaine. La question de l'enseignement/apprentissage de la prononciation peut être déterminante notamment dans des perspectives professionnelles et d'intégration. Ce travail sera le reflet d'un parcours disciplinaire épistémologique mais également d'une prise de recul épistémique sur le « moi » en tant que chercheur comme individu social et émotionnel. L'existence de cette thèse en tant qu'objet du réel (dé)matérialisé est l'expression d'une réflexion individuelle dont les ruptures ont été catalysées par des rencontres nombreuses et enrichissantes, des lectures et des données. Il est clair que l'origine de ce travail prend sa source dans une expérience personnelle où l'objet musical ne semblait pas être si éloigné de la parole que la doxa pouvait y prétendre. Ils sont tous deux une création de l'Homme, par l'Homme et pour l'Homme qui peuvent être à la fois incluant ou excluant socialement. Ils peuvent être le reflet ou le déclencheur d'états émotionnels qui sont au centre de la performance orale. Il est certain que la parole se compose d'éléments linguistiques connus et étudiés depuis l'avènement du structuralisme à partir du XX<sup>ème</sup> siècle et qui sont au centre des études fondamentales en phonétique et phonologique mais ce qui nous intéresse, ce sont les processus et mécanismes, autres que linguistiques, qui peuvent soit favoriser soit limiter la performance orale d'une langue étrangère à court ou long terme. Les termes de « processus et mécanismes » sont utilisés à dessein car nous nous situerons dans le postulat central de la neuropsychologie selon lequel « *les conduites et les processus mentaux sont sous-tendus par des événements physico-chimiques ayant leur siège dans le cerveau* » (Seron, 1998: 35). A ce titre, nous verrons les « langues » comme des codes qui n'existent que parce qu'ils sont incarnés par des individus (biologique) et que parce qu'ils sont co-construits



(social). Nous assumerons, dès lors, que toute « *langue* » « *vivante* » n'est pas unique mais un ensemble de variétés régionales, sociales, générationnelles, géographiques, individuelles, etc. Les différents codes linguistiques mis en action chez des individus plurilingues seraient tous actifs mais certains s'activeraient de manière volontaire et/ou involontaire lors de la performance orale (perception et production). Le but de notre approche de médiation de la perception et de la production des sons en langue étrangère ne se veut pas normatif mais tend à amener l'apprenant à prendre conscience de la manière dont il perçoit et produit les sons de ses « *langues* » et se sentir plus à l'aise dans la performance de chacune d'entre elles. De ce fait, nous ne nions pas la volonté de certains « *d'effacer* » leur accent et d'autres de le maintenir car nous ne pouvons contester le fait que l'accent est un marqueur social. Il est néanmoins important, en tant que chercheurs, d'amener les individus de nos sociétés à réfléchir sur leurs représentations. Nous tenons à préciser, dans cette introduction, que même si nous nous positionnons dans une perspective neurocognitiviste, nous prendrons une position critique sur les résultats réductionnistes dans ce domaine, notamment au regard des processus complexes en action lors de la performance orale. Notre travail de thèse se concentrera sur le français comme langue étrangère/seconde. Nous utiliserons l'abréviation FLE/S pour faire référence à un contexte d'apprentissage, c'est-à-dire les codes multiples et variés utilisés par les locuteurs francophones dont la/les langue(s) première(s) ne sont pas une variété de français. Cependant, nous tenons à rappeler que le FLE n'est pas une langue et son utilisation hypéronymique révèle souvent de la remise en question inachevée de la validité ontologique du construit de « *langue* » (Narcy-Combes, 2005).

L'individu « *je* » en tant que rédacteur de ce travail de recherche s'effacera au profit du « *nous* » qui par tradition symbolise une prise de recul académique. Cependant, même si l'ensemble de ce travail est le résultat de rencontres multiples, nous souhaitons assumer qu'il a été rédigé sous le prisme d'un individu et par conséquent, nous emploierons le « *je* » lorsque nous aborderons une prise de recul épistémique et notamment lors des synthèses de positionnement.

### **Ce que n'est pas cette thèse**

Avant de définir ce sur quoi portera notre étude, nous souhaitons préciser ce qui ne sera pas traité. Notre travail de thèse ne concernera pas les liens avec le rythme ou la prosodie car même si ces terrains semblent prometteurs, ils feraient l'objet d'une recherche à part entière. Nous ne traiterons pas non plus en détail l'apprentissage précoce de la musique et/ou des langues car notre expérience d'enseignement ne nous a pas amené à envisager ce terrain. Pour terminer, nous souhaitons préciser que nous ne développerons pas les possibilités de la musique

instrumentale comme accès au culturel (pour ne pas dire interculturel) ni au potentiel créatif et d'imagination. Nous nous focaliserons sur le système musical contemporain globalisé issu de la tradition occidentale principalement.

Nous tenons également à avertir le lecteur que cette thèse se veut être plus un début qu'une fin dans un champ où tout reste à faire. Nous avons dû à la fois construire un cadre théorique nouveau et le questionner à l'aide de données empiriques. L'ampleur de cette tâche nous a poussé à faire des choix méthodologiques afin de respecter les exigences de l'exercice doctoral. Etant donné le caractère complexe des liens musique-parole, nous avons souhaité les questionner à plusieurs niveaux en respectant les exigences inhérentes aux paradigmes de chaque domaine disciplinaire qui sont la base de notre réflexion. De ce fait, nous avons dû faire des choix méthodologiques en faveur tantôt de la phonétique et tantôt de la didactique des langues au détriment de certaines de nos volontés. Ces choix ont, malgré nous, amené notre thèse vers une analyse plus réductionniste que nous le souhaitions même si elle nous permet de mieux concevoir l'après-thèse. Les données obtenues nous ont permis de définir le caractère potentiellement développemental des outils pédagogiques proposés. Nous parlerons de « *dispositif potentiellement développemental* » car, dans notre cadre, l'apprentissage est vu comme une progression linéaire ponctuée de ruptures. Ces points de rupture s'apparentent plus à un développement langagier qu'à une acquisition qui révélerait une stabilité en contradiction avec notre vision du vivant. Le caractère « potentiel » rappelle l'épistémè de l'incertitude dans lequel nous nous plaçons.

## Questions de recherche

De manière concrète, notre thèse aura pour but de questionner les liens entre musique et parole, à travers un cadre théorique pluriel, dans le but de mesurer la ou les applications / implications possibles de la musique instrumentale dans la didactique de la prononciation en contexte FLE/S. Nos questionnements sont multiples mais quatre questions principales ont guidé ce travail:

- 1) A quels niveaux se situent les différences entre musique et parole ? Au niveau évolutif ? Représentationnel ? Social ? Cognitif ? Neurologique ? Tout à la fois ?
- 2) Le rapport que les individus portent sur l'objet verbal ou musical modifie-t-il la manière dont ils les conçoivent ? Que peut-on apprendre de l'un pour enrichir l'autre ?
- 3) Existe-t-il des différences entre musiciens et non-musiciens chez des apprenants du français en contexte FLE/S ? Si oui, à quel niveau et peuvent-elles être mesurées empiriquement ?

- 4) Une approche pédagogique qui lie musique et parole peut-elle avoir un apport potentiellement développemental ? Si oui, est-il supérieur à d'autres approches ?

### **Plan et méthodologie**

Cette thèse se compose de deux parties de trois chapitres chacune, soit au total six chapitres. Chaque chapitre est précédé par un résumé qui apporte ses points saillants et se termine par une synthèse de positionnement qui permet de replacer chaque chapitre dans la perspective générale de ce travail.

Le cadre théorique constituera notre première partie qui tentera de mettre en évidence le caractère complexe des liens qui unissent ces deux objets verbal et musical dans le but de répondre à la question 1 susmentionnée : « à quels niveaux se situent les différences entre musique et parole ? Au niveau évolutif ? Représentationnel ? Social ? Cognitif ? Neurologique ? Tout à la fois ? ». Le chapitre 1 aura pour but de réfléchir sur les liens entre musique et parole dans une perspective diachronique et synchronique. Pour cela, nous tenterons de déterminer la ou leurs origines en section 1.1. Puis, nous déterminerons les parallèles et différences entre le traitement de ces deux objets sur plan neurocognitif en section 1.2. La section 1.3 clôturera cette réflexion par les propriétés acoustiques et perceptives du son pour la musique et la parole et notamment des phonèmes vocaliques sur lesquels nos études porteront. L'ensemble de ces caractéristiques fera émerger un questionnement disciplinaire qui sera développé dans le chapitre 2. Il questionnera, en section 2.1, la notion de pluri-inter-transdisciplinarité qui définit le caractère pluriel de notre travail de thèse. La section 2.2 aura pour but de développer les théories émergentistes dans lesquelles nous nous plaçons tout en montrant qu'une approche intégrée avec le socioconstructivisme est nécessaire en section 2.3. Le chapitre 3, qui termine la première partie, s'intéresse plus particulièrement à la didactique de la prononciation en contexte FLE/S. La section 3.1 questionnera les liens entre perception et production pour montrer en section 3.2 que ces deux processus sont intimement liés dans l'acte de production orale. La section 3.3 clôturera notre réflexion théorique en montrant quelle place pourrait avoir la musique instrumentale dans la médiation de la prononciation.

La partie 2 s'appuiera sur notre cadre théorique tout en le remettant en question à travers des données empiriques. Etant donné notre volonté épistémologique et le peu de travaux dans le domaine, nous avons souhaité questionner les liens musique-parole à plusieurs niveaux afin de nous permettre d'ouvrir des voies pour l'après-thèse. Nous souhaitons avertir le lecteur que ces trois études ne reposent pas sur le même terrain mais permettent de prendre de la distance sur les liens musique-parole à travers une approche plurielle. Les trois études seront présentées de

la même manière, soit une présentation de la méthodologie expérimentale (section 4.1, 5.1 et 6.1) puis l'analyse des résultats (4.2, 5.2 et 6.2) et nous terminerons par une interprétation de ces derniers (4.3, 5.3 et 6.3). Le chapitre 4 représentera notre première étude psychosociale menée à l'aide de questionnaires et qui ont pour but de faire émerger des tendances dans les représentations de locuteurs plurilingues et de musiciens concernant leur rapport respectivement à l'objet verbal et musical. Elle nous permettra de répondre à la question 2 susmentionnée soit : « le rapport que les individus portent sur l'objet verbal ou musical modifie-t-il la manière dont ils les conçoivent ? Que peut-on apprendre de l'un pour enrichir l'autre ? ». Le chapitre 5 sera constitué de notre étude 2 psycho-sociale dont le but est d'apporter une réponse à notre questionnement 3 : « existe-t-il des différences entre musiciens et non-musicien chez des apprenants du français en contexte FLE/S ? Si oui, à quel niveau et peuvent-elles être mesurées empiriquement ? ». Cette étude prend la forme de deux tests de discrimination AXB des phonèmes vocaliques du français intercalés par une exposition à un document visuel ou audio de parole en français ou audio de musique instrumentale et passés par des apprenants du français. Nous tenterons de montrer des différences de traitement psychoacoustique (taux de discrimination et temps de réaction) entre musiciens et non-musiciens mais également l'influence de l'exposition à une musique instrumentale sur la progression entre les deux tests. La dernière étude qui constituera notre chapitre 6 portera sur notre dernier questionnement : « une approche pédagogique qui lie musique et parole peut-elle avoir un apport potentiellement développemental ? Si oui, est-il supérieur à d'autres approches ? ». Elle s'intègre dans une perspective de didactique expérimentale et interventionniste et tente de savoir si une association entre notes de musique et apertures peut avoir un caractère potentiellement développemental au sein d'un dispositif d'apprentissage et si cet apport est supérieur à d'autres approches. Nous lierons l'ensemble de ces points lors de la discussion ainsi que la conclusion générale de ce travail doctoral.

## **Première partie : perspectives croisées entre musique et parole**

Cette première partie « *Perspectives croisées entre musique et parole* » a pour but de porter une réflexion à plusieurs niveaux des liens entre musique et parole à travers la littérature scientifique des différents domaines concernés. Nous nous intéresserons, tout d'abord, à l'étude diachronique de ces liens par l'origine des langages puis à leur étude synchronique en nous questionnant sur le caractère inné ou acquis des langages verbal et musical au regard de différents paramètres environnementaux, neurocognitifs, acoustiques et phonétiques. A ce titre, nous justifierons les raisons qui nous ont poussés à focaliser notre étude sur les sons vocaliques dans le cadre de l'enseignement du français comme langue seconde ou étrangère. Nous légitimerons également ce travail de thèse en menant une réflexion sur les problématiques épistémologiques que posent les recherches pluri-inter-transdisciplinaires en didactique des langues. Nous présenterons, en outre, un regard croisé de l'émergentisme et du socioconstructivisme en tant qu'éclairage possible de ce genre de recherche. Pour terminer, nous nous concentrerons sur la didactique de la prononciation dans un contexte de français comme langue seconde ou étrangère en recentrant notre étude sur les processus perceptifs (et leurs liens avec ceux articulatoires) afin de mettre en évidence les utilisations de la musique instrumentale dans ce domaine de l'enseignement.

### **1. Musique et parole : diachronie et synchronie**

#### **Résumé**

Le chapitre « *Musique et parole : diachronie et synchronie* » permet de faire un bilan diachronique puis synchronique des liens qui relient les objets musical et verbal en tant que langages.

La section 1.1 s'interrogera sur les différentes théories qui s'intéressent à l'origine de la faculté de communiquer de l'Homme afin de mettre en évidence une origine probable commune entre les langages verbal et musical. Nous montrerons, en outre, que ces deux langages ont certainement un « ancêtre » commun (un protolangage) caractérisé par des propriétés holistique, multimodale, intentionnelle et musicale (Mithen, 2006 : 120-1). La musicalité ainsi que la gestuelle auraient joué un rôle important vers un passage à un langage verbal de type computationnel. (Mithen, 2006 : 147-9 ; Wray, 2002 ; entre autres). Cependant, malgré cette origine commune probable, il n'est pas possible actuellement de rejeter l'hypothèse nulle pour la musique, contrairement à la parole, selon

laquelle les humains seraient directement adaptés à ce langage par la sélection naturelle (Patel, 2008 : 359-67 ; 371-77). Ces résultats mitigés se retrouvent dans le cas d'études cliniques des patients aphasiques ou amusiques qui mettent en évidence un traitement plutôt indépendant car ces patients peuvent ne pas (totalement) développer une des deux capacités sans que l'autre semble être touchée par la lésion (Mithen, 2006 : 62 ; Hebert & Cuddy 2006 : 200 ; Peretz *et al.*, 2004 : 373-390 ; entre autres).

La section 1.2 a pour but de prolonger la réflexion synchronique sur les liens entre musique et parole en analysant les processus neurocognitifs sous-jacents à leurs traitements. Nous montrerons qu'actuellement l'ensemble des travaux sur le sujet tend vers des ressources partagées en ce qui concerne les processus d'intégration structurale des événements, (par exemple *l'hypothèse de ressources d'intégration syntaxique partagées* ; Grahn, 2011), mais des réseaux distincts en ce qui concerne le stockage des représentations musicales et linguistiques à long terme. Afin de compléter ce tableau, nous mettrons en évidence l'influence de certaines pressions environnementales (expertise musicale, âge, langues parlées, etc.) sur les capacités dans les deux domaines qui nous intéressent : la musique et la parole (Tillmann *et al.*, 2005 ; Bigand 2004 : 215 ; Deutsch & Dooley, 2009 : 2400 ; entre autres). Pour terminer cette section, nous traiterons brièvement des émotions déclenchées par ces deux langages puisqu'ils intègrent tous les deux une intentionnalité émotionnelle conditionnée par des filtres cognitifs d'une part, et la transmission d'un message, d'autre part.

La section 1.3 rappelle les principaux mécanismes physiques, acoustiques et physiologiques liés aux productions sonores musicale et verbale (Leipp, 1989 ; Vaissière, 2006 ; Renard, 1978 ; entre autres). Nous prendrons le temps de montrer les similarités et différences entre ces deux domaines. Nous finirons en montrant que certaines voyelles du français standard possèdent des caractéristiques acoustiques et phonétiques particulières : cardinales, focales et rapport formantique plutôt stable (Jones, 1956 ; Vaissière, 2011 ; Gendrot & Adda, 2005 ; Georgeton *et al.*, 2012 ; entre autres) ; ce qui en font des objets pertinents dans notre étude qui se focalise sur les processus neurocognitifs aux niveaux acoustique et phonétique.

## 1.1 Une origine ou des origines ?

Dans cette section, nous nous intéresserons aux origines, peut-être communes, des langages musical et verbal de manière diachronique mais également synchronique. Nous développerons l'approche synchronique en essayant de terminer la part d'acquis et d'innée dans la construction neurocognitive de ces deux objets et à travers les travaux portant sur l'analyse clinique de patients amusiques et aphasiques.

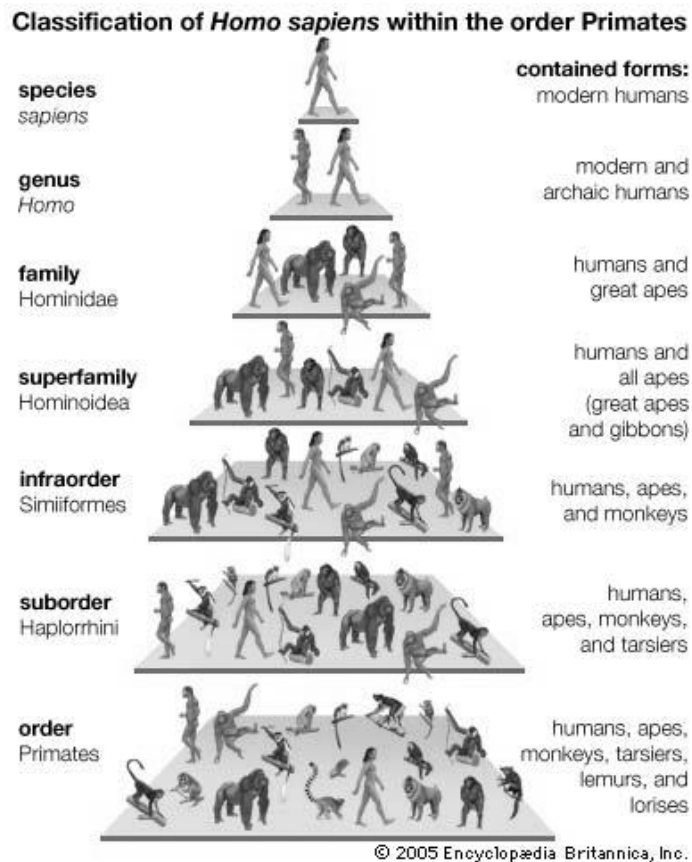
### 1.1.1 Diachronie et inter-espèce en synchronie

Saussure définit le **langage** comme « *la capacité humaine à communiquer* » (Saussure, 1967 : 23-35). Nous parlerons, dans ce travail de thèse, de « *langue* »<sup>1</sup> comme l'objectivation descriptive du système langagier par une société et de **parole** (ou **discours**) comme la réalisation physique/concrète de ce langage (*Ibidem*). Néanmoins, il semblerait que, dans une perspective élargie, la définition saussurienne du langage soit lacunaire en ce qui concerne le caractère « *humain* » de la communication. Comme le rappellent Li et Hombert (2002 : 175-176), l'étude de l'évolution diachronique du langage est le domaine de la linguistique diachronique car elle étudie les évolutions linguistiques à partir d'un point d'apparition supposé. Or nous traiterons, ici, de ce qui se passe avant l'apparition du langage humain et par conséquent de comportements communicationnels. Nous noterons, toutefois, que la définition du langage comme technologie (Auroux, 2001 dans Narcy-Combes, 2006b : 307) dans la relation transductive (Simondon, 1989) « *humain/société/technologie(s)* » (Narcy-Combes, *ibidem*) permet de remettre en cause cette apparente opposition terminologique. Même si aucun autre organisme vivant du monde animal ne possède les mêmes caractéristiques de langage que celles de *Homo sapiens*, nous ne pouvons nier l'existence d'autres types de langage dans le monde animal. La communication animale est plus limitée dans le sens où ce langage ne dissocie pas la « *référence* » et la « *fonction* ». Les animaux ne dissocient pas, en effet, l'identité de l'objet avec la relation que le sujet lui porte ; c'est à dire que le signal sonore porte une fonction et non un sens ; il n'est donc pas symbolique (Li & Hombert, 2002 : 176). Il semblerait que cette forme de communication soit la forme archéologique de ce que nous appelons maintenant : le langage (humain). De la même manière, ce que nous appelons « *chant* » pour des animaux comme les baleines ou certains oiseaux fait apparaître le même type de limite

---

1 Le terme « *langue* » apparaîtra entre guillemets tout au long de notre travail de thèse puisque notre positionnement ira vers une approche non-symbolisme et que la notion de « *langue* », dans ce cadre, relève du construit.

comme le rappelle Patel (2008 : 356) : « *le chant des animaux prévient toujours des mêmes choses : période d'accouplement, avertissements territoriaux et statuts sociaux* »<sup>2</sup> (d'après Marler, 2000).



**Figure 1:** Classification de *Homo sapiens* dans l'ordre des primates (Britannica en ligne, 2013)

Afin de limiter notre champ de réflexion, cette dernière portera sur l'ordre des primates (Figure 1) qui montre, dans des conditions de vie sauvage, une habilité pour le langage. Cette dernière est caractérisée par quatre éléments (Mithen, 2006 : 120-1) :

- « *holistique* » : « *ils manquent de sens arbitraire et persistent et ne sont pas composés d'énoncés dont la grammaire apporte un niveau supplémentaire de sens* »<sup>3</sup>. Ce terme est repris des travaux de Wray (2000).
- « *intentionnel* » : « *ils tentent de produire un effet émotionnel attendu chez autrui* »<sup>4</sup>
- « *multimodal* » : « *ils utilisent autant la gestuelle que les vocalisations* »<sup>5</sup> (seulement dans la famille *Hominidae*)
- « *musical* » : « *ils utilisent, de manière substantielle, le rythme et la mélodie* »<sup>6</sup>

2 « *animal songs always advertise the same set of things, including readiness to mate, territorial warnings, and social status* »

3 « *holistic* » : « *they lack consistent and arbitrary meanings, and are not composed into utterances by a grammar that provides an additional level of meaning* »

4 « *manipulative* » : « *they are trying to generate some form of desired behavior in another individual* »

5 « *multi-modal* » : « *they use gesture as well as vocalization* »

6 « *musical* » : « *they make substantial use of rhythm and melody* »



L'origine du langage, selon Steven Mithen – professeur d'archéologie cognitive à l'Université de Reading, chez les premiers hominidés, aurait les mêmes caractéristiques que celles susmentionnées chez les (grands) singes actuels : « *holistique, intentionnel, multimodal et musical* » (H.M.M.M.)<sup>7</sup>. La question qui reste néanmoins, est de savoir quelles sont les origines évolutives qui ont poussé la branche « hominidé » à favoriser l'émergence d'un langage permettant la dissociation de la référence et de la fonction d'un objet.

Les découvertes les plus récentes remettent en question l'impact social sur le développement physiologique de l'homme moderne. Nous considérerons les théories à venir dans une perspective archéologique et par conséquent basées sur une reconstruction à posteriori d'événements en dehors de notre portée. La bipédie serait probablement à l'origine de modifications profondes de la physiologie humaine permettant une meilleure adaptation à l'environnement. Les étapes (tri non chronologique et possiblement non exhaustif) de la bipédie semblent être les suivantes (chronologie du genre *Homo* disponible en annexe 1) :

- 1) **Position corporelle 1** : « *Kevin Hunt suggère qu'un bassin large représentait une plateforme stable tandis que des doigts courbés et des bras puissants semblaient adaptés, non pas pour escalader des arbres, mais pour se laisser pendre par un bras, appuyé sur ses deux pattes et se nourrissant du bras libre.* »<sup>8</sup> (Hunt, 1994 dans Mithen, 2006: 144-145).
  
- 2) **Position corporelle 2** : « *la propagation d'un environnement du type "savane" a favorisé la nécessité de réduire le stress thermique provoqué par une exposition croissante au soleil dans un paysage où les arbres se font de plus en plus rares. Quand un hominidé est debout, il ne reçoit la lumière du soleil que par le haut du crâne et ses épaules, tandis qu'à quatre pattes, l'intégralité de son dos est exposée. De plus, l'air est plus frais loin du sol.* »<sup>9</sup> (Wheeler, 1988 in Mithen, 2006 : 145).

---

<sup>7</sup> « *holistic, manipulative, multi-modal and musical* »

<sup>8</sup> « *the broad pelvis, [Kevin Hunt] suggests, made a stable platform, while the curves fingers and powerful arms should be seen as adaptations not for climbing trees but for hanging by one hand while standing on two legs and feeding with the other.* »

<sup>9</sup> « *the spread of savannah-like environments [has favoured] the need to reduce heat stress caused by increased exposure to the sun in landscapes that had markedly fewer trees [...] When standing erect, a hominid would only absorb sunlight on the top of its head and shoulders, whereas when knuckle-walking its whole back would be exposed. Moreover, the air is cooler away from the soil itself.* »

Nous précisons que ces théories ne sont que des hypothèses et nous mentionnerons que d'autres hypothèses sous-tendent que la bipédie serait apparue dans des groupes vivants près d'étendues d'eau facilitant, par la poussée d'Archimède, la station debout (Schrenk *et al.*, 2004 dans Niemitz, 2010 : 241-263)

→ Amélioration pour le langage : aucune

- 3) **Position du larynx** : « Aiello argumente que la descente du larynx était simplement la conséquence d'adaptations anatomiques nécessaires à la bipédie. Comme la colonne vertébrale devait pénétrer dans la boîte crânienne par le bas plutôt que par l'arrière, il y avait moins d'espace, pour le larynx, entre la colonne vertébrale et la bouche. »<sup>10</sup> (Aiello, 1996 dans Mithen, 2006 : 146 ; Liberman, 1984: 287-329).
- 4) **Nature laryngée** : « la bipédie peut avoir provoqué ce changement en réduisant la pression sélective sur la fonction locomotrice d'un larynx valvulaire qui devient moins rigide »<sup>11</sup>. (Mithen, 2006 : 147)
- 5) **Physiologie buccale** : il est constaté « une réduction de la taille des dents et de la mâchoire [qui serait due] au changement de comportement alimentaire en faveur d'un régime carnivore. »<sup>12</sup> (Mithen, 2006 : 129)
- 6) **Élargissement de la colonne vertébrale** : l'augmentation de l'innervation des muscles intercostaux aurait permis de contrôler la respiration pendant l'acte de parole augmentant, ainsi, la capacité de vocalisation. Cette innervation accrue serait la conséquence directe de la bipédie imposant un meilleur maintien du tronc du corps ainsi qu'une puissance supérieure due à la position debout des organes de la respiration (MacLarnon et Hewitt, 1999 : 341-363).

---

10 « Aiello argued that the low larynx was merely a consequence of the anatomical adaptations necessary for bipedalism. Because the spinal cord now had to enter the braincase from below rather than from behind, there was less space between the spinal cord and the mouth for the larynx »

11 « bipedalism may have caused the change by relaxing the selective pressure on the locomotor function of the valvular larynx [becoming less rigid] »

12 « the reduction in the size of the teeth and jaws because of the dietary trend towards meat-eating. »

→ Amélioration pour le langage : on constaterait une augmentation du choix possible de productions sonores à travers un meilleur contrôle de la fonction phonatoire, laryngée et pulmonaire dû à des changements physiologiques provoqués par la bipédie.

- 7) **Neurocognition** : « *le fait que Homo ergaster ait un cerveau plus grand que celui de l'australopithèque, non seulement dans l'absolu mais aussi relativement à la taille du corps, pourrait être expliqué, selon Leslie Aiello, par les nouvelles exigences du contrôle sensorimoteur que la bipédie requiert et permet. Se tenir debout et marcher sur deux jambes nécessitent un contrôle permanent du centre de gravité du corps à travers la coordination de petits groupes de muscles.* »<sup>13</sup> (Aiello, 1996 dans Mithen, 2006 : 146)
- 8) **Augmentation de la taille du néocortex** : l'acquisition du langage nécessite une capacité de mémorisation cognitive importante de par la nécessité de mémoriser un grand nombre d'énoncés idiomatiques ou ce que Wray (2002a, 2002b) appelle des « *formulaic sentences* » (énoncés formulaires ; traduction proposée par Legallois, 2009 : 14).
- 9) **Réduction de l'appareil digestif** : elle est nécessaire à l'augmentation de l'encéphale des hominidés. Le changement de régime alimentaire, à travers la cuisson, a permis une réduction des besoins de l'appareil digestif indispensable à la dépense d'énergie nécessaire à une densité encéphalique croissante (Aiello & Wheeler, 1995).

→ Amélioration pour le langage : la bipédie aurait favorisé des changements neurophysiologiques permettant un traitement accru d'unités linguistiques complexes et abstraites.

- 10) **Régime alimentaire** : « *les comportements de chasseur et de charognard pour obtenir de la viande dans la savane aurait provoqué une pression sélective sur les hominidés à augmenter leur répertoire holistique, leurs gestes intentionnels et leurs cris.* »<sup>14</sup> (Mithen, 2006 : 134)

---

13 « *the fact that Homo ergaster had a larger brain than the australopithecines – not just in absolute terms but also relative to body size - could, [Leslie Aiello] argues, be explained by the new demands on sensorimotor control that bipedalism both required and enabled. Standing and walking on two legs requires that the center of gravity is constantly monitored and small groups of muscles frequently recruited and changed to correct its position.* »

14 « *the lifestyle of hunting and scavenging for meat on the savannah would have placed pressure on hominids to expand their repertoire of holistic, manipulative gestures and calls.* »

11) **Vie sociale** : « pendant la période de *Homo habilis* et *Homo rudolfensis*, la taille des groupes de premiers hominidés avait tellement augmenté que les grognements n'étaient plus suffisant comme seul moyen d'expression d'engagement envers les autres membres du groupe. »<sup>15</sup> (Aiello & Dunbar, 1993 dans Mithen, 2006: 135).

12) **Psychocognition** : « la taille relativement importante du cerveau de *Homo habilis* et *Homo rudolfensis*, comparé à celle des australopithèque ou des grands singes d'aujourd'hui, peut aller en faveur de la théorie de l'esprit. »<sup>16</sup>. (Dunbar, 2004 dans Mithen, 2006 : 128)

→ Amélioration pour le langage : l'augmentation des pressions sociales sur des facteurs physiologiques aurait favorisé l'apparition d'un langage complexe dans une perspective évolutionniste.

L'ensemble du processus de développement du langage humain, tel que nous le considérons actuellement, est résumé par Mithen (2006 : 147) comme « *la complexité sociale des premiers hominidés et humains a créé des pressions sélectives en faveur de comportements communicationnels. Ils furent réalisés par des changements physiologiques et plus particulièrement par ceux associés à la bipédie qui ont permis à *Homo ergaster* de produire un plus grand choix de vocalisations que les ancêtres hominidés les plus proches des grands singes actuels* ». <sup>17</sup>

Nous avons pu comprendre dans cette première partie que l'apparition d'un langage, inédite dans le monde animal, s'est effectuée par des transformations physiologiques dues à un environnement modifié. Ces transformations sont devenues des critères sélectifs dans un contexte où la communication devenait nécessaire pour un système social se complexifiant. Il reste à se demander quel est le « *chaînon manquant* » entre la forme « *primitive* » de langage,

---

15 « *by the time of *Homo habilis* and *Homo rudolfensis*, the size of the early hominid groups had grown so large that grooming was no longer feasible as the only means to express one's social commitment to other members of the group.* »

16 « *The relatively large brains of *Homo habilis* and *Homo rudolfensis*, when compared with the australopithecines and modern-day apes, might reflect an enhanced theory of mind capability.* »

17 « *the social complexity of the early hominids and Early Humans [...] created selective pressures for enhanced communication. These were realized by the physiological changes [...] especially those associates with bipedalism, which enabled *Homo ergaster* to produce a significantly larger array of vocalizations than their immediate hominid ancestors of modern-day apes.* »

semblable à celle observée dans le monde actuel animalier (hors *Homo sapiens*), et la forme complexe de l'homme actuel.

Rappelons tout d'abord qu'une des particularités du langage humain est la capacité à conceptualiser des faits qui se déroulent en dehors de la deixis « *je* », « *ici* » et « *maintenant* », contrairement au langage animal qui comme le rappellent Li et Hombert (2002, 178) « *tend à associer une réaction émotionnelle ou motivée à une situation particulière en incluant le propre état interne hormonal de l'animal.* »<sup>18</sup>. A l'aide des éléments susmentionnés et dans le prolongement des travaux d'A. Wray, Mithen avance l'hypothèse que le protolangage aurait les mêmes caractéristiques que celle relevées dans le langage des primates : holistique, multimodal, intentionnel et musical. La thèse d'un protolangage holistique semble socio-anthropologiquement plus probable que celle d'un protolangage computationnel (cf. Mithen, 2006 : 147-9). Le caractère holistique de ce langage prendrait la forme d'énoncés, porteurs de sens, mais non décomposables en sous-unités. Ces énoncés holistiques auraient une forme plus musicale que langagière : les limites d'un tel langage auraient été compensées par des liens intrinsèques au corps (mouvement des bras, mains, etc.) et des variations prosodiques fortes (variations d'intensité, de fréquence fondamentale, rythmiques, etc.) dans le but de provoquer des émotions pertinentes chez le récepteur. L'importance de ces deux aspects, que sont la multimodalité et la musicalité, aurait émergé de la bipédie. Tandis que la gestuelle aurait été facilitée par la libération des membres supérieurs de la marche, le genre *Homo* aurait fait ses premiers pas en rythme condition nécessaire pour la coordination des membres (cf. la maladie de Parkinson pour laquelle un rythme d'*input* auditif peut réguler la marche et les mouvements chez certains patients : Thaut *et al.*, 2001).

Nous avons donc pu définir les termes du protolangage et ceux du langage humain moderne ; il nous reste donc à comprendre quelles sont les étapes qui ont permis le passage d'un langage holistique à un langage computationnel. Selon la théorie développée par Wray (2002), cette transition aurait été possible par « *segmentation* » d'unités perceptivement proches. La segmentation de certains phonèmes/syllabes aurait été renforcée par des variations prosodiques omniprésentes et des associations son/geste. Cette théorie est renforcée par les travaux de modélisation sur ordinateur de Kirby (2001, 2003) permettant la mise en avant de l'impact de la transition générationnelle sur la stabilisation des formes holistiques afin de constituer un seul

---

18 « *tends to be associated with the emotional or motivational reaction to a particular situation including the animal's own internal hormonal state.* »

et même langage computationnel. Cette étape de segmentation se serait déroulée il y a approximativement 200 000 ans, pour deux raisons :

- 1) jusqu'aux premiers *Homo sapiens*, les *Homo ergaster*, *heidelbergensis* ou *neanderthalensis* vivaient dans des groupes fermés et limités dont les échanges sociaux ne nécessitaient pas d'autre forme que le « *H.M.M.M.* » langage. *Homo sapiens*, en Afrique, quant à lui, semble s'organiser dans des groupes sociaux plus importants avec des rôles économiques et plusieurs positions sociales. Ces caractéristiques auraient nécessité des échanges d'information plus importants, favorisant le passage vers un système computationnel. Il n'y a pas de consensus pour le moment, pour déterminer si les changements sociaux ont été la cause ou l'effet de cette transition.
- 2) des mutations génétiques auraient pu avoir lieu, facilitant le traitement du langage en tant que segments. Prenons pour exemple le gène FOXP2 présent dans de nombreuses espèces mais dont la variante humaine (à quelques acides aminés près) semble faire émerger chez l'homme, en cas d'absence ou mutation, des troubles majeurs du langage (Lai *et al.*, 2001 ; Enard *et al.*, 2002). Nous précisons que ce gène semble être important dans le langage humain mais bien évidemment ni le seul ni l'unique.

Le caractère sélectif de la parole devient dès lors clair ; il se pose alors la question du caractère sélectif de la musique en tant que langage. Plusieurs hypothèses adaptationnistes coexistent actuellement car aucune ne semble complètement suffisante comme le rappelle Patel (2008 : 368-71) :

- La Sélection sexuelle Darwin (1871) puis reprise par Miller (2000), Mithen (2006) et Levitin (2010) :

**Point fort** : la musique permet d'effectuer des parades nuptiales complexes.

**Point faible** : « *faire la cour* » n'est qu'une fonction parmi tant d'autres de la musique telle que « *soigner, faire le deuil, fêter, mémoriser* »<sup>19</sup> (Cross, 2003 dans Wallin *et al.*, 2003 et repris par Patel, 2008 : 369).

- Le développement mental et social (Cross, 2003) :

**Point fort** : la musique permettrait aux individus de développer un ensemble de capacité cognitives et motrices à travers des explorations sociales « *séduire* » ;

**Point faible**: des personnes ayant des déficits congénitaux touchant la musique ne montrent pas de déficit fort sur les autres plans cognitifs et sociaux (Patel, 2008 : 369).

---

<sup>19</sup> « *healing, mourning, celebrating, memorizing* »

– La cohésion sociale (Morley, 2003) :

**Point fort** : la musique permettrait de renforcer les liens sociaux à travers les rituels de création musicale qu'elle implique. Ce lien peut être à plusieurs niveaux : inter-individus (membres d'une communauté), intergénérationnel (mère-enfant) ou mystique (vivants-défunts, hommes-dieux).

**Point faible** : la musique est également une activité solitaire. Les individus autistes ont des déficiences dans « *créer du lien avec autrui* » mais ont une forte sensibilité à la musique. Les mères ont bien d'autres média de communication avec leurs enfants que la musique (Patel, 2008 : 370-1).

Dans une perspective diachronique et inter-espèce en synchronie, il semblerait donc, selon de récents travaux en paléanthropologie cognitive, que le protolangage ait été holistique, intentionnel, multimodal et musical. Cette forme de communication liait, de manière indissociable, la production sonore avec le corps (la gestuelle) et la musicalité (proéminence des aspects prosodiques). Il semblerait que, dans une approche paléanthropologique, la musique, le rythme, le corps et le discours aient une origine commune forte comme le rappelle Pernon (1998 : 6) :

*« La musique n'a pas d'origine connue et il est probable qu'elle fut fille de la danse, elle-même de la marche. Le chant est dérivé de la déclamation et les premiers instruments de musique ont été les mains ou les pieds ».*

L'ensemble des sons productibles aurait été augmenté à travers des modifications physiologiques dues à la bipédie. Ces changements seraient devenus sélectifs dans de nouveaux schémas sociaux et environnementaux. Le passage d'une forme de langage holistique vers une forme de langage computationnel trouverait son origine dans des besoins linguistiques croissants et probablement des mutations aléatoires. La segmentation aurait été permise par la prosodie et la gestuelle de ce protolangage. L'épopée du langage humain moderne s'est déroulée sur une échelle archéologique importante. Pour terminer, nous noterons qu'il semblerait, selon Mithen (2006: 263-264) à partir des travaux de P. Carruthers (2002 : 657-674), que cette nouvelle forme de langage aurait facilité l'émergence d'une fluidité cognitive permettant notamment à l'*Homo sapiens* de lier cognitivement différents modules/intelligences (cf. les intelligences multiples : Gardner, 1997) tels que l'intelligence sociale (gérer un groupe social), du monde naturel (compréhension des phénomènes naturels) et technique (création et utilisation d'outils). La question du caractère sélectif des langages verbal et musical mène à se poser la question de leur caractère inné ou acquis à l'heure actuelle.

### 1.1.2 Un inné, des acquis

Même si les théories de l'évolution du protolangage semblent particulièrement intéressantes et renseignent sur l'importance de la capacité de communiquer de l'être humain, il reste à déterminer si la parole/discours et la musique partagent la même implication évolutive. Cela revient à poser la question de la part d'inné et d'acquis de ces deux objets de langage.

Il est largement admis à l'heure actuelle et comme le précise Deacon (1997 dans Patel, 2008 : 359) que « *la parole et le cerveau humain ont agi l'un sur l'autre à travers un processus de coévolution* »<sup>20</sup>. Cela revient donc à se demander s'il existe une sélection naturelle pour le langage verbal, c'est-à-dire si la sélection naturelle a joué un rôle direct sur les mécanismes d'acquisition du langage. Notre position sera celle de Deacon, susmentionnée, car elle permet de sortir du clivage manichéen scientifique entre les théories adaptationnistes (Pinker et Jackendoff, 2005 : 201-36) en faveur d'une sélection directe et les théories constructivistes (Tomasello *et al.*, 2005 : 675-691) pour un lien indirect. Nous nous positionnerons dans la perspective d'un « *nativisme minimal* » du langage (Clark & Misyak 2009 : 253-261) selon lequel, une aire cérébrale pourrait être vue comme représentant un certain universel langagier si et seulement si elle est prédisposée à traiter un certain nombre suffisant de fonctions linguistiques générales (ex : l'aire de Broca).

Dans cette perspective, Patel (2008: 359-367) reprend onze grandes idées-clés permettant de penser que la sélection naturelle aurait influencé les mécanismes d'acquisition du langage verbal :

- 1) **Babillages** : les jeunes enfants produisent des sons de manière spontanée même chez des enfants atteints de surdité que ce soit de manière sonore ou gestuelle. Il semblerait, par conséquent, que les babillages émergent du processus de maturation entre l'*output* moteur et l'*input* sensoriel afin de découvrir les structures phonologiques du langage.
- 2) **Anatomie du conduit vocal** : l'abaissement du larynx permet, entre autre, d'augmenter la palette de sons productibles au risque d'augmenter le risque d'étouffement lors de la déglutition.
- 3) **Apprentissage vocal** : l'homme est l'un des rares animaux et l'un des seuls primates à profiter d'un apprentissage langagier complexe permettant la gestion des différents articulateurs. Cette capacité est largement en place à l'âge de 3 ou 4 ans.

---

20 « *language and the human brain have acted as selective forces on each other via a process of co-evolution* »



- 4) **Apprentissage de la structure linguistique des sons de la parole** : les unités sonores de la parole s'organisent en un système complexe qui se met en place assez tôt dans le développement de l'enfant tant en production qu'en perception.
- 5) **Période critique d'acquisition du langage verbal** : de nombreux résultats tendent à prouver l'existence d'une période critique pendant laquelle l'accès à un *input* langagier peut avoir des effets importants sur la structure cérébrale.
- 6) **Structures comparables dans le développement du langage oral et des signes** : le langage oral et celui des signes partageraient les mêmes composantes linguistiques et activeraient des régions similaires. Ceci prouverait le rôle de la sélection naturelle dans l'organisation d'un système de communication ne dépendant pas de la modalité.
- 7) **Robustesse de l'acquisition du langage langagier** : malgré un *input* langagier assez limité, les enfants sont capables d'intégrer un code linguistique de manière rapide et robuste dans le temps.
- 8) **Ajout de complexité dans la pauvreté de l'input linguistique** : malgré un *input* linguistique limité, les enfants sont capables de complexifier leur système linguistique sans accès à un système déjà établi.
- 9) **Gène pertinent dans la capacité langagière** : la présence d'une forme déficiente d'un gène (FOXP2) chez l'homme peut impliquer de nombreux troubles de la parole
- 10) **Coût biologique d'un déficit langagier** : un déficit langagier semble être dommageable à un individu pour survivre et se reproduire.

L'ensemble de ces éléments apporte de fortes preuves en faveur du fait que le langage verbal aurait été influencé par la sélection naturelle. Il reste à se demander s'il est possible de trouver autant d'indices en ce qui concerne la musique. Malgré le fait que la musique semble exister dans un très grand nombre de cultures humaines (pour ne pas dire toutes), il n'existe pas de preuve évidente du caractère sélectif de la musique. Une approche empirique des organismes vivants semble être plus prometteuse qu'une approche adaptationniste pour la détermination d'une pression sélective dans l'apparition de la musique dans les sociétés humaines (Fitch, 2006). Nous reprendrons les quatre grandes idées-clés de Patel (2008: 371-377) :

- 1) **Rythme d'apprentissage des structures musicales** : jusqu'à présent, aucune étude n'a directement montré un développement rapide de capacités musicales qui ne seraient pas reliées au langage ou toute autre fonction générale auditive.

- 2) **Effet de la période critique** : il n'y a pas de preuve concrète pour une période critique des habilités musicales même si un apprentissage précoce de la musique peut avoir un impact important sur la neuroanatomie du système moteur et notamment sur le développement des capacités d'« *oreille absolue* ».
- 3) **Robustesse de l'acquisition des capacités musicales** : actuellement, aucune donnée n'a permis de montrer que le développement d'habilités musicales est aussi robuste que celui d'habilités linguistiques.
- 4) **Coût biologique d'une impossibilité d'acquérir des capacités musicales** : il n'y a pas de preuve apparente que des déficiences congénitales musicales soient un poids biologique en termes reproductifs, par exemple.

Notre position conclusive sera celle de Patel (2008 : 412) lorsqu'il dit que :

*« dans le cas du langage, il semble y avoir suffisamment de preuve pour rejeter l'hypothèse nulle, selon laquelle, les humains ne seraient pas directement adaptés, par la sélection naturelle, pour cette capacité. Dans le cas de la musique, cependant, je ne pense pas qu'il y ait suffisamment de preuve pour rejeter l'hypothèse nulle. De plus, la recherche doit se prononcer à savoir si les hommes se sont adaptés par l'évolution pour acquérir des capacités musicales. »<sup>21</sup>*

Son hypothèse est que la musique est une création humaine (à l'image du feu) dont la pratique (passive ou active) modifie nos structures cérébrales et démontrent la capacité humaine à nous modifier nous-mêmes. Nous nous intéresserons maintenant, plus particulièrement, aux cas cliniques afin de réfléchir sur les liens neurophysiologiques entre musique et parole.

### 1.1.3 Sujets déviants : Amusies, aphasies

Jusqu'à présent, la compréhension de l'organisation cérébrale passait par l'analyse comportementale de sujets atteints de lésions et l'analyse médicale du cerveau *post mortem*. Depuis les années 1990 et l'apparition de l'Imagerie par Résonance Magnétique Fonctionnelle (IRMf), il a été possible de visualiser en temps réel l'activation des différentes zones cérébrales *in vivo*. Néanmoins, l'analyse des capacités neurocognitives de sujets déviants reste encore un point majeur dans la compréhension croisée des processus en œuvre dans le traitement de différents objets. Dans le cadre de notre réflexion, nous focaliserons notre attention sur les cas acquis ou congénitaux de :

---

21 « *in the case of language, there appears to be enough evidence to reject the null hypothesis that humans have not been directly shaped by natural selection for this ability. In the case of music, however, I do not think enough evidence has accumulated to reject the null hypothesis [...] Further research is needed to address whether humans have been specifically shaped by evolution to acquire musical abilities* »

- Aphasie : « *déficit dans les fonctions de base du langage qui touche toutes les modalités* »<sup>22</sup> (Caplan, 1992 dans Patel, 2008 : 270)
- Amusie : « *déficit dans les capacités liées à la perception et/ou production de la musique* »<sup>23</sup> (Peretz, 2006 dans Patel, 2008 : 226)

Nous pouvons nous demander si les patients aphasiques et amusiques présentent des déficiences communes de manière récurrente, partielle ou nulle. L'étude de patients ayant subi des lésions cérébrales n'a pas apporté de précision satisfaisante sur le sujet puisqu'il a été relevé des cas de dissociation entre amusie et aphasie mais également d'associations sélectives et modérées dans les capacités musicales chez des patients aphasiques ou des capacités langagières chez des patients amusiques (Hebert & Cuddy 2006 : 200). Par exemple, des patients aphasiques ne pouvant produire que quelques mots peuvent produire des mélodies avec une précision importante (Hébert *et al.* 2003 : 1838-50; Peretz *et al.*, 2004 : 373-390), tandis que certains patients amusiques peuvent reconnaître des mots dans une chanson mais ne pourront pas les chanter (Peretz *et al.*, 1997 : 354-367). Selon un travail de synthèse mené par Mithen (2007) sur les cas d'aphasie et d'amusie, il semblerait que :

*« Les réseaux neuronaux qui opèrent le traitement langagier et musical ont un degré certain d'indépendance, il est possible de « perdre » ou ne jamais développer une des deux capacités tandis que l'autre semble normale, du moins en comparaison à la première »* (traduit par l'auteur de Mithen, 2006 : 62)

Afin de renforcer notre propos, nous porterons brièvement notre attention sur la lecture. Hebert et Cuddy (2006 : 201) ont repéré onze cas sur seize répertoriés, présentant des déficiences de lecture de texte et de musique chez des musiciens (semi-)professionnels. Sur les cinq patients restants, seul un présente une claire dissociation entre la lecture musicale et de texte. Il apparaîtrait néanmoins que :

*« La lecture de texte et musicale, bien que souvent associée, semble reposer sur des processus distincts partageant des corrélats anatomiques adjacents. De plus, la nature sélective des déficiences dans la lecture musicale à la suite de lésions cérébrales suggère un modèle à plusieurs composantes. »*<sup>24</sup> (Hebert & Cuddy 2006: 204)

Il faudrait prendre en compte une réalité complexe de la lecture musicale pour un traitement dissocié de la hauteur tonale et du rythme par exemple (Peretz & Kolinsky 1993 : 301-325)

---

22 « *deficit of core language functions that cuts across modalities* »

23 « *deficit in musical perception and/or production abilities* »

24 « *music and text reading, though often associated, may rely on distinct processes that share adjacent anatomical correlates. Moreover, the selective nature of music-reading deficiencies following brain damage suggests a model with multiple components* »

mais également trois modalités principales de lecture : visuelle-motrice (lire et jouer), visuel-chanter (lire et chanter la note) et visuel-verbal (lire et nommer la note) (Hebert & Cuddy, 2006 : 204 ; Schön *et al.* 2002 : 2286). Ainsi, Schön *et al.* (2002 : 2288) proposent l'existence d'une route de décodage sémantique via des représentations internes abstraites, une route directe via un algorithme de décodage asémantique et une route combinant les deux. Néanmoins, d'après des corrélats anatomiques, les lectures musicale et textuelle reposent sur des aires adjacentes dans l'hémisphère gauche (Hebert & Cuddy, 2006 : 203). Selon toute évidence, « *un réseau neuronal dédié peut être responsable de cette capacité sophistiquée qui permet de transformer la notation musical en une réponse fine motrice* »<sup>25</sup> (Schön *et al.* 2002: 2288).

La réflexion croisée de données sur les liens musique et parole dans une perspective diachronique et synchronique de patients amusiques et aphasiques, nous a permis de mieux appréhender la directionnalité des recherches interdisciplinaires sur le sujet. L'ensemble des techniques de neuro-imagerie actuelles permet d'affiner ces travaux.

## **1.2 Parallèles et différences au niveau neurocognitif**

Dans cette section, nous allons dresser un bilan des parallèles et différences au niveau neurocognitif entre le traitement de la musique et de la parole. Nous nous intéresserons, plus particulièrement, aux processus impliqués, à l'influence de pressions environnementales et de la place des émotions dans ces traitements.

### **1.2.1 Traitement(s) neurocognitifs**

De manière générale, les deux hémisphères du cerveau sont comparables en termes de taille et de surface. Les quatre lobes principaux ont un aspect macroscopique sensiblement similaire (Gazzaniga *et al.*, 2001 : 325). Cependant, depuis le XIX<sup>ème</sup> siècle, il est reconnu que le cerveau présente une asymétrie fonctionnelle. Ce constat a notamment été conforté par les travaux de Geschwind sur les cerveaux de cent individus droitiers (Geschwind et Levitsky, 1968 : 186-7). Il n'est dès lors plus remis en question que les aires de Broca et Wernicke (Broca, 1861 ; Wernicke, 1874/1977) traiteraient de manière préférentielle respectivement les processus phono-articulatoires et les processus phono-sémantiques. De plus, Hesling (2002 : 7) rappelle que :

---

25 « *a dedicated neural network might be responsible for this sophisticated skill that allows transforming music notation into a precise motor response* »

« L'hémisphère droit se caractérise par des zones plus importantes de cortex associatif (aires qui n'ont pas de projections hors du cerveau et dont les neurones communiquent entre eux et avec d'autres neurones du cortex), tandis que l'hémisphère gauche se caractérise par des zones motrices et sensorielles plus importantes. »

Cependant, Peretz (1985 : 575) émet des doutes quant à une dichotomie trop tranchée sur la question en ce qui concerne la musique :

« Les résultats des travaux qui ont étudiés les différences latérales dans la perception des sons simultanés sont dans l'ensemble cohérents avec l'idée que l'hémisphère gauche se caractériserait par un mode analytique de traitement et l'hémisphère droit par un traitement de nature différente. Cependant, le manque d'indications indépendantes des asymétries observées sur le traitement opéré par les sujets laisse encore obscure la nature des opérations mieux effectuées dans l'hémisphère droit. De ce fait, la distinction en termes de modes de traitement analytique et holistique demeure une hypothèse plausible mais non convaincante quant à la division des fonctions entre hémisphères cérébraux dans la perception des sons musicaux. »

Néanmoins, il est communément admis, de nos jours, que les habiletés cognitives comme le langage représentent un ensemble de sous-éléments spécifiques. De cette manière, les recherches actuelles ne permettent pas de déterminer avec précision le rôle de chaque hémisphère dans le traitement de la prosodie, par exemple. Différentes hypothèses se confrontent ou se complètent sur le sujet (Di Cristo, 2002:17) :

- la **latéralisation radicale** : la prosodie implique majoritairement hémisphère droit (HD),
- la **latéralisation fonctionnelle** : les aspects affectifs sont traités par HD et les aspects linguistiques de la prosodie par l'hémisphère gauche (HG),
- la **distribution du traitement des émotions** : la prosodie implique majoritairement HD mais les émotions positives HG et les émotions négatives HD,
- la **latéralisation paramétrique** : les différences de hauteur sont traitées par HD et les différences de durée par HG.

Malgré ces théories parfois opposées, il semblerait que « les deux hémisphères soient concernés par le traitement de la prosodie en général [...] mais on note une nette préférence de HD pour effectuer cette tâche cognitive » (Di Cristo, 2002 : 17). Le traitement de l'information est considéré non pas en fonction de la tâche (langage, musique, syntaxe, sémantique, etc..) mais plutôt en terme d'élément référentiel ou non. Bien qu'il existe une interdépendance des deux hémisphères, le gauche aurait une prédilection pour les éléments référentiels tandis que le droit traiterait l'information de manière non-référentielle.

Nous nous intéressons plus particulièrement à une région nommée la zone de Broca. Cette partie du cerveau est également mise en action pour l'organisation visuelle des mots par le langage des signes (Hickok *et al.*, 1998 : 129-36). En outre, « *l'émergence de la faculté créative de langage chez l'homme est sans doute liée à l'accroissement de ses capacités cognitives et des aires cérébrales de Broca et Wernicke* » (Vaissière, 2006 : 47). C'est donc le centre d'organisation des informations dans le temps quelle que soit l'origine sensorielle de l'impulse et quelle que soit la région qui l'analysera ensuite. Cette zone est responsable de la structure, la syntaxe des éléments communicatifs. La région de Broca est directement reliée à celle de Wernicke via le faisceau arqué. Cette dernière précède l'aire de Broca dans l'analyse d'un signal sonore, et permet de reconnaître un ensemble de phonèmes connus: « *un mot* » ou « *groupe de sens* ».

Il reste à se demander si dans cette perspective, musique et parole peuvent posséder des corrélats neurophysiologiques communs. Depuis les années 1980, le « *dual factor model* » (modèle du double facteur) propose que la parole soit traitée simultanément de manière linguistique et acoustique. Ces différents traitements peuvent être mis en évidence à travers la variation des caractéristiques de stimuli et du type de tâche dans des expérimentations phonéto-phonologiques. Werker & Tee (1984b : 1876) dissocient le traitement linguistique en deux entités :

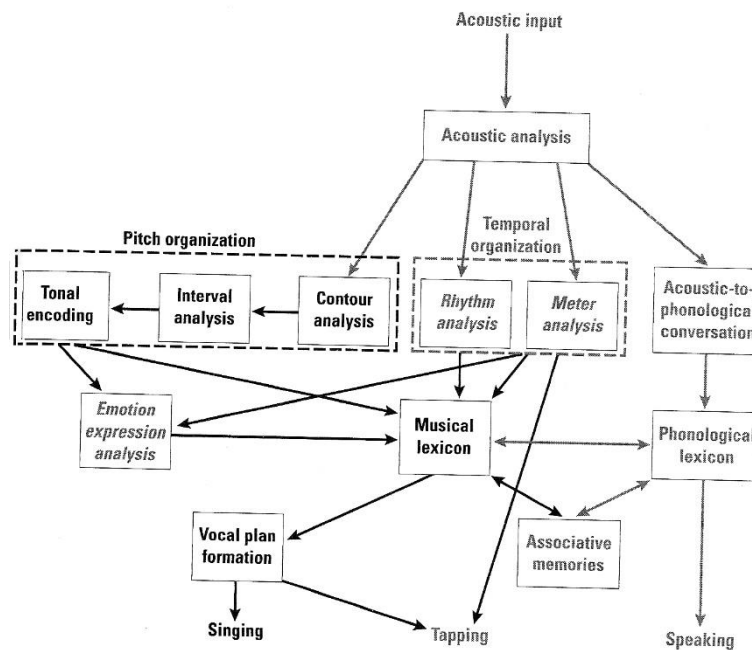
*« **le niveau phonémique**: les stimuli sont traités comme des événements de sens qui peuvent être encodés efficacement et représentés dans la mémoire pour une longue période. **Le niveau phonétique**: le son peut être perçu comme des percepts synthétisés et sans sens et peuvent ainsi être stockés dans la mémoire pour une courte période mais s'amenuise plus rapidement qu'au niveau phonémique. **L'information phonétique et phonémique** peut être disponible simultanément mais pendant une courte période de temps. »*<sup>26</sup>

Il est probable que nous puissions retrouver des corrélats neurophysiologiques, entre musique et parole, dans le traitement acoustique et éventuellement phonétique, mais un traitement dissocié au niveau phonologique. Précisons que toutes les informations acoustiques sont traitées de manière égale au niveau des aires primaires corticales, et ce quelle que soit leur nature. Le traitement phonologique ou musical qui est mis en place au cours du temps permet de traiter certaines informations pertinentes plus rapidement et plus efficacement dans certaines conditions à travers un système de *feedback*, notamment (Simmonds *et al.*, 2011).

---

26 « **phonemic level**, stimuli are treated as meaningful events that can be efficiently encoded and represented in memory for a long duration. **Phonetic level**, the sound may be perceived as synthesized (nonmeaningful) percepts, and thus may be retained in memory for a short duration but have a more rapid decay period than phonemic stimuli. **Phonemic and phonetic information** may be available simultaneously for a short period of time. »

En ce qui concerne la musique, le modèle le plus récent proposé par Peretz & Coltheart (2003 ; Figure 2) propose une représentation modulaire de la musique.



**Figure 2:** Une représentation modulaire de la musique (Peretz & Coltheart, 2003 : 690). Chaque rectangle représente un bloc processuel et chaque flèche un flux d'information. Les blocs spécifiques à la musique apparaissent en noir foncé et les autres en noir/gris clair.

Néanmoins, cette «modularité» serait de nature complexe, mettant en avant une multiplicité de traitements pour la musique et la parole. Patel (2008 : 357) tient à préciser que « *les modules peuvent être le produit du développement [cérébral] plutôt que le reflet de spécialisations déterminées de manière innée* »<sup>27</sup>. Des liens au niveau acoustique et prosodique existeraient, comme le proposent Peretz & Coltheart (2003 : 690) dans le modèle ci-dessus (Figure 2). De plus, des travaux tendent en faveur d'une SSIRH : « *Shared Syntactic Integration Resources Hypothesis*<sup>28</sup> » (Grahn, 2011 : 235) proposant que :

« *la musique et le langage partagent des ressources impliquées dans les processus d'intégration structurale des événements (musicaux ou linguistiques et des processus de mémoire de travail) [...] En revanche, la musique et le langage disposeraient de réseaux neuronaux distincts pour le stockage des représentations musicales et linguistiques à long terme, autorisant ainsi des déficits sélectifs du traitement de la musique (i.e., amusie) et du langage (i.e., aphasie)* » (Hoch, 2008 : 39).

<sup>27</sup> « *modules can be a product of development, rather than reflecting innately specified brain specialization* »

<sup>28</sup> L'hypothèse de ressources d'intégration syntaxique partagées

Il semble clair qu'il faille dissocier deux ensembles : la syntaxe et la sémantique. Un grand nombre de protocoles expérimentaux ont mis en évidence des processus similaires dans le traitement syntaxique de ces systèmes sonores. Dans les deux cas, une violation syntaxique provoque, lors d'enregistrements électrophysiologiques, l'émission d'une onde P600 (onde positive autour de 600 ms ; Kutas & Hillyard, 1983 ; entre autres). Le processus de violation repose sur le développement d'attentes musicales ou langagières qui facilitent la compréhension. Elles permettent donc d'affirmer que « *ces études suggèrent des corrélats électrophysiologiques communs au traitement de la syntaxe en musique et en langage* » (Hoch *et al.*, 2008: 37). Cependant, aucun travaux, jusqu'à présent, n'ont permis de montrer de tendance stable en ce qui concerne le plan sémantique. Tandis que la violation sémantique langagière met en évidence l'émission d'une onde N400 (onde négative autour de 400 ms ; Kutas & Hillyard, 1983 ; entre autres), la violation sémantique musicale provoque une onde positive autour de 800 ms (*Ibidem* : 38) ; cette différence « *suggère des ressources distinctes pour les traitements musical et sémantique* » (*Ibidem* : 43), même si le chant montrait des cas d'interaction entre les sémantiques musicales et linguistiques. Le caractère des activités pourrait expliquer les résultats contradictoires. En effet, il semble plus aisé de mettre en corrélation des tâches syntaxiques entre musique et parole que des tâches sémantiques.

Nous mentionnerons que l'approche modulaire proposée par Peretz & Coltheart (2003) peut être reconsidérée au regard des travaux de Chobert & Besson (2012) selon lesquels la pratique musicale chez l'enfant peut avoir des répercussions sur les autres capacités perceptives et cognitives :

*« l'apprentissage de la musique influence le traitement des unités de base du langage, les syllabes [allant] à l'encontre d'une conception modulaire et informationnellement encapsulée du langage qui serait indépendante des autres capacités perceptives et cognitives. Ces résultats indiquent que l'apprentissage musical ne facilite pas seulement le traitement des paramètres acoustiques tel que la durée ou la fréquence, mais améliore également la perception de paramètres linguistique plus abstraits tel que le VOT »<sup>29</sup> (Chobert & Besson, 2012 : 71).*

Ces résultats ont été confirmés par Kraus & Chandrasekaran (2010 : 600) qui affirment que « *l'entraînement musical a des effets sur la capacité auditive non-exclusivement avec la musique* »<sup>30</sup>. Ils ont également démontré que les musiciens ont de meilleures représentations (correspondance stimuli – réponse neuronale) en ce qui concerne la perception de la fréquence

---

<sup>29</sup> *Voice Onset time* : intervalle de temps entre le début du relâchement de la consonne et le début du voisement de la voyelle. Il intervient notamment dans la perception voisée ou non-voisée des consonnes occlusives, en anglais par exemple (Lisker and Abramson, 1964).

<sup>30</sup> « *music training improves auditory skills that are not exclusively related to music* »



fondamentale et les harmoniques du spectre de la parole (*Ibidem*).

Même si l'ensemble de ces résultats scientifiques donne l'impression du caractère immuable en place dans le traitement des informations par le cerveau, il nous reste à montrer l'impact de l'environnement sur ces structures apparemment fixes.

### 1.2.2 Pressions environnementales

Après avoir montré que les traitements de la musique et de la parole semblent posséder certains corrélats similaires mais que les différences semblent profondément marquées chez un normo-individu, nous présenterons des situations pouvant modifier cet ordre apparent. La plasticité cérébrale est l'un des mécanismes à l'œuvre dans cette problématique. Nous la définirons comme « *l'ensemble [des] processus de modification, de remodelage subtil du système nerveux* » (Schiffmann, 2001 : 14). Pendant longtemps, il a été considéré comme acquis que la plasticité cérébrale était une des particularités de la jeune enfance. Cependant, il est aujourd'hui communément admis que cette dernière est toujours active à l'âge adulte, notamment en ce qui concerne la mémoire et l'apprentissage (Schiffmann, 2001 : 15). Nous nous intéresserons aux cas de musiciens afin de comprendre en quoi l'apprentissage et la pratique musicaux peuvent induire des modifications à plusieurs niveaux sur le cerveau humain.

- **sur le plan fonctionnel** : de nombreuses études (Elbert *et al.*, 1995 ; Pantev *et al.*, 1998 entre autres) ont pu démontrer l'impact de la pratique musicale sur la représentation corticale des mains, par exemple. L'utilisation dissociée des doigts pour les instruments à cordes frottées (violon, alto, etc.) chez des droitiers augmente la représentation corticale des doigts dans l'hémisphère droit. Des résultats similaires ne sont pas trouvés dans l'hémisphère gauche car la main droite est moins sollicitée dans la tenue de l'archet (Münste *et al.*, 2002 : 474). De la même manière, les musiciens sont plus sensibles aux notes de musique jouées par l'instrument qu'ils pratiquent (Pantev *et al.*, 2001).
- **sur le plan anatomique** : il a été constaté que certaines aires corticales étaient plus étendues chez les musiciens en général, tels que le *planum temporale*, le corps calleux, l'air motrice primaire des mains et le cervelet (Münste *et al.*, 2002 : 474). Ces découvertes ont relancé le débat, sans y apporter de réponse précise, sur les différences structurales dues directement ou non aux capacités musicales.

A ce sujet, il est possible de trouver dans certains travaux de vulgarisation scientifique (Levitin, 2006 : 160 ; entre autres) que l'expertise musicale ferait passer certaines facultés d'analyse de l'hémisphère droit (centre dit de l'intuition) vers l'hémisphère gauche (centre dit de la logique).

Il est bien évident qu'une telle affirmation ne peut représenter la complexité à l'œuvre dans ce genre de traitement. De manière plus spécifique, il est possible d'avancer que « *les experts sont plus performants pour distinguer la dimension élémentaire du son musical (la hauteur, la durée ou l'intensité), mais lorsqu'il s'agit de comprendre des structures des extraits, la perception musicale des experts et des non-experts est proche* » (extrait de « *la symphonie neurale* », CNRS, 2012). La notion d'expertise a récemment été revue dans le domaine en ce qui concerne l'apprentissage implicite de la musique par exposition. Il semblerait que cette dernière réduise la distance qui peut exister entre des « *non-musiciens* » et des « *musiciens experts* » (Tillmann *et al.*, 2005). Bigand (2004 : 215) tient à préciser que :

*« l'auditeur musicalement entraîné ne se différencie pas nettement de celui sans formation musicale, cela signifie bien qu'un long apprentissage musical ne provoque pas de fortes réorganisations cérébrales. Une façon de dépasser cette apparente contradiction consiste cependant à envisager que l'apprentissage implicite entraîne des réorganisations cérébrales plus massives que les apprentissages explicites. Si tel est le cas, l'apprentissage explicite de la musique provoquerait bien des différences entre les cerveaux des musiciens et des non-musiciens, comme cela est actuellement fortement médiatisé, mais ces différences resteraient proportionnellement faibles par rapport aux immenses similitudes neuronales forgées dans les deux groupes d'auditeurs par des années d'apprentissages implicites de la musique. »*

La notion d'expertise fait toutefois intervenir, de manière indirecte, la question de la pratique. A ce sujet, il a été montré pour la musique que (annexe 2) :

*« l'activité (cérébrale) dans le groupe de (musiciens) professionnels était beaucoup plus concentrée spatialement et, dans le cortex primaire moteur, était beaucoup plus intense et confinée à la partie droite du cerveau (le côté qui contrôle la main gauche) tandis que chez les amateurs, l'activité était plus diffuse et présente dans les deux hémisphères. »<sup>31</sup> (Watson, 2006 : 532).*

Hyde *et al.* (2009 dans Kraus & Chandrasekaran, 2010 : 599) ont démontré que quinze mois d'entraînement musical intensif provoquent des changements structuraux dans les aires primaires auditives et motrices. En ce qui concerne l'apprentissage des langues :

*« quand un locuteur utilise sa L1, il a été montré une suppression paradoxale de l'activité neuronale dans le cortex associatif auditif et somatosensoriel [...] qui peut indiquer l'efficacité par laquelle la suppression de tout traitement local augmenterait la sensibilité d'un sous-ensemble de neurones au feedback sensoriel. »<sup>32</sup> (Simmonds, 2011: 470)*

---

31 « *the activity in the professional group was much more tightly focused spatially, and in the primary motor cortex was more intense and confined to the right side of the brain (the side which controls the left hand), whereas in the amateurs it was more diffuse and present on both sides.* »

32 « *when speakers use L1, it has been shown that there is paradoxical suppression of neural activity in auditory*

L'expert serait donc défini, en comparaison au non expert, par un traitement de l'information plus intense et plus confiné dans les régions concernées et également par la mise en place d'un système de réponse *feedback* sensorielle. Nous verrons plus tard que dans l'apprentissage des langues, la mise en place du système de *feedback* est liée à la notion de période critique. La pratique aurait donc une influence sur le traitement de l'information sonore en quantitatif et qualitatif.

Dans une perspective développementale, la possibilité d'une vision inter-perceptuelle s'inscrit dans la notion d'état synesthésique défini, chez le bébé, comme son incapacité à dissocier les signaux sensoriels de telle façon qu'une couleur peut avoir du goût, un son une odeur, etc. (Levitin, 2006 : 163) mais elle peut également apparaître chez l'adulte synesthète (Rosenthal, 2011). Nous comprenons bien que l'organisation encapsulée du monde qui nous entoure dépend de la maturation biologique cérébrale mais aussi l'expérience dans la dissociation du traitement des stimuli sensoriels.

Les cas d' « *oreille absolue* » (AP<sup>33</sup>) définie comme « *la capacité à identifier ou produire la hauteur tonale d'un son sans aucun point de référence* »<sup>34</sup> (Zatorre, 2003 : 692) apparaît comme un bon cas d'influence environnementale dans les capacités de représentation d'un système sonore. Il a été montré que recevoir un *input* musical précoce (avant l'âge de 6 ans) est une variable très répandue pour ne pas dire *sine qua non* au développement de capacités d'AP (Levitin & Zatorre, 2003). Zatorre (2003 : 693) précise qu'« *il est clair qu'un éveil musical est indispensable au développement de capacité d'oreille absolue mais il est encore plus important que cela se produise de manière précoce au cours de la vie*<sup>35</sup> ». Toutefois, il faut préciser que les individus présentant un profil AP ne possèdent pas nécessairement des capacités générales accrues en rapport à la musique. Cette capacité particulière étant fortement encapsulée. Il est important également de préciser que les musiciens tardifs présentent souvent une « *oreille relative* » (RP)<sup>36</sup>, capacité à identifier une note en référence à une autre (Gardner, 2009). La question de l'AP/RP apporte, de manière non attendue, des réponses en ce qui concerne les liens musique-parole. Nombreux sont ceux qui se demandent si les locuteurs de langues tonales (ex :

---

*and somatosensory association cortex [...] it may indicate efficiency, whereby overall local suppression increases the sensitivity of a subset of responsive neurons to sensory feedback.* »

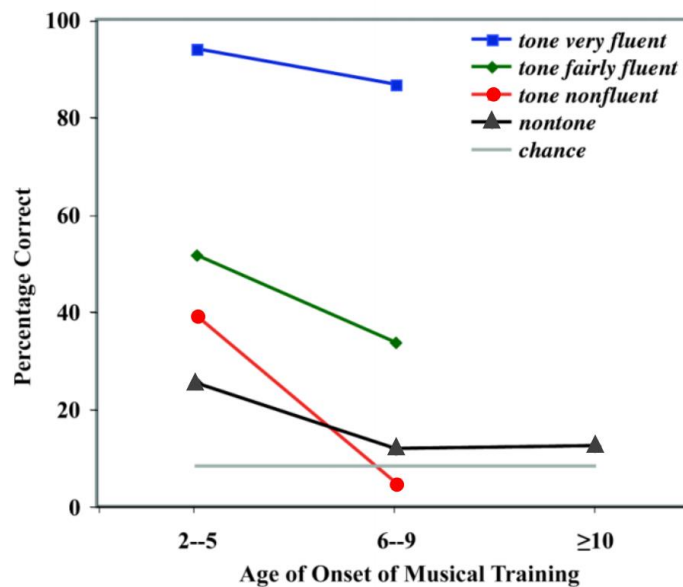
33 Absolute Pitch (AP)

34 « *the ability to identify or produce the pitch of a sound without any reference point* »

35 « *clearly, musical training is essential for AP [absolute pitch] to manifest itself, but most importantly, it must happen early in life* »

36 Relative Pitch (RP)

chinois) ont de plus grandes capacités en musique. Comme nous avons pu le voir, le paradigme dans lequel nous nous plaçons est en faveur d'un traitement partiellement modulaire des caractéristiques acoustiques musicale et linguistique. De ce fait, certains traitements peuvent être partagés et d'autres non, et dont les conditionnements restent encore très complexes. Même si à l'heure actuelle, aucune recherche n'a pu montrer que les capacités générales musicales de locuteurs à langues tonales étaient supérieures à celles d'autres types de langues. Il semblerait, pour le moment, comme l'a montré l'étude menée par Alexander *et al.* (2008 : 4) que « comparés aux locuteurs anglophones, les locuteurs sinophones (mandarin) discriminent mieux les mélodies musicales mais les identifient moins bien »<sup>37</sup>. Néanmoins, il a pu être démontré que les locuteurs de langues tonales, à variables égales (âge d'*input* musical, etc.), ont une prévalence accrue à présenter un profil AP (Figure 3 ; Deutsch & Dooley, 2009).



**Figure 3:** Pourcentage de réponses correctes pour le test AP en fonction de l'âge de premier entraînement musical et le niveau de performance à parler une langue à ton. « Les individus des groupes “tone very fluent” (parle couramment une langue à ton), “tone fairly fluent” (parle plutôt couramment une langue à ton) et “tone non-fluent” (parle une langue à ton mais de manière non courante) étaient tous d'origine d'Asie de l'Est. Les individus du groupe “nontone” (ne parle pas de langue à ton) étaient tous caucasiens et ne parlaient aucune langue à ton. La courbe “chance” (hasard) représente une réponse au hasard à la tâche. »<sup>38</sup> (Deutsch & Dooley, 2009 : 2400).

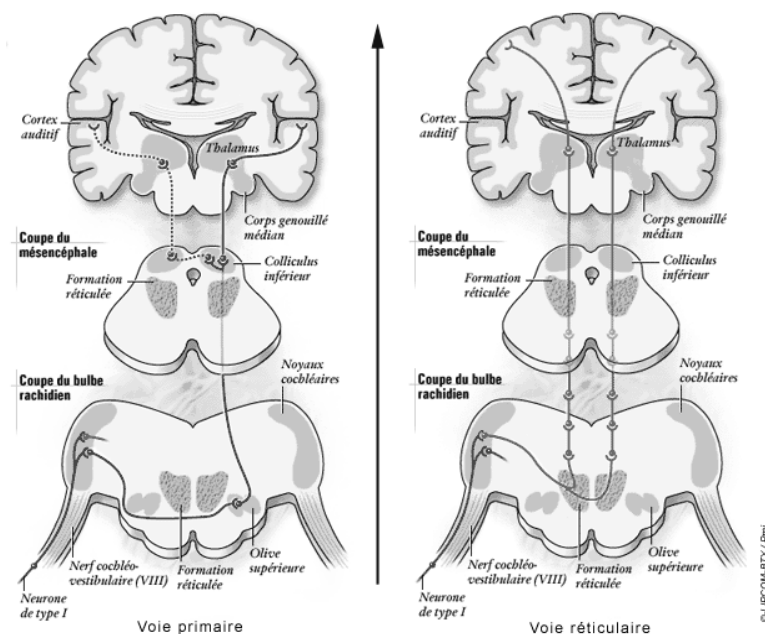
37 « relative to the English speakers, the Mandarin speakers more easily discriminated, but less easily identified, the music-melodies »

38 « Percentage of correct responses on the test of AP, as a function of age of onset of musical training and fluency in speaking a tone language. Those in groups tone very fluent, tone fairly fluent, and tone non-fluent were all of East Asian ethnic heritage. Those in group nontone were Caucasian and spoke only nontone language. The line labeled chance represents chance performance on the task. »

La question des liens musique-parole est maintenant plus claire en ce qui concerne les traitements neurocognitifs en tant que processus mais également dans une perspective de neuroplasticité. Il nous reste à élucider la question des émotions dans l'activation de certains de ces processus.

### 1.2.3 Émotions

Avant de traiter, en profondeur, la question des émotions dans la perspective comparée de la musique et de la parole, revenons sur quelques éléments neurophysiologiques. En effet, l'audition fait intervenir deux voies particulières (Figure 4) :



**Figure 4:** les deux voies de la perception auditive (med.univ-montp1, 2013).

– les voies auditives primaires :

« (i) La voie « essentielle » est une voie rapide qui se projette sur le cortex. Son rôle est de transporter l'information auditive discriminative.

(ii) La voie « de contrôle » est une voie plus lente, qui se projette sur le cortex auditif secondaire

–. L'information transportée est dépourvue de sélectivité fréquentielle » (Perrot, 2010 : 10)

– les voies auditives non-primaires/ secondaires :

« Les voies auditives non primaires interviendraient dans les phénomènes comportementaux impliquant des stimuli auditifs, tels que la réaction d'éveil, les réactions émotionnelles et les réactions végétatives » (Ibidem)

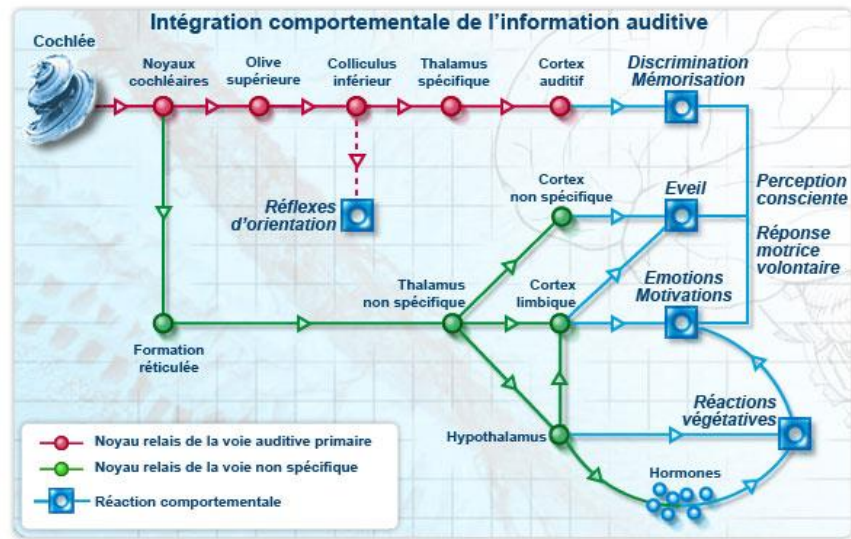


Figure 5: Intégration comportementale de l'information auditive (Cochlea.org, 2013).

Ainsi comme la Figure 5 le représente, il est facilement compréhensible que l'audition et les émotions soient étroitement liées à travers des réponses comportementales favorisées par une voie auditive non-primaire dédiée à ces réponses. Cette vision plurimodale fait écho au double encodage linguistique d'Ivan Fonagy (1983 : 14). Nous précisons que Fonagy mentionne à l'époque un traitement sériel tandis que nous nous positionnerons dans la perspective de Vaissière (2008) d'un traitement « en parallèle » de ces deux codes (Figure 6).

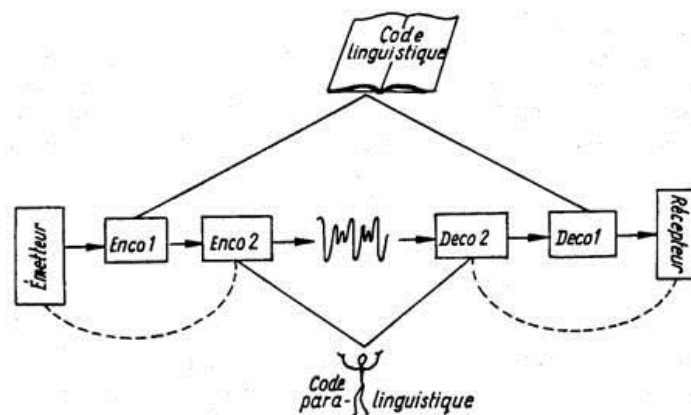


FIG. 2. — Schéma de la communication verbale à vive voix. — L'encodage linguistique est suivi d'un deuxième acte d'encodage à partir d'un « code naturel », pré-linguistique.

Figure 6: Le double encodage de la communication verbale (Fonagy, 1983 : 14).

Nous ne développerons pas plus en profondeur l'ensemble des mécanismes liés aux émotions car ils sont hors de propos dans notre réflexion générale. Nous tenons cependant à préciser que notre conception des émotions reprendra celle développée par Damasio (2010 : 136-7) définissant les **émotions** comme « *des programmes complexes et en grande partie automatisés d'actions qui ont été concoctés par l'évolution* », les **sentiments émotionnels** « *qui sont des perceptions composites de ce qui se passe dans notre corps et notre esprit quand nous éprouvons des émotions* » et les **états émotionnels** « *agrégat de toutes les réponses physiologiques provoquées par une émotion* ». La notion d'automatisme faisant sens chez Damasio (2010 : 198), en lien avec la conscience, lorsqu'il la définit comme « *un sentiment automatique, spontané et immédiat de soi, en tant que protagoniste de l'expérience de soi-même* ».

Dans cette perspective, il est clair que la musique n'est pas plus « *émotionnelle* » que la parole comme l'on pourrait présenter cette dichotomie à travers la *doxa* (Darsel, 2010: 15). Les émotions sont une réponse automatique au monde qui nous entoure comme le rappellent Deonna & Teroni (2009 : 25) : « *les émotions nous mettent en lien avec notre environnement et plus spécifiquement avec certains de ses aspects : ressentir une émotion, n'est-ce pas réagir à certains traits saillants des situations auxquelles nous faisons face ?* ». Notre positionnement sera en faveur d'un filtre émotionnel entre le monde qui nous entoure et nos représentations cognitives à travers l'espace corporel. Musique et parole sont deux technologies suscitant les réponses émotionnelles mais de manière différente. Tandis que la parole provoquera une réponse émotionnelle à travers une intentionnalité pragmatique par le « *sens* » majoritairement, la musique chez les non-musiciens impliquera les voies auditives non-primaires avec plus d'intensité. L'intentionnalité musicale portera directement sur le fait de générer des émotions tandis que l'intentionnalité discursive portera sur l'interaction avec autrui.

Les émotions sont à la fois un filtre entre nos représentations du monde et la manière dont nous le percevons, mais également un déclencheur de processus neurocognitifs. Même si les émotions participent à la manière dont nous cartographions le monde qui nous entoure (Frith, 2010 : chapitre 6), il nous reste à comprendre l'ensemble des mécanismes acoustiques et physiologiques qui s'y rattachent.

### 1.3 Propriétés acoustiques et perceptives

Dans cette section, nous définirons une partie des mécanismes à l'œuvre dans la perception et la production des sons qu'ils soient musicaux ou verbaux. Nous traiterons principalement des mécanismes physiques, physiologiques et phonético-phonologiques.

#### 1.3.1 Le son

Le son est un élément physique qui se manifeste par une onde résultant de la vibration mécanique d'un milieu ou support (Leipp, 1989 : 41). « *Quand les particules d'air sont au repos, elles se meuvent rapidement dans toutes les directions et sont équidistantes. Un choc les met en mouvement, créant des alternances de zones de raréfactions de l'air et de surpression locales* » (Vaissière, 2006 : 58).

Le son est un type de vibration particulière caractérisée par la possibilité de la percevoir par le système auditif. Le son n'existe dès lors que par l'existence d'un système auditif. Il est une formalisation créée par le cerveau en réponse à des molécules en vibration. Les deux principales mesures du son se réalisent en Hertz (Hz) et en décibels (dB) (Leipp, 1989 : 17).

Les Hertz mesurent la fréquence d'un son, c'est à dire la périodicité de récurrence de l'onde produite en fonction du temps en rapport au nombre de cycles vibratoires par seconde (Leipp, 1989 : 20).

La fréquence peut se calculer ( $f$ ) selon la relation mathématique :

$$f = \frac{1}{T} \text{ (Équation 1)}$$

T = période temporelle, c'est à dire le temps nécessaire pour que le phénomène se reproduise identiquement à lui-même.

Sur le schéma ci-après (Figure 7), il faut 0.5 secondes à l'onde pour faire une période. La fréquence représente alors  $f = 1/0,5 = 2$  Hz. La fréquence de l'onde est alors de 2 Hz.

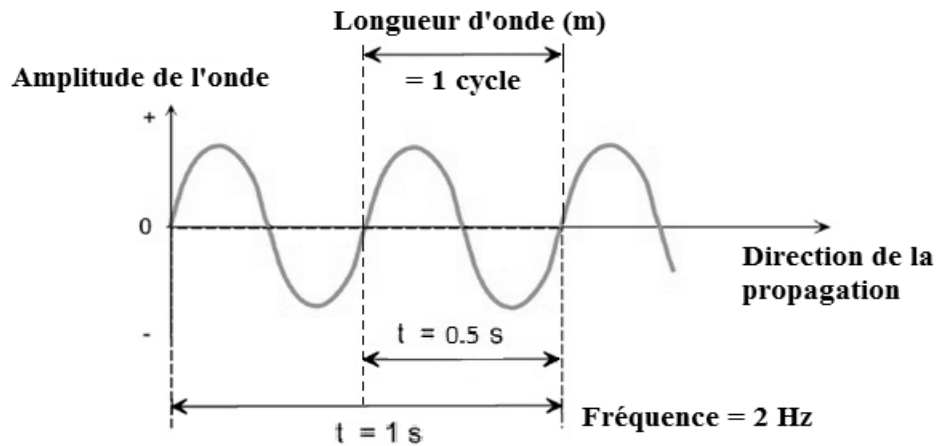
On peut également déduire la fréquence ( $f$ ) de l'équation :

$$f = \frac{c}{\lambda} \text{ (Équation 2)}$$

c étant la célérité exprimée en mètre par seconde et  $\lambda$  la longueur d'onde (en mètres).

Dans l'air ambiant, et sur Terre, on prend comme vitesse du son ou célérité une valeur moyenne de 340 mètres par seconde (Vaissière, 2006 : 59). Un son aigu sera caractérisé par une fréquence élevée et un son grave par une fréquence basse. Les sons perçus par l'audition humaine se situent entre 16-20 Hz et 16-000-20 000 Hz (Vaissière, 2006 : 85).





**Figure 7:** Calcul de la fréquence d'une onde (bbemg.ulg.ac, 2012).

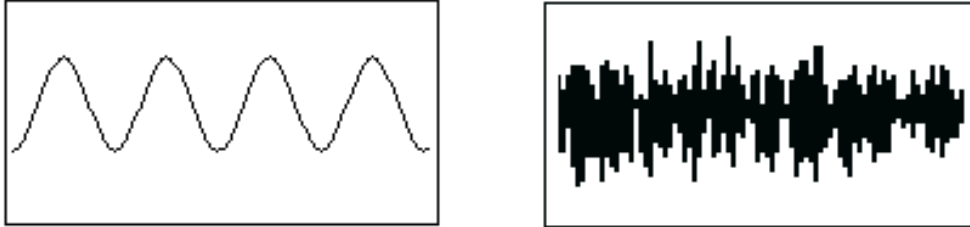
Les décibels (dB) permettent de mesurer l'intensité ou pression sonore d'un son. Dans le langage courant, on qualifiera cette donnée par « *c'est fort* », « *ce n'est pas assez fort* » pour ce son.

0 dB est considéré comme le minimum que l'oreille humaine peut percevoir appelé seuil d'audibilité et non comme le silence absolu pour obtenir une échelle avec des valeurs traitables. Toutefois, le seuil d'audibilité est généralement de 4 dB. Le seuil de douleur est de 130 dB, mais l'oreille peut subir des dommages à partir de 85 dB en fonction de la durée et des conditions d'exposition (Leipp, 1989 : 12-15).

Le timbre est la troisième caractéristique majeure du son. Bien qu'il soit déterminant en musique comme pour en « langue », les recherches actuelles n'ont pas réussi à comprendre l'ensemble des éléments qui le déterminent. En musique, le timbre est ce qui permet de reconnaître un instrument d'un autre. Dans ce cas, ce sont les propriétés des harmoniques qui interviennent (annexe 3, figures a et b) mais également la nature acoustique du son. Une note jouée au piano sera caractérisée par une attaque forte et une décroissance d'intensité rapide tandis que une note jouée au violon aura une attaque plus progressive puis une intensité régulière (Patel, 2007 : 29) (annexe 3, figure c).<sup>39</sup> Pour les langues, le timbre peut être ce qui permet de reconnaître l'identité d'un individu par le seul son de sa voix, c'est le principe de la reconnaissance perceptuelle par la voix (Conein, 1990 : 146) ; mais également ce qui permet de discriminer plusieurs phonèmes entre eux en fonction des résonateurs utilisés.

<sup>39</sup> Ces remarques ne prennent pas en compte un jeu particulier tel que le pizzicato, etc.

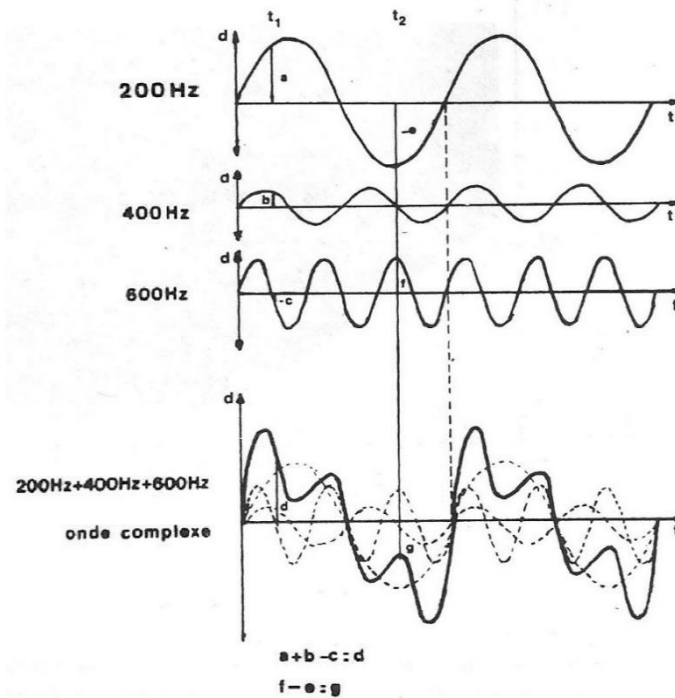
Sur un plan acoustique, le son est caractérisé par des ondes périodiques (Figure 8), c'est-à-dire que les vibrations se répètent identiquement à elles-mêmes à des intervalles réguliers (Chouard, 2001 : 87-98).



**Figure 8:** Représentation d'une onde périodique simple (à gauche) et d'une onde apériodique (à droite) en fonction de l'amplitude (en ordonnée) et du temps (en abscisse) (Ciral.ulaval.ca en ligne, 2012).

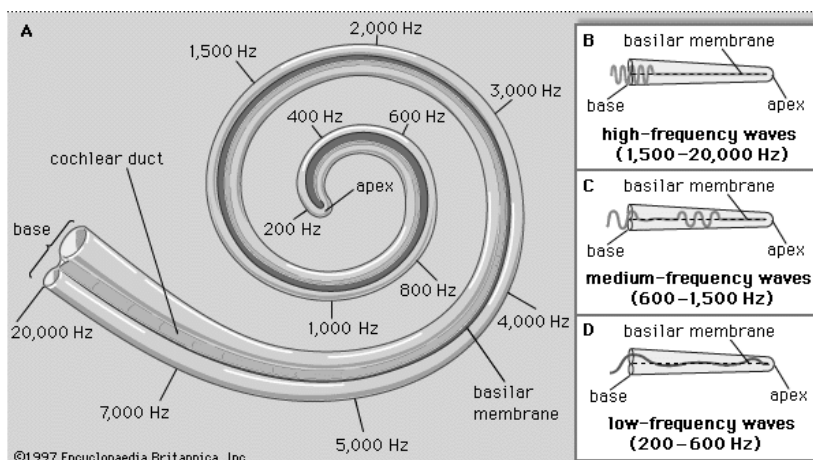
L'onde périodique s'oppose à l'onde apériodique. Cette dernière est alors irrégulière et anarchique. C'est que l'on appelle généralement un bruit (Chouard, 2001 : 99-101). Aubin rappelle en citant Carton que « *dans la parole, il n'y jamais de son simple : les vibrations périodiques y sont toujours composées et résultent de la superposition de nombreuses vibrations simples* » (Aubin, 1996 : 41-42). Cette simple distinction ne permet donc pas d'expliquer la nature de tous les sons.

Les ondes simples sont des ondes dont une seule fréquence audible et mesurable est constatée. Une onde complexe à l'inverse se compose d'une multitude de fréquences secondaires que l'on appelle les harmoniques (Leipp, 1989 : 25 ; Vaissière, 2006 : 59). Il est possible de spécifier deux types d'ondes complexes: celles périodiques et celles qui sont apériodiques. Les ondes complexes périodiques sont la somme de plusieurs fréquences pures, on peut alors décomposer l'onde complexe pour en déterminer les différentes simples (Figure 9). La Loi de Fourier montre que « *un phénomène périodique quelconque peut toujours se décomposer en une somme de sinusoïdes élémentaires (harmoniques), dont les fréquences respectives sont des multiples entiers de la composante la plus grave appelée fondamentale* » (Leipp, 1989 : 25). Les ondes complexes apériodiques au contraire sont une association anarchique d'ondes.



**Figure 9:** Représentation d'une onde complexe et ses harmoniques (Renard, 1978 : 6).

Les ondes complexes possèdent une fréquence fondamentale (F0) qui se trouve être la fréquence la plus basse de toutes. Les harmoniques sont quant à elles, pour les ondes complexes périodiques, des multiples entiers aux fréquences supérieures à celle de F0. L'ensemble de ces ondes est traité par l'oreille externe (pavillon, conduit auditif) puis l'oreille moyenne (tympan et osselets) et l'oreille interne (cochlée, vestibule, canaux semi-circulaires, etc.). La cochlée est l'organe ayant pour mission de transformer l'information acoustique en impulsion électrique vers le cortex auditif primaire. Nous précisons seulement que la cochlée possède la particularité de décomposer et traiter la hauteur des fréquences, à travers la répartition des cellules sensorielles (cellules ciliées) de l'organe de Corti, de manière dissociée (Figure 10).



**Figure 10:** Représentation de la répartition tonotopique de la cochlée. Les sons sont traités dans la cochlée des fréquences aiguës vers les graves respectivement de la base vers l'apex de la membrane basilaire (Encyclopaedia Britannica en ligne, 1997).

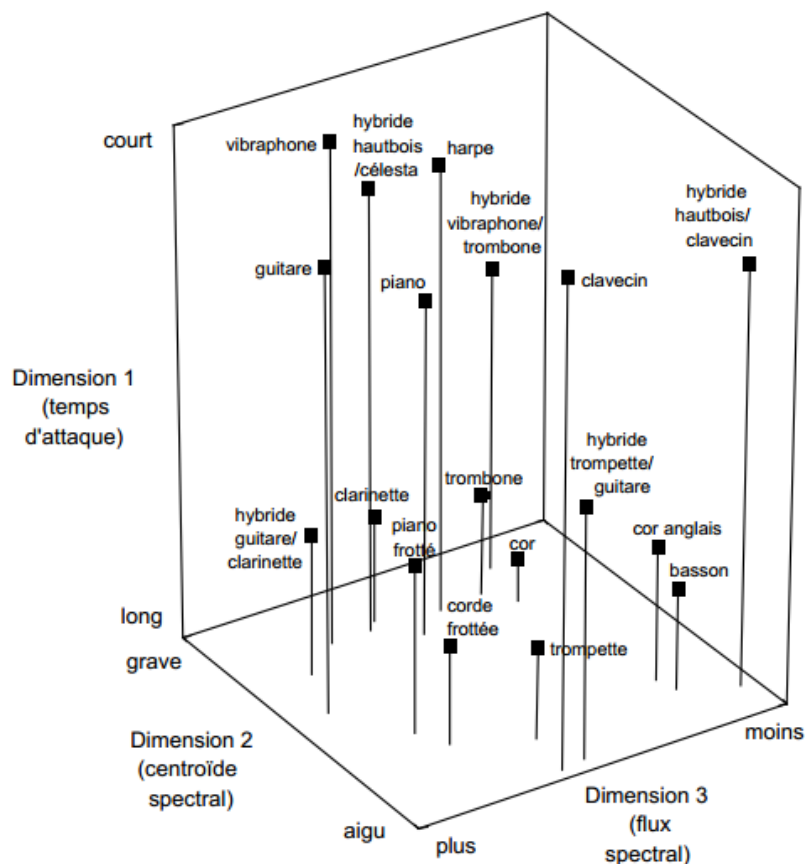
Cette particularité permet de mieux comprendre d'une part, que la cochlée et le cerveau traitent de manière primaire toute information sonore de manière égale, et d'autre part, que le cerveau reçoit une information cartographiée, de manière tonotopique, par la cochlée.

Après cette courte présentation du son en tant qu'objet scientifique acoustique et perceptif, nous nous intéresserons à la production de ces sons.

### **1.3.2 Instruments : appareil phonatoire et instruments de musique**

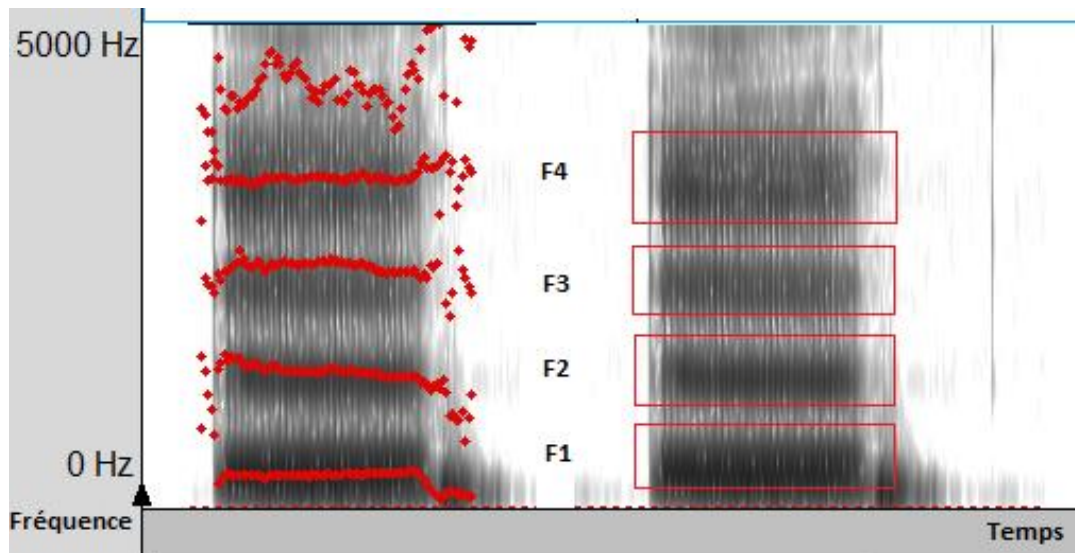
L'appareil phonatoire humain est souvent vu comme une « *machine à sons* », à l'instar de la phonétique articulatoire, comme le souligne Aubin « *[beaucoup de phonéticiens] se passent difficilement du recours à l'analogie instrumentale* » (Aubin, 1996 : 37). Leipp ajoute que « *parmi les instruments à vent, l'appareil phonatoire apparaît comme un des plus raffinés, pour peu qu'on sache en « jouer »* » (Leipp, 1989 : 285). L'appareil phonatoire et les instruments de musique ont en commun la présence d'un excitateur, d'une structure vibrante et éventuellement d'un ou plusieurs résonateurs. En parole, la hauteur tonale est le produit de la vibration des plis vocaux situés dans la cavité laryngée. La variation de la hauteur tonale fait partie des éléments qui définissent la prosodie en parole (avec l'accentuation et le rythme, entre autres). Le son produit par cette source sera modulé à travers les différentes cavités (résonateurs) de la phonation telles que la cavité nasale, buccale ou pharyngée. Certaines fréquences seront atténuées ou amplifiées selon le volume, la consistance de ces résonateurs qu'il est possible de faire varier en bougeant la langue ou les lèvres, par exemple. La théorie source-filtre (Fant, 1960 : 15) permet d'expliquer ce phénomène (annexe 4). La hauteur tonale, en musique, variera en fonction du type d'instrument, par exemple : longueur et diamètre de la corde, longueur du tube, occlusion des clés, etc.

La notion de timbre dépend intrinsèquement des propriétés physico-acoustiques. Il faut néanmoins prendre conscience que la notion de timbre en musique n'est pas strictement la même en parole. Il en va de même en ce qui concerne la note (éventuellement sonème) et le phonème. Le « *timbre* » en musique est un attribut multidimensionnel complexe et encore mal compris. Il intègre l'ensemble des caractéristiques acoustiques qui permettent de définir l'instrument qui a joué un son (*Figure 11*). Selon Schaeffer (1966 : 231): « *le timbre perçu est une synthèse des variations de contenu harmonique et de l'évolution dynamique; en particulier, il est donné dès l'attaque lorsque le reste du son découle directement de cette attaque* ».



**Figure 11:** Une représentation du timbre musical (McAdams et al. 1995 dans Caclin, 2004 : diapositive 2). Le timbre musical se compose d'au moins trois dimensions que sont le temps d'attaque (court/long), le centroïde spectral (aigu/grave) et le flux spectral (plus/moins).

Traditionnellement le « *timbre* » dans la parole fait référence à l'ensemble des caractéristiques perceptives qui permet de déterminer l'identité d'un individu au simple son de sa voix. Il dépend de nombreuses variables physiologiques telles que la physiologie des plis vocaux et des différentes cavités, entre autres. Nous nous intéresserons, dans notre étude, plus particulièrement à l'analyse des sons vocaliques. Dans cette perspective, le timbre peut faire référence aux différences caractéristiques acoustiques qui permettent une représentation spectrographique vocalique. Dans la théorie source-filtre, mentionnée précédemment, après la mise en vibration de l'air par une source (les plis vocaux), l'onde sonore va être transformée par les filtres (les résonateurs) et présentent des zones d'harmoniques renforcées, un renforcement d'énergie, que l'on appellera *formant* (Vaissière, 2006 : 60). Quatre formants sont pertinents pour les sons vocaliques et renseignent indirectement sur la forme des cavités qui les ont créés (Figure 12).

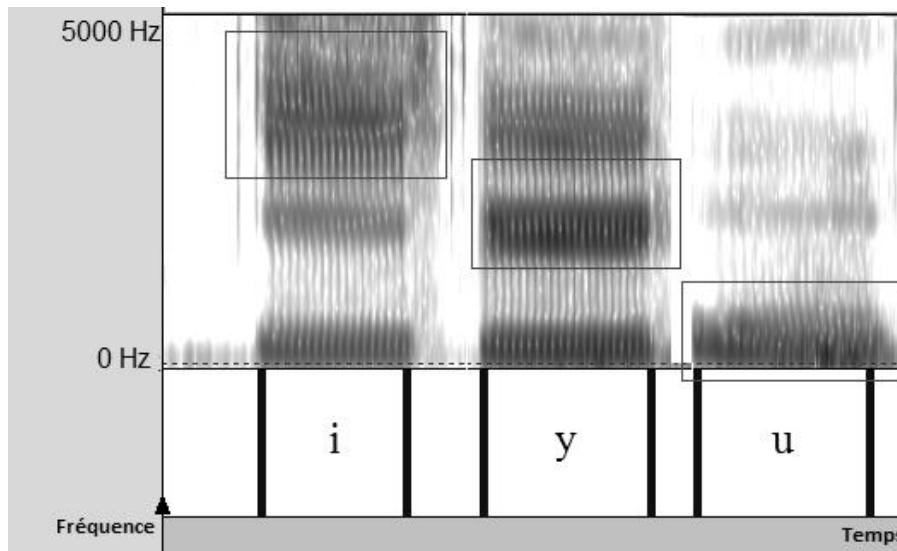


**Figure 12:** Représentation spectrale d'une voyelle. Analyse formantique du son vocalique isolé /ø / avec le logiciel Praat. Les 4 premiers formants ont été calculés automatiquement par le programme dans la représentation de gauche. Le [ø] du français standard se caractérise par quatre formants (plus ou moins) équidistants.

Vaissière (2006 : 62) rappelle que la simplification qui consiste à dire que F1 représente l'aperture, F2 l'antériorité et F3 la labialité (arrondissement) est erronée :

*« tous les formants sont modifiés par la forme générale du conduit vocal, mais certains sont plus sensibles que d'autres aux mouvements de certains articulateurs. F1 augmente rapidement lorsque la mandibule et/ou la langue s'abaissent. F2 est sensible à la position de la langue sur l'axe antérieur-postérieur, et tout autant à la configuration des lèvres lorsque la langue est massée vers l'arrière. F3 est particulièrement sensible à la longueur de la cavité antérieure lorsque la langue est massée vers l'avant. F4 est plus difficilement manipulable. »*

La hauteur spectrale pour un son vocalique représentera la répartition d'énergie qui compose le phonème (Figure 13). Ainsi à hauteur tonale/intonation égale, certains phonèmes vocaliques seront perçus comment plus « aigu » ou plus « grave » en fonction de la répartition d'énergie dans l'empreinte spectrale. Un /i/ sera généralement perçu comme plus « aigu » qu'un /y/ lui-même plus « aigu » qu'un /u/. Il ne faudra pas confondre la hauteur tonale qui dépend de la vibration des plis vocaux et la hauteur spectrale qui dépend des caractéristiques du spectre de la voyelle. La hauteur spectrale est traditionnellement appelée « acuité » en didactique des langues (Callamand, 1981 : 7).



**Figure 13:** Représentation spectrale de trois voyelles isolés /i y u/ avec le logiciel Praat. Les rectangles représentent les renforcements d'énergie. Le /i/ est caractérisé par un renforcement d'énergie dans les hautes fréquences, le /y/ dans les moyennes fréquences et le /u/ dans les basses fréquences.

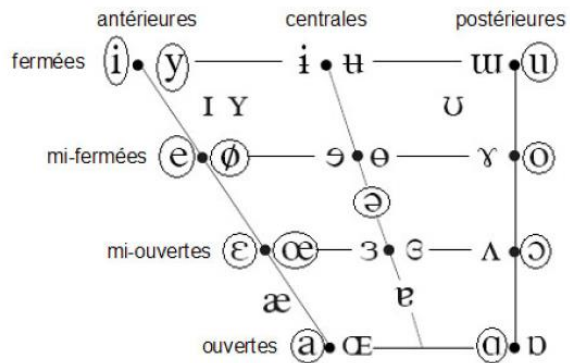
Il semble donc clair que l'appareil phonatoire humain repose sur des caractéristiques physico-acoustiques similaires à celle des instruments de musique. Néanmoins, les particularités et besoins de chacun de ces ensembles ont favorisé l'émergence de traits particuliers permettant la création d'un système complexe mais pertinent. Les caractéristiques acoustiques sont traitées de manière similaire pour la musique et la parole par le cortex auditif primaire. Cependant, les réseaux phonétique et phonologique ne donneront une prévalence qu'aux informations pertinentes au système global langagier.

Notre réflexion portant sur le français et la perception des voyelles, nous continuerons sur les particularités perceptives des sons vocaliques du français standard afin de montrer en quoi elles possèdent des caractéristiques intéressantes pour notre étude.

### 1.3.3 Particularités vocaliques du français standard

En français standard, nous pouvons dénombrer 10 voyelles orales et 3 voyelles nasales (Lauret, 2007), les voyelles peuvent être (Figure 14) :

- antérieures non arrondies /i e ε/,
- antérieures arrondies /y ø œ/,
- postérieures /u o ɔ/,
- basse /a/,
- nasales /ẽãõ/



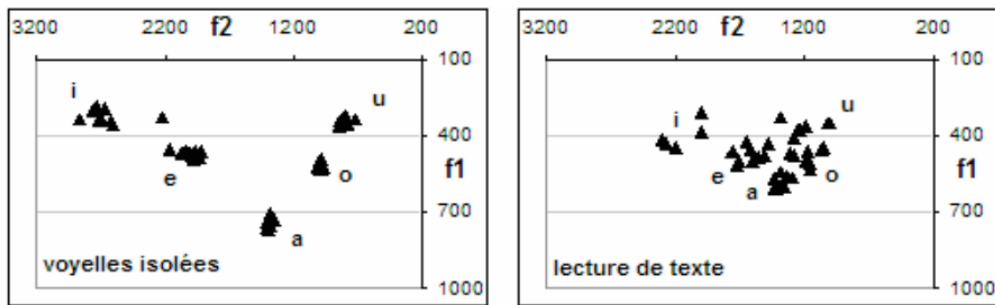
**Figure 14:** Représentation du trapèze général vocalique (à droite). Les phonèmes vocaliques du français sont entourés (Meunier, 2007 : 2).

Le /ɑ/ dit postérieur qui s'oppose historiquement au /a/ antérieur dans les paires minimales /pat/ (« pâte ») et /pat/ (« patte ») a disparu pour une grande partie de la population française métropolitaine au profit d'une opposition en termes de durée vocalique dans un contexte d'opposition explicite (Lauret, 1998 : 21). L'opposition entre le /œ̃/ et /ẽ/ n'est plus phonologisée à l'exception des locuteurs du français méridional, par exemple. Le /ə/ apparaît uniquement dans des syllabes non-accentuées (la syllabe finale en français). Le schwa est une réalisation phonétique mais n'est pas un phonème en français : pour preuve, il n'existe pas de paire minimale l'incluant. Une neutralisation dans les oppositions entre les voyelles moyennes /e ε/, /ø œ/, et /o ɔ/ est observée. En général, les syllabes ouvertes font apparaître les voyelles mi-fermées (/e ø o/) et les syllabes fermées, les voyelles mi-ouvertes (/ε œ ɔ/). En Français Langue Etrangère (FLE), les archiphonèmes /E/, /Ẽ/, et /O/ sont souvent utilisés.

Traditionnellement, les voyelles sont représentées selon des critères articulatoires (antérieur/central/postérieur en abscisse et fermé/mi-fermé/mi-ouvert/ouvert en ordonnée) telles que présentées dans l'Alphabet Phonétique International (IPA) (*International Phonetic Association*, 1999). Le terme « trapèze vocalique » apparaît lorsque l'on considère les onze voyelles du français et « triangle vocalique » en substituant le /ɑ/. Les représentations articulatoire et acoustique des voyelles montrent un grand nombre de similarités spatiales ; néanmoins, comme nous l'avons vu, le lien F1-Aperture et F2-Antériorité est erroné à

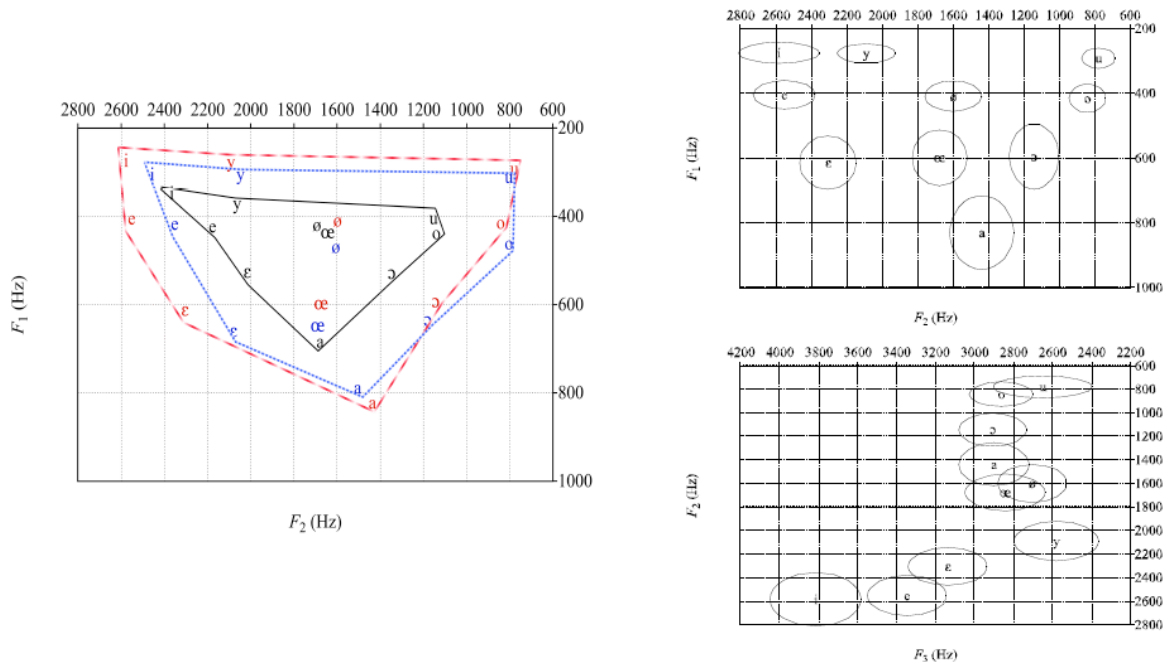


l'exception du premier pour les voyelles postérieures. Comme le rappelle Vaissière (2006 : 74) « il existe de grandes possibilités de compensation entre les articulateurs, souvent méconnues ».



**Figure 15:** Réalisations acoustiques des voyelles /i/, /e/, /a/, /o/, /u/ prononcées isolément (à gauche) et dans un texte (à droite). On observe la réduction considérable de la taille globale du système vocalique; les valeurs moyennes de formants observées en lecture de texte sont très éloignées des valeurs prototypiques (Meunier, 2007 : 9).

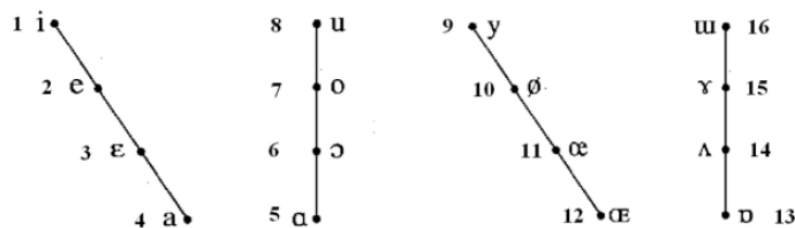
Les voyelles du français présentent des caractéristiques acoustiques particulières (Georgeton *et al.*, 2012). Elles montrent avant tout une stabilité significative dans le temps (pas de diphtongue, triptongue, etc.) et dans le profil formantique, et ce particulièrement pour les voyelles prononcées à l'isolée.



**Figure 16:** Comparaison des triangles vocaliques sur le plan F1-F2 de Gendrot & Adda (2005) (en noir), Calliope (Tubach, 1989) (en bleu) et Groupe Didactique (GD) (Georgeton et al., 2012) (en rouge) à gauche. Valeurs des formants sur un plan F1/ F2 et F2/ F3 de GD (à droite) (Georgeton et al., 2012 : 150).

Cette stabilité reste néanmoins constante si nous considérons le triangle/trapèze vocalique **dans son ensemble** dans de la lecture (Figure 15, Meunier, 2007 : 9) ou de la parole journalistique (Figure 16, Gendrot & Adda-Decker, 2005 dans Georgeton *et al.*, 2012 : 150) ou encore « *dans le contexte /pVI/ où VI est /e o u y ø/ et /pV2R/ où /i ε a ə oe/* » (Figure 16, Tubach, 1989 dans Georgeton *et al.*, 2012: 146).

Les voyelles du français sont considérées par Jones (1956) comme de bons prototypes de voyelles cardinales. Il les définit comme des sons vocaliques produits dans une position extrême de la langue (Figure 17).



**Figure 17:** Voyelles cardinales primaires (à gauche) et secondaires (à droite) telles que définies par Jones (Vaissière, 2011 : 53).

Il donna une définition articulatoire pour les voyelles cardinales primaires :

« [i], [a] et [u]. [i] sont les voyelles les plus hautes et antérieures que le conduit vocal humain peut produire, avec les lèvres étirées. [u] est réalisé avec la langue en position la plus postérieure et haute qu'il est possible dans la bouche, avec les lèvres protruses. [a] est prononcé avec la langue dans la position la plus basse et postérieure que possible. »<sup>40</sup> (Vaissière, 2011 : 53)

et une définition perceptive pour les voyelles cardinales secondaires comme :

« « perceptivement équidistantes » entre ces trois voyelles aux extrémités : [e], [ε] et [a] sont à égale distance perceptivement l'une de l'autre entre [i] et [a]; tout comme [o] et [ɔ], entre [u] et [a]. La distance auditive a été jugée comme directement liée à la hauteur de la langue. »<sup>41</sup> (Vaissière, 2011 : 53)

Kamiyama (2009 : 81) précise que :

« sur le plan acoustique, Vaissière (2006) distingue trois séries de voyelles, définies sur une base uniquement acoustique : 1) antérieures /i e ε y/, dans lesquelles la distance F1-F2 est supérieure à celle de F2-F3 ; 2) labio-postérieures /u o ɔ/, dans lesquelles la distance F1-F2 est inférieure à celle de F2-F3 ; 3) acoustiquement centrales /ø œ a/, dans lesquelles le F2 se situe à mi-chemin entre F1 et F3 »

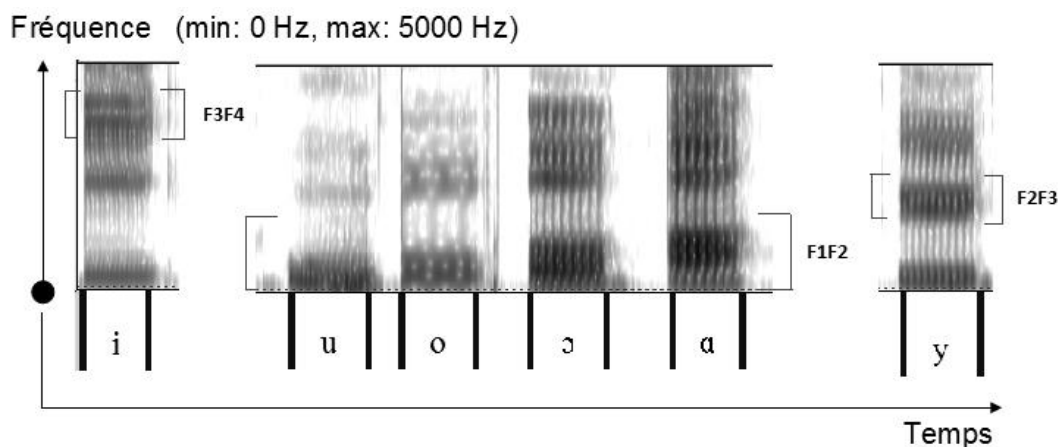
40 « [i], [a] and [u]. [i] is the highest and most fronted vowel that a human vocal tract (VT) can produce, with spread lips. [u] is realized with the tongue as "back" and as high as possible in the mouth, with pursed lips. [a] is uttered with the tongue as "low" and "back" as possible in the mouth. »

41 « « auditorily equidistant » between these three « corner vowels »: [e], [ε] and [a] are auditorily at an equal distance from each other between [i] and [a]; likewise for [o] and [ɔ], between [u] and [a]. The auditory distance was judged to be directly related to tongue height. »

En plus de faire apparaître un certain nombre de voyelles cardinales, le système phonéto-phonologique du français présente un certain nombre de voyelles focales. Elles sont définies comme des sons vocaliques présentant un rapprochement de deux formants (Schwartz *et al.*, 1997 : 255-286). Vaissière (2006 : 63) précise que :

*« la prégnance perceptive des formants peut être modifiée : lorsque deux résonances (deux formants) se rapprochent (ce qui est possible dans le cas d'une très forte constriction ou lorsque les cavités antérieures et postérieures ont un diamètre très différent), leur amplitude se renforce mutuellement, et avec elle leur prégnance perceptive ».*

On dénote six voyelles focales en français : /i y u o ɔ α/ aux focalisations F3F4 pour /i/, F2F3 pour /y/ et F1F2 pour /u o ɔ α/ (Figure 18).



**Figure 18:** Représentation spectrographique des six voyelles focales du français. Les renforcements d'énergie sont signalés par des crochets et les formants impliqués inscrits à côté. Le /i/ montre une focalisation F3F4, le /y/ F2F3 et /u o ɔ α/ F1F2.

La présence de ces voyelles focales a été mesurée sur un grand corpus de parole journalistique et notamment pour le /i/ et /y/ en comparaison respectivement à sept<sup>42</sup> et deux<sup>43</sup> langues (Gendrot *et al.*, 2008). Il apparaît donc que le système phonétique du français possède des caractéristiques particulières au niveau vocalique qui justifient un intérêt particulier au niveau segmental avant une intégration des variations dues à la coarticulation. Nous ne traiterons pas spécifiquement de ce dernier point car ce sujet se trouve en dehors de notre réflexion de thèse, mais par exemple (voir annexe 5) :

*« la présence d'un /R/ en coda aura tendance à allonger la voyelle (Léon, 2000, entre autres) et à augmenter la valeur du F1 et baisser la valeur du F2 (F2 dans la plupart des cas, mais F2, F3 pour le /i/) (Vaissière, 2007) » (cités dans Georgeton *et al.*, 2012 : 146).*

42 Allemand, anglais, arabe, espagnol, italien, mandarin et portugais

43 Allemand et mandarin

Les caractéristiques cardinales et focales des voyelles du français en font de bons objets d'étude pour notre travail de thèse qui se focalise sur un traitement acoustique et phonétique de l'*input* sonore, au regard des travaux de Chobert & Besson (2012) (voir section 1.2.1). Les deux caractéristiques mentionnées pourraient être des points acoustiques saillants pertinents en didactique de la prononciation en contexte FLE/S afin de travailler sur les capacités générales de traitement de l'*input* auditif.

En conclusion, nous avons pu voir de manière diachronique et synchronique que les avis restent partagés en ce qui concerne les liens musique-parole. Toutefois, nous nous positionnerons en faveur d'une origine commune de la musique et de la parole à travers un protolangage holistique, intentionnel, multimodal et musical. L'importance de la prosodie et la gestuelle aurait permis de passer d'une forme holistique de langage vers un système semi-computationnel en référence aux « *formulaic sentences* » (énoncés formulaires) d'A. Wray. Aucune preuve actuelle n'est assez forte en faveur d'un caractère sélectif de la musique sur le cerveau humain. La position générale, de nos jours, tend en faveur d'une vision plutôt modulaire (Peretz & Coltheart, 2003 : 688-91) mais non encapsulée (Chobert & Besson, 2012 : 71) de la musique et de la parole. Néanmoins, cette modularité ne porterait pas sur les ensembles eux-mêmes (module musique *versus* module parole) mais pour des traitements dissociés de la syntaxe, sémantique, prosodie, etc. En outre, le SSRIH apporte des preuves en faveur d'un traitement syntaxique commun pour la musique et la parole. L'ensemble de ces remarques est soumis aux réorganisations cérébrales à l'œuvre par plasticité cérébrale.

Sur le plan acoustique et perceptif, musique et parole profitent des mêmes phénomènes physiques et physiologiques. Sur le plan de la production sonore, ils profitent d'un même organe dans les cas du chant et de la parole. Néanmoins, leurs caractéristiques en système (phonologique et musical) ont favorisé l'émergence de traits distinctifs pertinents (cf. le timbre) pour un traitement efficace et rapide de l'information. Les caractéristiques des sons vocaliques du français, en outre leur stabilité, aspects cardinal et focal, justifient l'intérêt que nous leur porterons tout au long de cette thèse.

## 2. Musique et parole : un ou des domaines ?

### Résumé

Le chapitre « *Musique et parole : un ou des domaines ?* » s'articule autour de la question de la légitimité des recherches portant sur plusieurs domaines et en didactique des langues, plus particulièrement.

La section 2.1 tentera de définir les différences entre pluri-, inter- et transdisciplinarité en nous appuyant sur la notion de pluri-inter-transdisciplinarité définie par Claverie (2010). Nous verrons notamment que contrairement à la pensée monodisciplinaire, les pensées inter- et transdisciplinaires sont instables et toujours en action ce qui demande au chercheur une énergie supplémentaire pour ne pas abandonner un espace non-réservé dans la société. Cette réflexion nous mènera à remettre en question l'écart qui peut exister entre la musique et la parole notamment dans les domaines scientifique et académique.

La section 2.2 ouvrira la problématique des recherches pluri-inter-transdisciplinaires à différentes théories et approches comme le connexionnisme, les théories du chaos (Larsen-Freeman, 1997), la complexité (Morris, 1999), la « simplicité » (Berthoz, 2009) et enfin l'émergentisme (O'Grady, 2010), en tant qu'éclairages théoriques permettant de légitimer une approche variée dans la compréhension des phénomènes du vivant. Nous justifierons cette analyse par l'affirmation que « *le tout est plus que la simple somme des parties* » (O'Grady, 2010 : 1) et qui met en porte-à-faux une approche purement réductionniste.

Néanmoins, la section 2.3 contrebalancera la section précédente en se focalisant sur les pressions descendantes à travers le socioconstructivisme et les travaux de Piaget (1964) et Vygotski (1934/1997). En didactique des langues, nous proposerons une modélisation de l'approche par tâche en tâches sociale et d'entraînement (Narcy-Combes, 2006). Tandis que les premières motivent l'usage de la « langue » en cours d'acquisition par les interactions authentiques et la contextualisation de l'apprentissage ; les dernières permettent, si besoin, de renforcer des points spécifiques de manière individuelle et personnalisée. Ces points spécifiques seront traités en fonction du caractère non-linéaire des ruptures dans l'apprentissage de chacun mises en évidence lors des tâches sociales. Nous intégrerons l'enseignement/apprentissage de la prononciation dans ce cadre en montrant les spécificités.

## 2.1 Pluri-inter-transdisciplinarité

Dans cette section, nous définirons la notion de pluri-inter-transdisciplinarité afin de nous questionner sur la légitimité de ce type de recherche et dans notre cas sur les liens disciplinaires entre les sciences rattachées à la musique et à la parole.

### 2.1.1 Pluri-inter-transdisciplinarité

« *Penser consiste entre autres à organiser ses représentations du monde en catégories cognitives* » (Claverie, 2010 : 2). La catégorisation est l'un des fonctionnements humains les plus primitifs et toujours au cœur de nos sociétés modernes. Cette tendance, reliée au monde scientifique, s'est traduite en une vision structuraliste des éléments d'étude reposant sur « *un idéal ou des espoirs d'intelligibilité intrinsèque, fondés sur le postulat qu'une structure se suffit à elle-même et ne requiert pas [...] le recours à toutes sortes d'éléments étrangers à sa nature* » (Piaget, 1968 : 6). Elle a permis la compréhension et la segmentation des « *langues* » en différents champs d'analyse : phonétique, phonologie, sémantique, pragmatique, syntaxique, etc. en (re)créant des sous-ensembles catégoriels. Cette méthodologie porte un grand nombre d'avantages sur le plan analytique notamment. Cependant, les avancées technologiques de recherche en neurosciences de la cognition, depuis les années 1990, amènent vers une nouvelle conception des domaines du monde vivant. La didactique n'échappe pas à cette tendance.

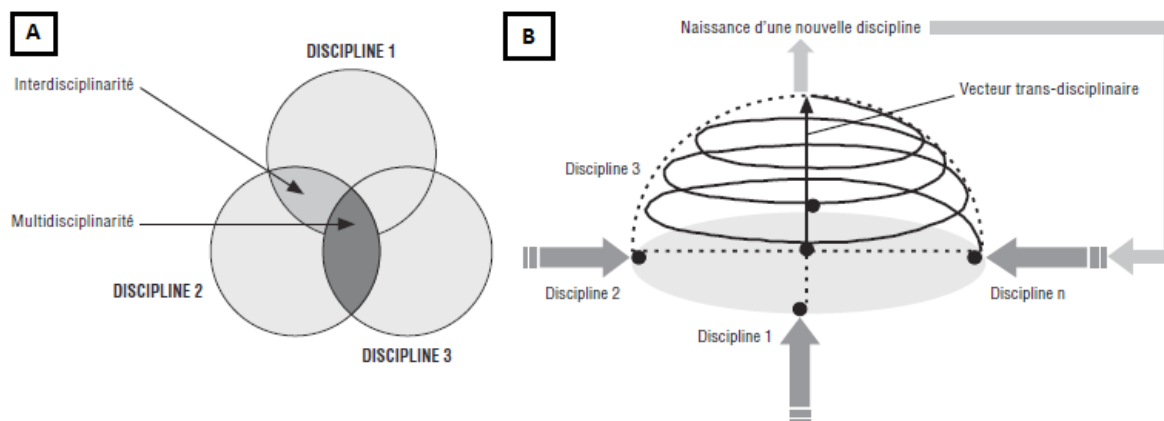
Tout porte donc à démontrer que le cerveau s'adapte aux contraintes environnementales grâce à la neuroplasticité (Shaw, 2001 : chapitre 1) lui permettant ainsi un traitement rapide, économique et donc efficace de l'information. Ces phénomènes dépendent en partie de notre conception du monde mais dépassent les clivages catégoriels conventionnels. Ces derniers prennent leur source dans la pensée prototypique que Bernard Claverie définit comme « *celle qui nous permet une conscience du monde à partir des groupes d'éléments que l'homme classe dans son système représentationnel* » (Claverie, 2010 : 2). A l'instar des supermarchés qui s'organisent en « *hiérarchie sémantique* » avec en position hypéronimique les rayons, puis les gondoles pour finir par les objets eux-mêmes et de ce fait « *l'espace vient lui-même structurer la représentation [pour faciliter la démarche d'achat]* » (Claverie, 2010 : 6). Les sciences n'échappent pas à ce processus de catégorisation. Même si nous pouvons reconnaître, dans l'organisation du monde savant, une puissance explicative et investigatrice, elle pose la question de la place de l'interdisciplinarité dans la recherche actuelle. Bien que le savoir se découpe en différentes catégories depuis Aristote (les pratiques, les sciences et les théologies) (Della Chiesa, 2007 : 144) en passant par le Moyen-Age, il a fallu attendre le XVII<sup>ème</sup> siècle et celui

des Lumières afin de voir émerger le spécialiste : « *expert d'une catégorie de savoir : c'est un être disciplinaire* » (Claverie, 2010 : 7). Cette hyperspécialisation fut notamment le catalyseur des grandes découvertes scientifiques du XX<sup>ème</sup> siècle. Le structuralisme de Saussure prendra dès lors la forme du structuralisme piagétien en faveur d'une « *conception des disciplines non plus linéaire mais circulaire [...] et qui sont donc de fait interdépendantes* » (Claverie, 2010 : 8). Nous pouvons ajouter que la science est restreinte par les structures sociales dont elle dépend (Callon & Latour, 1991), comme le précise B. Claverie (2010 : 8) pour la France :

*« le CNU, organe d'évaluation des universitaires, est organisé en sections, et présente une organisation hiérarchique très structurée, telle que celle du supermarché. Difficile d'y repérer des transversalités, et les nouveaux textes réglementaires tendent à confirmer le cloisonnement en disciplines, sous disciplines, spécialités, etc., sans trop privilégier les intersections, ni malheureusement privilégier l'apparition de nouveaux champs transversaux. »*

Nous pouvons donc nous demander si l'hyperspécialisation, la fragmentation et la certitude pourront continuer, à eux seuls, de pourvoir le matériel nécessaire à la compréhension des phénomènes complexes de notre monde, à l'instar de Morin, selon qui l'hyperspécialisation isole la discipline par rapport aux autres et par rapport aux problèmes qui chevauchent les disciplines (Morin, 1994 : en ligne). Étrangement, ce sont les sciences dites « *dures* » qui ont ouvert la voie en théorisant sur l'existence de matière dont nous ne sommes pas encore capable de mesurer l'existence mais seulement l'effet (Bourguignon, 1997 : en ligne) ; donnons pour exemple la matière noire ou sombre (Milgrom, 2002 : 42-52).

A l'heure où certains parlent de la nécessité de l'interdisciplinarité (Figure 19), nous pensons que la notion est déjà trop limitée et celle de pluri-inter-transdisciplinarité (Claverie, 2010 : 1-14) paraît plus complète. La pluridisciplinarité (ou multidisciplinarité) est intéressante dans le fait qu'elle associe des disciplines qui évoluent, de manière indépendante, mais dans une organisation commune. L'interdisciplinarité « *élabore un formalisme suffisamment précis pour permettre d'exprimer dans ce langage unique les concepts [...] d'un nombre plus ou moins grand de disciplines* » (Claverie, 2010 : 11).



**Figure 19:** Deux modélisations (A et B) de l'interdisciplinarité, multidisciplinarité et transdisciplinarité (Della Chiesa, 2007 : 148-149).

Cette approche du monde ne remet, en aucun cas, en cause les travaux structuralistes effectués jusqu'à présent car en effet, le réductionnisme instrumental a permis et permet encore de comprendre les éléments minimaux qui constituent la complexité des systèmes supérieurs. Les enjeux de la pluri-inter-transdisciplinarité s'articulent autour de l'importance « *d'élaborer des interclasses [...] qui n'est plus une contrainte, mais une nécessité pour une pensée organisée, créative et ouverte à de nouvelles acquisitions, à de nouvelles conceptions, à de nouvelles représentations* » (Claverie, 2010 : 12). Toutefois, ce travail n'est pas aisé car il requiert une adaptation au croisement de réseaux remettant en question sa propre représentation du monde. Cet effort constant, s'il n'est pas motivé, tendra vers un retour à une vision cognitive simple « *monodisciplinaire* ». Nous devons cependant opposer les démarches pluri- et interdisciplinaire qui peuvent créer de nouveaux mécanismes de pensée à la transdisciplinarité (Della Chiesa, 2007 : 144-148) qui demande un effort constant de dépasser les niveaux précédents tout en s'appuyant continuellement sur ces derniers. De plus, la transdisciplinarité implique la naissance d'une nouvelle discipline (Figure 19, B) dans les champs transdisciplinaires qui peuvent ainsi « *contribuer à l'évolution qui leur a donné naissance, mais également faire évoluer en retour les disciplines dont ils sont issus* » (Della Chiesa, 2007 : 149). Comme conclusion de cette première approche de la pluri-inter-transdisciplinarité, nous pouvons citer que :

*« penser la transdisciplinarité est donc une tâche instable [...] La pensée monodisciplinaire est rassurante, la pensée pluridisciplinaire et la pensée interdisciplinaire sont plus difficiles, mais elles peuvent s'appuyer sur une pensée stabilisée. La pensée transdisciplinaire est toujours en action, toujours en dépense d'énergie, toujours donc facile à abandonner d'autant qu'elle ne s'inscrit pas bien dans les espaces que nous réserve la société »* (Claverie, 2010 : 13).



Dépasser la monodisciplinarité n'est plus un simple désir marginal mais repose sur la nécessité actuelle de prendre en compte l'Humain et ses productions sociales dans un tout économique, culturel et sociobiologique. Ils s'intègrent dans des systèmes complexes qui interagissent perpétuellement. La psychologie et la physique moléculaire ou astronomique ont « *étrangement* » été les précurseurs dans la nécessité de dépasser les clivages catégoriels afin d'aller puiser dans la pluri-inter-transdisciplinarité les éléments dont les technologies ne fournissent pas encore de réponse claire. Le besoin de reconnaissance des sciences dites « *molles* » les a éloignés de cette perspective. Les neurosciences actuelles permettent de remettre en question ces acquis afin de réorganiser le monde dans une vision peut-être déstabilisante mais sûrement enrichissante.

### **2.1.2 Musique et parole : une science ou des sciences ?**

Pour étudier les sons des langages qui nous intéressent, il est impératif de dissocier deux champs de connaissance : les sciences du langage et la musicologie. Les objets étudiés par ces deux domaines proviennent certainement de la même origine même s'il n'y a pas de consensus actuellement comme nous avons pu le voir précédemment. Il semblerait donc que musique et parole trouvent une même origine communicative (annexe 6). Cependant l'étude du langage verbal est appelée « *science du langage* » ou « *linguistique* » (de manière hypéronymique) tandis que nous parlons de « *musicologie* », terme formel. Nous noterons cependant que depuis quelques années, la notion de « *science du langage musical* » fait son apparition, à l'instar du « *Premier Symposium International sur les Sciences du Langage Musical* » qui s'est déroulé à l'Université de Provence, Aix-en-Provence, les 15 et 16 Mai 1998. Nous ne pourrions, selon nous, qu'encourager la démocratisation de ce terme.

La phonétique est l'une des « *branches les plus anciennes de la linguistique* » (Vaissière, 2006 : 29). C'est un sous-domaine des sciences du langage parmi la pragmatique, la morphologie, la sémantique, etc. Elle est vue comme la science qui étudie les sons utilisés dans la communication verbale. Il serait plus juste de dire « *les* » phonétiques étant donné que Vaissière reconnaît l'existence d'au moins neuf branches de la phonétique (Vaissière, 2006 : 31-35): générale, articulatoire et physiologique, perceptive (auditive), acoustique, orthophonique et didactique, développementale, statistique et computationnelle, clinique et la neurophonétique. Bien que phonétique et phonologie aient longtemps été des domaines bien distincts dans la forme, nous remarquons que « *le rapprochement entre phonéticiens et*

*phonologues s'est concrétisé depuis plusieurs années par l'organisation régulières de rencontres internationales* » (Vaissière, 2006 : 27). La phonologie a pour base la phonétique mais portera son analyse sur la fonction des sons dans la communication verbale, de la manière dont les sons s'organisent pour établir une fonction distinctive. « *Le cercle linguistique de Prague conseillait de séparer clairement l'étude des sons, objet de la phonétique, de l'étude du système, objet de la phonologie* » (Vaissière, 2006 : 21). Nous reviendrons plus tard sur la dichotomie phonétique/phonologie.

L'entrée « *musicologie* » du Larousse (2013) offre la définition suivante : « *discipline qui étudie de manière scientifique et historique tout ce qui relève de la musique.* ». Dans notre démarche, l'aspect historique concernant les compositeurs et les évolutions des styles ne seront pas développés. Contrairement aux sciences du langage, la musique, qui s'est constituée comme discipline au début du XX<sup>ème</sup> siècle, n'a pas réussi à s'inscrire dans les sciences humaines. Campos, Donin & Keck (2006) dressent un panorama diachronique expliquant ces « *rendez-vous manqués* ». Néanmoins, l'homme étant très facilement « *mélomane* »<sup>44</sup>, la musique a toujours été intégrée aux autres domaines des sciences humaines telles que la sociologie, l'ethnologie, etc. Les domaines de la rééducation ont également montré une certaine appétence dans l'utilisation de la musique dans les Art Thérapies, l'apprentissage, etc. C'est à ce titre que nous constatons que la musique et la parole en tant que langages s'intègrent dans les mêmes domaines des sciences cognitives et de la didactique, entre autres. Nous trouvons, par ailleurs, les termes de syntaxe et sémantique musicale, par exemple. Depuis quelques années, la musicologie est revenue sur le devant de la scène avec la prolifération croissante de travaux sur le cerveau musicien. La « *European Science Foundation* » (ESF) fait même état de « *la musicologie et le cerveau humain, deux priorités stratégiques* » (La lettre de l'INSHS, N°21, Janvier 2013)<sup>45</sup>.

Il apparaît donc que, de manière historique, musique et parole ont gardé un lien intrinsèque de par leur matière acoustique. La poésie et la chanson en sont un bon exemple. Néanmoins, dans la recherche scientifique, les sciences du langage et la musicologie restent deux domaines distincts qui ne s'entrecroisent que dans de rares occasions. Toutefois, les neurosciences de la cognition semblent remettre en cause la notion de disciplinarité avec l'émergence d'un grand nombre de recherche et de travaux pluri-inter-transdisciplinaires (cf. 1.2.1). Nous préciserons

---

44 Dans son acception la plus hypéronymique signifie « qui aime la musique »

45 [http://www.cnrs.fr/inshs/Letres-information-INSHS/lettre\\_infoINSHS\\_21.pdf](http://www.cnrs.fr/inshs/Letres-information-INSHS/lettre_infoINSHS_21.pdf) (consulté le 03/08/13)

également que cette opposition est affaiblie dans des domaines comme la didactique, par exemple, de par sa nature intrinsèquement interdisciplinaire. Il nous reste à comprendre en quoi les travaux pluri-inter-transdisciplinaires nous permettent, en tant que chercheurs, de nous positionner dans un nouveau paradigme épistémologique.

## 2.2 L'émergentisme

Dans cette section, nous mettrons en avant différentes théories qui permettent de dépasser un cadre purement réductionniste qui peut, très souvent, tendre vers une monodisciplinarité. Nous mettrons particulièrement en avant les atouts du connexionnisme, des théories du chaos, de la complexité, de la « simplicité » et enfin des théories émergentistes dans le cadre de notre problématique générale.

### 2.2.1 Connexionnisme, théories du chaos, complexité et

#### « simplicité »

Le passage d'une approche purement réductionniste à une approche connexionniste a permis de prendre en compte les phénomènes ascendants et descendants interprétant le(s) langage(s) comme un ensemble de processus. Nous noterons, néanmoins, que certains auteurs admettraient une part de réductionnisme dans le connexionnisme (Memmi, 1995 : 262-265). Ces processus interagissent au sein même du système complexe mais également avec des variables environnementales, individuelles et dynamiques. La notion de complexité dans l'approche des sciences humaines apparaît pour la première fois dans les travaux d'Edgar Morin en appliquant les lois des systèmes aux sociétés (Morin, 1999 dans Ait Abdelmalek, 2004 :102). Un système complexe se définit par un comportement conséquent d'un grand nombre de composantes qui réagissent entre elles et dont émerge une nouvelle organisation (Davies, 1988 : 22). Le langage verbal est vu comme un système complexe par le fait qu'il intègre différents sous-systèmes : phonologie, lexique, syntaxe, etc.

La complexité se définit selon deux autres grandes propriétés qui s'intègrent dans les théories du chaos (Larsen-Freeman, 1997: 143-146) :

- **la non-linéarité** : les effets au sein d'un tel système sont toujours disproportionnés par rapport aux effets initiaux.
- **le chaos et l'imprévisibilité** : l'évolution des systèmes complexes est soumise à des changements d'états déclenchés de manière imprévisible. Nous pouvons simplement prédire que cela arrivera mais pas à quel moment. Ces moments chaotiques sont entrecoupés d'épisodes où règnent ordre et régularité.

De ces deux premières propriétés découle ce qui est communément appelé « *l'effet papillon* » caractérisé dans l'acquisition des langues par « *de légères différences dans l'input peuvent rapidement prendre un effet majeur dans l'output* » (Gleick, 1987 : 8). Ces effets tendent à être réduits à mesure que le système évolue car les biologistes avancent que les systèmes ouverts s'organisent du désordre vers l'ordre au fur et à mesure des restructurations (Larsen-Freeman, 1997 : 144). Il est en effet possible de voir émerger une tendance : désordre vers ordre dans une vision globale, mais de manière locale il est impossible de prédire une quelconque évolution. Cette évolution s'articule autour d'attracteurs (*Ibidem* : 146) qui influencent une reproduction du mouvement, tel un pendule, mais sans qu'aucun cycle ne soit totalement le même que le précédent. En ce sens, Rutherford (1987 : 37) propose de comparer la « *langue* » à un organisme plutôt qu'à une machine.

Il est, dès à présent, indispensable de rappeler que, dans cette perspective, le langage est vu comme un processus dynamique qui évolue chaque fois qu'il est utilisé. Cette caractéristique est reprise dans les travaux portant sur la théorie des systèmes dynamiques (« *dynamic system theory* » ; DST) (Van Geert, 1994 ; Verspoor, de Bot & Lowie, 2011; entre autres). De Bot, Lowie & Verspoor (2007 : 8) définissent cette approche par :

« *la propriété principale d'un système dynamique est ses changements au cours du temps qui sont exprimés avec l'équation fondamentale  $x(t+1)=f(x(t))$ , pour toute fonction décrivant comment un état  $x$  au moment  $t$  se transforme en un nouvel état  $x$  au moment  $t+1$ .* »<sup>46</sup>

Néanmoins, nous ne développerons pas plus les caractéristiques de la DST qui nous paraît pertinente principalement dans des études longitudinales, ce qui ne sera pas le cas pour ce travail de thèse. Cet état changeant, présent dans la DST, reflète le caractère non-linéaire du développement langagier qui ne peut plus être simplifié par l'apprentissage d'un élément puis d'un autre. L'*intake* tend perpétuellement vers l'ordre à travers des périodes de désordre stimulées par l'émergence de caractéristiques engagées par l'*input* voire l'*output*. Une étude menée par Verspoor *et al.* (2012) a permis de mettre en évidence, dans l'apprentissage de l'anglais comme langue étrangère, que l'apprentissage s'organise autour de progressions et régressions. Les niveaux débutants montraient de grandes restructurations lexicales, les niveaux intermédiaires des restructurations syntaxiques et les niveaux avancés des restructurations tant lexicales que syntaxiques. Chaque apprenant fait apparaître un patron de progression individuel et différent. Les auteurs rappellent notamment que :

---

46 « *the major property of a DS (Dynamic System) is its change over time, which is expressed in the fundamental equation  $x(t+1) = f(x(t))$ , for any function describing how a state  $x$  at  $t$  is transformed into a new state  $x$  at time  $t+1$ .* »

*« le langage se développe dans tellement de dimensions au même moment et il y a un tel degré de variation dans la manière où les apprenants agissent que nous pourrions remplacer la métaphore de la “règle” par celle d’un “brocoli” ; plutôt que de regarder en terme de durée, nous devrions regarder aux changements dans toutes les directions et mesurer la taille de la “tête” en s’assurant que toutes les branches se sont développées. »<sup>47</sup> (Verspoor et al., 2012 : 258)*

La fossilisation serait, dans cette perspective, la fermeture du système, par exemple syntaxique ou phonologique, autour d'un attracteur fixe. Ce dernier serait en parti la résultante de pressions sociopsychologiques (Larsen-Freeman, 1997 : 152, 161). D'autre part, la complexité semble montrer, à travers l'approche connexionniste, que l'*input* peut être à l'origine de l'émergence d'une complexité du système supérieur à la complexité de l'*input* lui-même (Mohanam, 1992). Dans cette perspective, chaque individu crée sa propre « *langue* » ou du moins « *discours/parole* ». Cependant, Larsen-Freeman rappelle que cette construction est limitée par un cadre global (biologique, neurologique, etc.) mais aussi que l'interaction impose une adaptation réciproque et donc construire un langage commun. Nous rappellerons l'exemple des cibles perceptives du français réalisables selon plusieurs stratégies articulatoires. La créativité individuelle intervient également dans cette construction langagière (Larsen-Freeman, 1997 : 154). Une fois encore, la complexité renforce l'importance de la pluri-inter-transdisciplinarité en remettant en cause la capacité scientifique de mesurer et prévoir la « *langue* » comme système linéaire. Les théories du chaos et de la complexité encouragent la complémentarité et l'inclusion là où certains linguistes peuvent voir opposition et exclusion (*Ibidem.* :158) comme le présentait Edgar Morin en affirmant que « *la pensée complexe est tout d'abord une pensée qui relie* » (dans Vallejo-Gomez, 2008 : 249).

Plus récemment, la notion de « *simplicité* » développée par Berthoz (2009) remet au centre des processus cérébraux l'importance d'un traitement rapide et efficace de l'information. Il définit la « *simplicité* » comme :

*« une nécessité biologique apparue au cours de l'évolution pour permettre la survie des animaux et de l'homme sur notre planète : malgré la complexité des processus naturels, le cerveau doit trouver des solutions, et ces solutions relèvent de principes simplificateurs » (Berthoz, 2009 : 17).*

---

47 « *language develops in so many dimensions simultaneously and there is such a great deal of variation in the way learners behave that we might have to replace the yardstick metaphor with a broccoli one; rather than looking for length we should look at change in all directions, and measure the size of the head, making sure that all sides have developed.* »

Elle intégrerait six paramètres essentiels au vivant (Berthoz, 2009 : 20-4) :

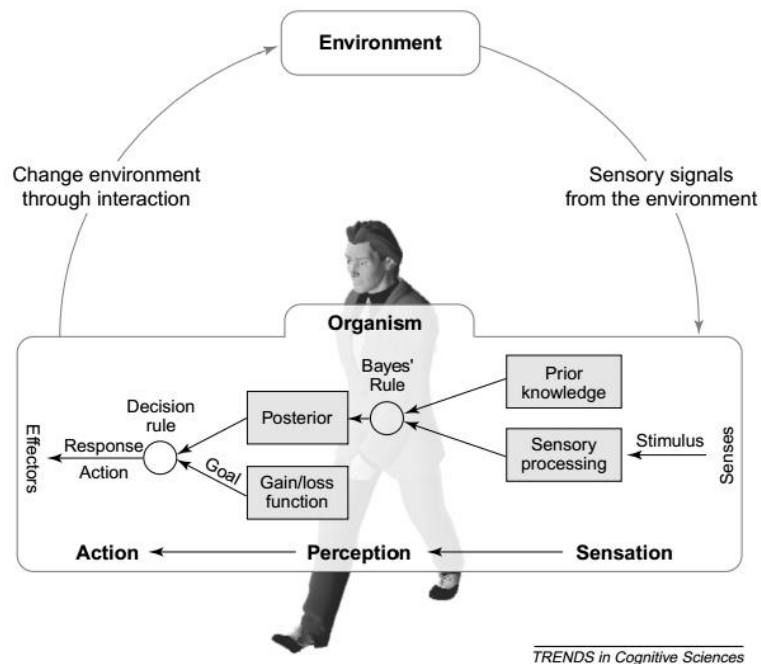
- **la séparation des fonctions et la modularité** : la diversité est un facteur de « *simplexité* »,
- **la rapidité** : la rapidité exige des solutions élégantes, pas nécessairement simples, mais efficaces,
- **la fiabilité** : les mécanismes neuronaux du cerveau et de ses annexes doivent être hautement fiables,
- **la flexibilité et l'adaptation au changement** : un organisme doit pouvoir s'adapter au contexte en compensant ses déficits et tirant profit de ses forces,
- **la mémoire** : les souvenirs d'une expérience passée doivent pouvoir être mis à profit pour une action présente ou tenter de prédire un fait à venir,
- **la généralisation** : un organisme doit être capable de programmer comme unité un ensemble de fonctions afin de limiter le poids d'actions potentiellement répétables.

L'un des paradigmes actuels de la perception repose sur la théorie bayésienne. Ce n'est qu'après la mort de Thomas Bayes, un prêtre non conformiste, que son théorème basé sur le principe qu'« *étant donné un certain phénomène A que nous souhaitons connaître et une observation X se rapportant à A, le théorème de Bayes nous dit de combien nous devrions actualiser la connaissance de A en fonction de la nouvelle observation X* » (Frith, 2010 : 165-6). Nous ne présenterons pas en détail l'ensemble des caractéristiques de cette vision probabiliste du cerveau ; néanmoins, Dehaene (2012 : diapositive 5) résume cette approche par :

*« l'inférence Bayésienne est une théorie mathématique simple qui caractérise le raisonnement plausible en présence d'incertitudes. Elle rend bien compte des processus de perception: étant données des entrées ambiguës, notre cerveau en reconstruit l'interprétation la plus probable. Nos décisions combinent un calcul Bayésien des probabilités avec une estimation de la valeur probable et des conséquences de nos choix. L'architecture du cortex pourrait avoir évolué pour réaliser, à très grande vitesse et de façon massivement parallèle, des inférences Bayésiennes. L'algorithme utilisé pourrait expliquer la manière dont notre cerveau anticipe sur le monde extérieur et dont il répond à la nouveauté. Le bébé semble doté, dès la naissance, de compétences pour le raisonnement plausible et l'apprentissage Bayésien, combinant de façon quasi optimale, les aprioris issus de notre évolution et les données reçues du monde extérieur. Ainsi la théorie Bayésienne résoudrait le dilemme classique entre les théories empiristes et nativistes. L'apprentissage du langage, la reconnaissance des mots, et la théorie de l'esprit pourraient également être modélisés comme des inférences Bayésiennes ».*

En ce qui concerne la boucle perception-action dans une approche bayésienne (Figure 20), Ernst et Bühlhoff (2004 : 164) affirment que notre perception du monde s'effectue dans une intentionnalité d'action et que nos actions affectent notre perception du monde qui nous entoure. Afin de permettre aux interactions d'avoir lieu, nous devons reconstruire notre environnement

à partir de l'ensemble des modalités sensorielles cartographiées par le cerveau.



**Figure 20:** La boucle perception-action dans un cadre Bayésien (Ernst et Bühlhoff, 2004 : 164). Dans un cadre bayésien, la perception émerge à travers l'organisation d'un ou plusieurs stimuli, émis par l'environnement, traités à travers des règles bayésiennes nécessitant un traitement sensoriel et une connaissance a priori. A partir de ce bit d'information, une décision sera prise provoquant une action ou non. Le choix pourra être motivé par une fonction gain/perte dont le poids dépendra du but à atteindre. L(a non-)action ainsi produite pourra avoir un impact sur l'environnement lui-même à travers ses interactions avec l'acteur.

Néanmoins, il est impossible d'enclencher ce processus par le seul traitement ascendant de l'information. De manière générale, les connaissances *a priori* (*prior knowledge*) sont les expériences passées reliées à un ou plusieurs média d'*input* sensoriel qui ont modifié le traitement neuronal de ces média à long terme (Bransford, *et al*, 2002 : 14–15). Dans le cadre Bayésien, elles sont l'ensemble des structures neuronales (pré-) établies qui permettent de traiter l'*input* sensoriel ; ces structures sont, d'une certaine manière, déjà en place chez le nouveau-né à travers l'organisation neurophysiologique de l'embryon et l'expérience intra-utérine. Les connaissances *a priori* (*prior knowledge*) sont indispensables pour donner du sens à cette reconstruction. Les inférences Bayésiennes nous permettent de modéliser un tel lien (Figure 20). Cette vision fait référence en psychocognition au rôle de la mémoire de travail (Baddeley et Hitch, 1974) comme :

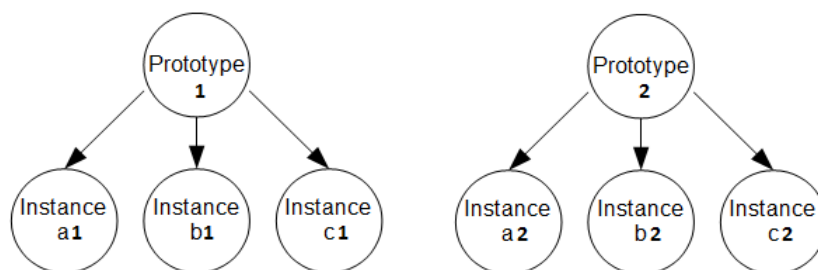
« à l'interface entre la cognition et la mémoire à long terme, et l'administrateur central permettrait à l'individu d'organiser son comportement non pas en fonction de stimulations immédiates mais en fonction de représentations internes de l'environnement » (Ollat, 1998 : 43).

Tiberghien (1999 : 255) rappelle que « *la mémoire ne concerne pas seulement le passé, elle détermine largement ce que sera notre présent perceptif. La mémoire génère en outre, de façon permanente, des schémas, des cadres d'interprétation qui façonnent nos anticipations.* ».

D'après l'ensemble de ces travaux, il semblerait qu'un grand nombre de nouvelles approches tendent en faveur de l'épistémè actuelle de l'incertitude (Morin, 2000 : 12) à travers la notion de non-linéarité dans l'évolution des organismes vivants. Elles permettent également de faire le pont entre les structures physiologiques, les pressions environnementales et la représentation mentale du monde. Elles permettent, en outre, de montrer l'impact des connaissances préalables sur la perception de nouvelles données. Néanmoins, nous noterons qu'en ce qui concerne la didactique et l'acquisition des langues, les différentes théories ou approches susmentionnées semblent trop fragmentées, dématérialisées pour proposer un modèle théoriquement pertinent et applicable.

### 2.2.2 Les théories émergentistes

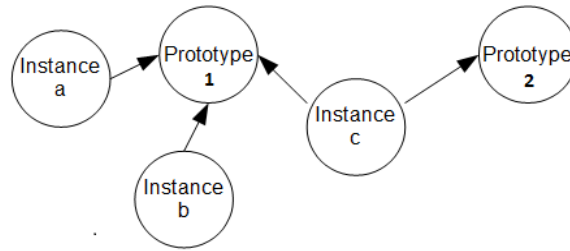
Contrairement au structuralisme qui décompose de manière descendante les systèmes (Figure 21), les théories émergentistes font intervenir des propriétés « *émergentistes* » qui vont au-delà de la simple émergence traditionnellement définie comme l'apparition d'un élément x.



**Figure 21:** Décomposition structuraliste des systèmes (par l'auteur).

Cette notion est apparue pour la première fois dans les travaux de John Stuart Mill, philosophe, logicien et économiste britannique qui propose que les systèmes complexes chimiques et physiques sont plus que la somme de leurs parties (O'Grady, 2010 : 1). De ce fait, l'organisation et l'évolution des organismes complexes reposent sur leurs structures (Figure 22). L'action principale est dès lors de recomposer le système plutôt que de le décomposer.





**Figure 22:** *Recomposition émergentiste des systèmes (par l'auteur).*

Nous repérons 6 grandes composantes émergentistes universelles que sont l'espace (densité, proximité), le temps (développement, accroissement), la force (type d'interaction), l'énergie (type d'énergie interactionnelle), la matière (atomes, molécules, ions) et la structure (organisation, morphologie). Les théories émergentistes reposent sur l'existence de propriétés émergentistes des parties qui émergent, de manière différente, dans le système complexe qu'elles composent. Pour comprendre cette influence, il est indispensable de repérer deux types d'émergentisme (Stephan, 1999):

- l'émergentisme faible (*Idem* : 1-2), proche du réductionnisme, fait l'état de propriétés émergentistes (certaines propriétés du système complexe ne se retrouvent pas dans les propriétés des parties) et des propriétés non-émergentistes (certaines propriétés du système se retrouvent dans celles des parties).
- l'émergentisme fort met en avant la notion d'irréductibilité (*Idem* : 2-5). On dissocie deux formes :
  - « *synchronic emergentism* » (émergentisme synchronique) : une propriété est émergentiste, si et seulement si, elle n'est pas réductible à la somme de propriétés des parties et la notion de nouveauté
  - « *diachronic emergentism* » (émergentisme diachronique) : il n'est pas possible de prédire le moment et le type de propriétés émergentistes qui apparaîtront d'un tel système.

Même si l'ensemble de ces recherches a trouvé un public important dans les sciences dites dures, à travers le connexionnisme par exemple, le rapport entre théories émergentistes et phénomènes langagiers ne s'est fait que plus récemment. L'idée principale repose sur le postulat que ces phénomènes sont mieux expliqués s'ils font référence à des facteurs simples non-grammaticaux et leurs interactions (O'Grady, 1997 : 1). Ces éléments sont, de manière non-exhaustive, les propriétés physiologique, perceptive, de la mémoire de travail, de la pragmatique, des

interactions sociales, des propriétés de l'*input* et celle des mécanismes d'apprentissage (O'Grady, 2010 : 2). Avant d'aller plus loin dans cette réflexion, il nous semble important d'ajouter que l'émergentisme ne récuse pas totalement la notion de nativisme en ce sens qu'il reconnaît l'existence de structures cérébrales innées mais qu'il n'existe pas de structuration cérébrale codant des principes ou contraintes grammaticales (Jalilzadeh, 2011 : 153). Nous pouvons, en ce sens, dénombrer deux approches émergentistes pour l'acquisition des langues :

- « *input-based emergentism* » (émergentisme fondé sur l'*input* linguistique) (O'Grady, 1997 : 2-4) : grandement basé sur le « *competition model* » (modèle fondé sur la compétition linguistique) (MacWhinney, 1998) figurant que la fréquence du contact à un *input* langagier aura un impact majeur sur les processus développementaux et acquisitionnels.
- « *processor-based emergentism* » (émergentisme fondé sur la notion de processeur linguistique) (*Idem* : 5-6) : théorisant que l'existence des composantes du langage (syntaxe, phonologie, etc...) est en réalité un ensemble de processus visant un traitement neurocognitif simple, économique et efficace.

Cette vision des éléments linguistiques comme processus visant à réduire le poids du traitement neurocognitif et de la mémorisation est au centre des théories émergentistes. Ainsi, la syntaxe est considérée comme émergeant de la conversation afin de faciliter le traitement sémantique et pragmatique ; la morphologie s'opère face aux régularités statistiques en contact ; le lexique est déterminé par la manière dont le cerveau stocke et gère les expériences mais aussi face à la fréquence de mise en contact à une occurrence (cette vision défie la notion de « *mot* » et de « *phrase* ») ; enfin la phonétique émerge en réponse aux différentes contraintes physiques des organes phonatoires et auditifs et la phonologie est un système d'actualisation des formes sonores en fonction de l'expérience (O'Grady, 2010 : 4-6).

Même si les théories émergentistes permettent de comprendre comment les propriétés de la « *langue* », qu'elles soient biologiques (physiologie, propriétés neuronales) ou cognitives, les « *compétences* » langagières sont vues comme des processus nés de l'émergence des parties du système en faveur d'un traitement efficace de l'*input*. Cependant, dans cette perspective, elle fait écho au connexionnisme (O'Grady, 2010 : 2) qui assume un système en réseau d'unités simples dans la modélisation des phénomènes mentaux. Il faut donc se questionner sur l'existence de systèmes complexes et de leur organisation. Néanmoins, nous ajouterons que, dans son article quelque peu provocateur « *Emergentism – Use often and with Care* », MacWhinney (2006 : 738-9) précise que l'éclairage émergentiste :

- « *n'est pas qu'une alternative à la Grammaire Universelle de Chomsky mais une méthode qui touche toutes les sciences étant donné que la nature est faite de processus émergents*
- *fournit des lignes générales de recherche afin de comprendre les mécanismes complexes mais qu'il est de la responsabilité de chaque chercheur d'appliquer ces lignes aux cas spécifiques*
- *permettra de mieux comprendre les phénomènes complexes à travers la prolifération d'études de corpora, de bases de données multimédia, des réseaux neuronaux, apprentissage à distance et les travaux en neurosciences de la cognition »*

L'émergentisme a eu un écho important en dans le domaine de l'acquisition d'une langue seconde ou étrangère, dans les années 1990 et dans la littérature anglophone, à travers le connexionnisme, les théories du chaos, de la complexité ou encore des systèmes dynamiques. En France, cet impact a été moindre, pour ne pas dire nul. L'un des défauts majeurs de l'émergentisme a été de traiter majoritairement des processus ascendants sans y intégrer de manière concrète ceux descendants. Intégrer le socioconstructivisme à l'émergentisme nous semble donc pertinent dans cette perspective.

## **2.3 Une approche intégrée**

Dans cette section, nous verrons en quoi le socioconstructivisme associé aux théories vues dans la section 2.2 peut apporter un regard complémentaire en tant que cadre théorique en didactique des langues et plus largement dans les sciences du langage.

### **2.3.1 Le socioconstructivisme**

Le constructivisme, en tant que théorie de l'apprentissage, est apparu, entre autres, dans les travaux de Piaget. Il s'inscrit en opposition au béhaviorisme ne présentant le processus d'apprentissage que comme une association stimulus-réponse (Skinner, 1904-1990 dans Rolland, 2011: 14-16). Brief (1977 : 195) rappelle que :

*« le constructivisme piagétien se développe le long de deux axes distincts. Le premier peut être inséré avec des ajustements mineurs dans la panoplie des diverses théories de la connaissance tandis que le second axe est proprement dominé par des soucis psychologiques et serait plutôt un objet d'étude pour l'épistémologie analytique. »*

L'idée principale repose sur le fait que les structures internes du sujet se construisent de manière progressive à travers les interactions avec l'objet. Selon Piaget (1967 : 56-57) : « *les idées sont antérieures à la langue, et la langue se construit au fur et à mesure des interactions* ». Nous ne nous positionnerons pas précisément sur ce point, au regard des différentes théories qui

coexistent actuellement, comme le synthétisent Marschark *et al.* (1997, chapitre 1) :

- « *language equals thought* » (le langage est concomitant à la pensée): position prise par les béhavioristes américains postulant que le langage et la pensée sont la même chose soit de la parole subvocale,
- « *language and thought are independent* » (le langage et la pensée sont indépendants): les développements du langage et de la cognition sont distincts (cf. Chomsky N., 1975 ; Fodor J., 1975),
- « *language determines thoughts* » (le langage détermine la pensée): liées au déterminisme linguistique et de la relativité linguistique (cf. Hypothèse de Sapir-Whorf dans Whorf, 1956 ; Sapir, 1985), les capacités langagières détermineraient les capacités cognitives d'un individu,
- « *thought determines language* » (la pensée détermine le langage): le développement cognitif chez l'enfant se mettrait en place en premier puis soutiendrait la mise en place des structures langagières (cf. Piaget, 1964).

Le socioconstructivisme apparaît avec les notions de « *constructivisme social* » et de « *zone proximale de développement* » pour Vygotski (1934/1997). Cette dernière ne s'inscrit pas dans le constructivisme piagétien par le fait que :

*« les recherches psychologiques liées au problème de l'apprentissage scolaire se bornaient ordinairement à établir le niveau de développement mental de l'enfant. Mais ce niveau est insuffisant pour déterminer l'état de son développement. »*  
(Vygotski, 1934/1997 : 350-1)

Il propose alors, afin de déterminer l'état du développement, de « *prendre en considération non seulement les fonctions venues à maturité mais aussi celles qui sont au stade de la maturation, non seulement le niveau présent mais aussi la zone prochaine de développement* » (Vygotski, 1934/1997 : 351). Il définit concrètement la zone prochaine de développement comme :

*« [la] disparité entre l'âge mental, ou niveau présent de développement, qui est déterminé à l'aide des problèmes résolus de manière autonome, et le niveau qu'atteint l'enfant lorsqu'il résout des problèmes non plus tout seul mais en collaboration »* (Vygotski, 1934/997 : 351).

En rupture avec une approche behavioriste de l'apprentissage, le constructivisme puis le socioconstructivisme ont permis de réintégrer dans le processus d'apprentissage/développement l'importance et l'influence de l'environnement à travers les interactions que l'individu opère tout au long de sa vie. C'est à ce propos que Vygotski dissocie deux processus psychiques :

*« toute fonction apparaît deux fois dans le comportement social de l'enfant : d'abord au niveau social, ensuite au niveau individuel ; d'abord entre les personnes (interpsychologique), ensuite à l'intérieur de l'enfant (intrapsychologique). [...] Toutes les fonctions supérieures ont leurs origines dans les relations réelles entre individus humains. »* (Vygotski, 1931/1974 : 201 dans Schneuwly, 2008 : 20-1).

C'est à travers une approche socioconstructiviste, que l'apprentissage des langues s'est développé autour des interactions avec autrui. Cependant, une approche purement sociale des processus mentaux ne permet pas de prendre en compte les différentes limites et contraintes que la biologie (génétique, physiologique, etc.) impose. Nous proposons, de ce fait, une approche conjointe entre émergentisme et théories socioconstructivistes pour la prise en compte des phénomènes ascendants et descendants dans le processus de développement langagier.

### **2.3.2 Socioconstructivisme, émergentisme et didactique des langues**

Intégrer les théories émergentistes, de la complexité et du chaos dans l'analyse de la « *langue* » nécessite une modélisation nouvelle permettant à chaque chercheur de l'intégrer dans son approche du monde langagier verbal. Comme le rappelle Narcy-Combes (2002), la modélisation en « *langue* », telles que les sciences humaines la conçoivent, permet d'accroître la légitimité des travaux en favorisant un positionnement cognitif et épistémologique. Toutefois, les trois théories dont nous traitons dans cette réflexion ne sont qu'à leurs débuts en acquisition des langues et encore plus en didactique en France. De plus, la modélisation semble indispensable car ces théories sont complémentaires et font émerger une « *méta* » approche qui va bien au-delà de l'émergentisme « *pur* » en interrogeant de manière plus certaine les relations « *top-down* » (descendante) et « *bottom-up* » (ascendante). En cela, la « *simplicité* » définie par Berthoz comme « *une nécessité biologique [de traiter l'information] très rapidement, avec élégance et efficacité, de situations complexes, en tenant compte de l'expérience passée et en anticipant l'avenir* » (Berthoz, 2009 : 17) impose de nouvelles règles de modélisation. Dans cette visée, un nouveau rôle de la modélisation émerge : « *[elle] permet de fournir, aux acteurs du dispositif, un cadre auquel se référer et s'appuyer afin de s'adapter aux réorganisations chaotiques internes et externes (environnement) dans le but de vérifier la validité des concepts qui justifient les interactions en action* »<sup>48</sup> (Narcy-Combes & Bertin, 2012a : section 3.1).

La complexité ne peut être simplifiée mais il est possible de ne pas la rendre compliquée. La schématisation en trois dimensions apparaît donc comme un moyen de rendre compte de cette complexité sans la simplifier tout en affichant sa valeur « *simplexe* ». La schématisation traditionnelle « *simplifiante* » n'est pas la solution à ce problème. Nous proposons ici de nous

---

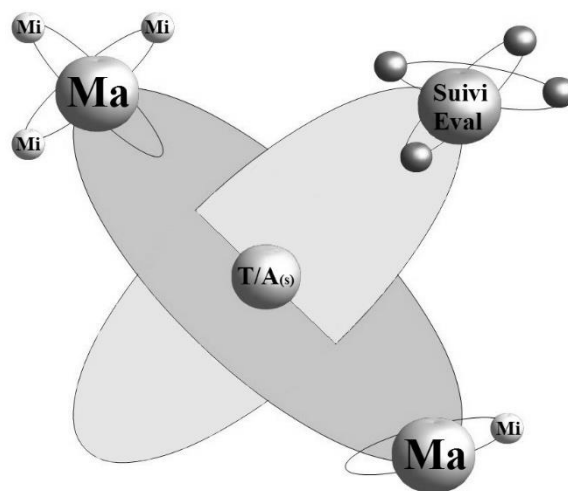
48 « *Modelling provides the actors with a framework they can refer to, guidelines to check the validity of the concepts that justify the interactions at work.* »

inspirer du monde atomique et/ou astronomique comme métaphore de la « langue » pour une modélisation du dispositif par tâche d'apprentissage des langues réparti en tâches d'entraînement (micro tâches) et tâches sociales (macro tâches) (Narcy-Combes, 2006) (Tableau 1).

	<b>Tâche sociale</b>	<b>Tâche d'entraînement</b>
Objectifs	- sensibilisation aux obstacles, aux écarts inter et intralinguaux - création de besoins d'apprentissage	- (re)mise en place d'une connaissance explicite adéquate
Caractéristiques	- produit socialement réaliste - la validation est faite par des membres externes au dispositif pour assurer une valeur sociale	- tâches d'interprétation, de conscientisation et d'entraînement - regroupées (ou non) en centre de ressources numériques ou physiques

**Tableau 1:** Caractéristiques principales des tâches sociales et d'entraînement. (D'après Bertin et al., 2010, chapitre 11 dans Miras & Narcy-Combes, 2014 : 22).

Dans un tel dispositif, la relation tuteur/apprenant(s) est au centre d'un système satellitaire composé de sous-systèmes : tâches sociales (macro tâches) et de suivi/évaluation eux-mêmes centres de gravité respectivement de micro tâches et micro-suivi/évaluation (Figure 23).



**Figure 23:** Dispositif par tâche (d'après Narcy-Combes & Miras, 2013 : 40). La relation tuteur-apprenant(s) (T/A(s)) est au centre du système d'apprentissage où gravitent des tâches sociales (ou macro-tâche : Ma) et un suivi/évaluation (Suivi Eval). Chaque tâche sociale peut motiver le déclenchement d'une ou plusieurs tâches d'entraînement (ou micro-tâche : Mi). Le suivi/évaluation pourra également intégrer des micro-étapes non modélisées ici.

Ces différents ensembles interagissent selon les principes suivants :

- chaque élément est pondéré en fonction de leur importance,
- cette pondération influe sur la nature et force d'interaction des éléments entre eux ainsi que la nature des éléments eux-mêmes,
- tous les éléments sont en interaction.

Ce système s'intègre dans une vision non-linéaire et dynamique, il est donc indispensable de le considérer en fonction du temps (Figure 24). Miras & Narcy-Combes (2014 : 22) rappellent que :

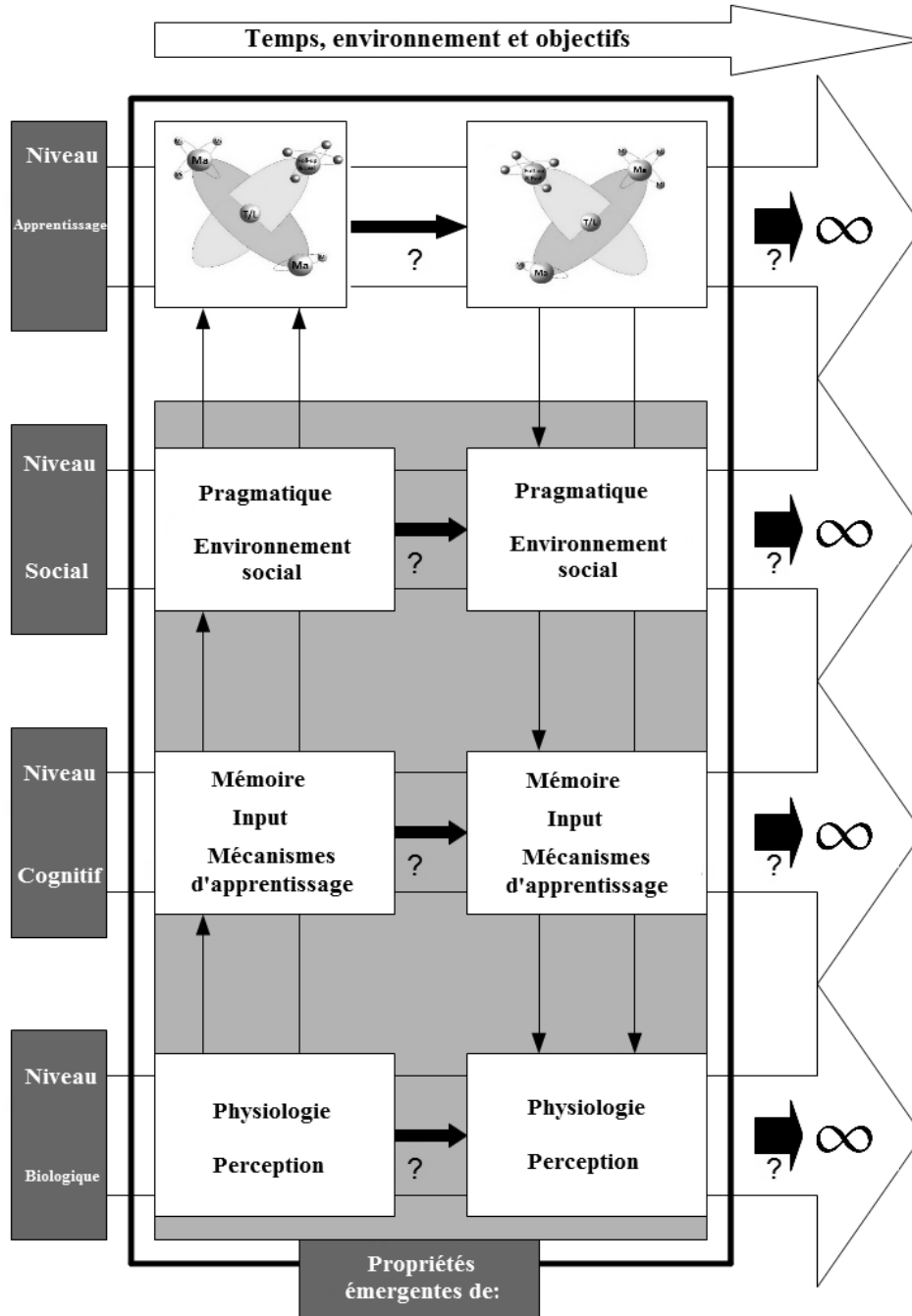
*« une tâche sociale (macro tâche) est mise en place : si (1) elle est validée, alors une autre tâche sociale est lancée, si (2) elle fait émerger des difficultés, un certain nombre de tâches d'entraînement (micro tâches) sur des points spécifiques sera mis en place jusqu'à ce que la tâche sociale puisse être validée. »*

A ce titre, nous tenons à rappeler le caractère particulier de la prononciation qui contrairement à la syntaxe, à la morphologie et au lexique, reste majoritairement intuitive, même en L1. Le travail de médiation de la prononciation, selon nous, portera principalement sur des tâches d'entraînement. Nous reviendrons dans la troisième partie de ce travail de thèse sur cette particularité.

Dans cette représentation globale (Figure 24), le dispositif est stimulé par les propriétés émergentes réparties en différents niveaux (biologique, cognitif et social) de manière directe ou indirecte. De cette stimulation émergentiste mais également de l'évolution interne au dispositif, le système va brusquement changer vers un autre état d'organisation (qualitatif et quantitatif) qui va pouvoir modifier (ou non) les propriétés initiales. Certaines de ces propriétés pourront être modifiées de manière instantanée (propriétés interactionnelles) ou sur du plus long terme (propriétés biologiques). Cette évolution s'effectue à l'infini et de manière non-prévisible, respectivement représentée par les signes ( $\infty$ ) et (?). L'ensemble de ces changements est stimulé et cadré par le temps, l'environnement et les objectifs (institutionnels, personnels, etc..).

Cette modélisation a pour intérêt de mettre en évidence la mise en place de processus langagiers dans un apprentissage par tâche des langues. Le chaos à l'œuvre dans cette complexité émergentiste s'invite dans un cadre mêlant environnement et objectifs dans le temps. La relation tuteur/apprenant mise au centre du dispositif renforce le rôle de médiateur (Demaizière, 2007 : 5) dans l'individualisation des apprentissages. Il peut être déstabilisant pour certains enseignants de concevoir un apprentissage « chaotique » dans un cadre institutionnel où il faut suivre un


programme. Toutefois, si nous pouvons admettre les biais de ce cadre, il est possible de médier l'apprentissage en favorisant l'émergence de nouvelles performances.



**Figure 24:** Dispositif global sous le prisme de l'émergentisme (par l'auteur). Le dispositif évolue dans le temps et à travers les pressions environnementales et les objectifs. Ces changements sont stimulés et stimulent des niveaux inférieurs qu'ils soient biologiques, cognitifs, sociaux et/ou liés aux mécanismes d'apprentissage. Les changements sont imprévisibles car s'effectuent de manière non-linéaire.



De manière synthétique (Tableau 2), concevoir une approche conjointe des théories émergentiste et socioconstructiviste permet de prendre en compte dans le développement des facultés de langage (musical ou verbal) l'ensemble des contraintes génétiques et physiologiques du corps humain, tout en considérant que ce cadre est modulé par les interactions avec le monde qui nous entoure.

Emergentisme		Socioconstructivisme
<ul style="list-style-type: none"> <li>- Complémentarités imprévisibles des tâches</li> <li>- Liens processus/tâches</li> <li>- Nécessités d'interactions</li> <li>- Valeur et limites de l'entraînement</li> <li>- Performances et compétences</li> <li>- Ce qui est universel en neurophysiologie</li> </ul>		<ul style="list-style-type: none"> <li>- Déclenchement social des apprentissages</li> <li>- Motivation sociale des apprentissages</li> <li>- Nécessité d'interactions, mais peu d'attention à l'entraînement</li> <li>- Contextualisation de l'apprentissage</li> <li>- Authenticité du besoin de changer de code</li> <li>- Ce qui est idiographique en psychologie</li> </ul>

**Tableau 2:** Tableau synthétique des apports complémentaires de l'émergentisme et des théories socioconstructivistes (Miras & Narcy-Combes, 2014 : 24).

Cette approche combinée des théories émergentistes et socioconstructiviste permet donc de remettre en cause la catégorisation du monde qui nous entoure telle qu'elle a été présentée jusqu'à présent sous le prisme du structuralisme. Elle permet notamment de donner un cadre théorique dans lequel les pressions ascendantes et descendantes peuvent (co)exister avec la possibilité de mesurer les changements qu'ils engendrent dans un système mais sans qu'il soit possible de prédire l'émergence de ces ruptures. Dans cette vision, musique et parole peuvent faire l'objet d'un traitement comparé dans une perspective pluri-inter-transdisciplinaire et plus particulièrement en didactique des langues. La modélisation que nous proposons a pour but de montrer aux acteurs de l'enseignement qu'il est possible de prendre en compte le chaos dans un cadre (même institutionnel) de formation afin de prendre en compte les profils d'apprentissage de chaque apprenant. Il nous reste, par conséquent, à intégrer ce cadre dans l'optique de l'enseignement/apprentissage de la prononciation du français en contexte FLE/S.

### 3. Didactique de la prononciation en contexte FLE/S

#### Résumé

Le chapitre « *Didactique de la prononciation en contexte FLE/S* » porte sur les mécanismes qui sous-tendent l'acquisition d'une langue étrangère et plus particulièrement de sa prononciation. Nous focaliserons notre attention sur la didactique de la prononciation en contexte FLE/S et notamment de l'utilisation de la musique instrumentale dans ce domaine.

La section 3.1 dresse un panorama de la didactique de la prononciation en contexte FLE/S afin de mettre en évidence que ce domaine a favorisé très rapidement une approche de type behavioriste basée sur l'articulatoire (Wachs, 2011; Lauret, 2007) à l'exception de la méthode verbo-tonale (P. Guberina), pour ne citer que la plus (re)connue. Le paradigme actuel dans lequel se situent les approches communicative puis actionnelle n'a pas permis de faire avancer les considérations de la prononciation dans l'apprentissage/enseignement général des langues étrangères. Afin de contrebalancer ce constat, nous ferons le bilan des recherches en acquisition phonético-phonologique à travers les travaux de Kuhl (2004) ; Flege (1995a) ; Best *et al.* (1988) ; Honikman (1964) ; entre autres et ce afin de réfléchir à l'apparente opposition entre perception et production. Nous terminerons cette section en montrant que les caractéristiques cardinales et focales de certaines voyelles du français standard représentent des cibles perceptives pertinentes pouvant servir de point d'appui (Barth, 2004) à l'apprentissage.

La section 3.2 prendra le contre-pied de la section précédente en montrant que production et perception sont intimement liées dans la prononciation en « langue » et qu'il n'est pas possible par conséquent de travailler les deux processus indépendamment mais bien dans une approche intégrée. Pour ce faire, nous présenterons en détails ce qu'ont apporté à la question la compréhension de l'effet McGurk (McGurk & MacDonald, 1976), la théorie motrice de la perception (Liberman, 1967, 1985) et des neurones miroirs (Rizzolatti & Sinigaglia, 2006). Cette relation nous permettra également d'expliquer en quoi l'utilisation de la musique instrumentale dans l'amélioration des capacités acoustiques perceptives pourrait prendre part au travail de médiation de la perception et production en français langue étrangère.

La section 3.3 développera l'ensemble des pratiques pédagogiques en dehors des approches

traditionnelles qui prennent en compte d'autres paramètres dans l'apprentissage que celui purement réflexif. Nous commencerons par présenter plusieurs approches/techniques, qui mettent le corps au centre de l'apprentissage, telles que la suggestopédie (cf. G. Lozanov), le Silent Way (cf. C. Gattegno) ou le bonhomme sonore (Guimbretière & Kaneman-Pougatch, 1991 : 70). Nous poursuivrons par montrer qu'actuellement la musique est déjà présente dans l'enseignement/apprentissage des langues comme avec les Jazz Chants (Graham, 2000), les Ritmimots (Llorca, 1988), le « *European Music Portfolio : a creative way into languages* » (Ludke, *et al.*, 2012) ou d'autres que nous contesterons comme *SpeedLingua*©. Nous terminerons cette section en montrant que la didactique de la musique peut également apporter un regard différent sur l'enseignement/apprentissage de la prononciation en langue étrangère en questionnant la relation individu-son (Terrien, 2012 ; Malinen, 2000 ; Huhtanen, 2008 ; entre autres).

### **3.1 Articulatoire *versus* perceptif ?**

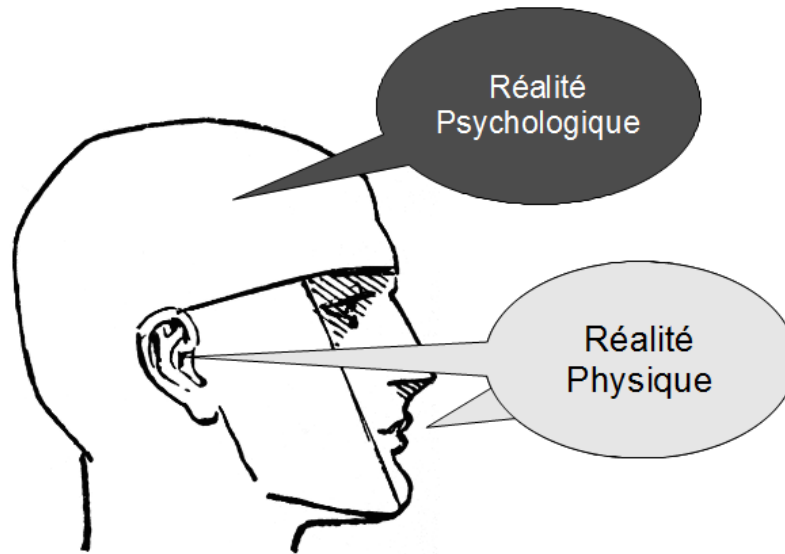
Dans cette section, nous dresserons un panorama de la didactique de la prononciation afin de montrer dans quelle épistémè elle se situe. A partir de ce constat, nous mettrons en évidence les recherches qui portent sur l'acquisition phonético-phonologique en langues étrangères. Nous terminerons en montrant que certaines caractéristiques des voyelles du français standard pourraient avoir un intérêt particulier dans l'enseignement/apprentissage de la prononciation du français en contexte FLE/S.

#### **3.1.1 Panorama de la didactique de la prononciation**

Ce travail de synthèse permettra de porter un regard épistémologique sur les différentes approches générales de la didactique de la prononciation dans la perspective qui est la nôtre. Dès le XIX<sup>ème</sup> siècle, il est fait référence à « *l'art de parler* ». Cette dénomination fait apparaître dans les classes sociales qui se doivent de le pratiquer, que parler (le français) n'est pas seulement un fait commun, mais requiert un enseignement au même titre que la musique, etc. De cette nécessité de maîtriser l'art de parler, les maîtres doivent adapter leur enseignement et l'on fait intervenir la notion de « *double action artistique (étudier et enseigner)* » (Aubin, 2008 : 103). Dans cette perspective, le maître de langue maîtrise sa parole, au même titre que le maître de chant. Cette conception du parler fait intervenir l'idée d'art mais dans un sens large où l'importance du lexique, de la prononciation mais aussi de la rhétorique en sont les éléments principaux. D'ailleurs, à cette époque, le français n'a pas une valeur incontestée de « *langue* »

musicale qui fait face à l'importance de l'italien et de l'allemand, de par leurs influences en termes culturels. Les analogies entre musique et parole, notamment à travers des expressions imagées ou des exercices empruntés à la musique (dictée, répétition, écoute, etc...), sont un phénomène qui a pris une ampleur particulière aux XVII<sup>ème</sup> et XVIII<sup>ème</sup> siècles, mais qui s'est vue décroissante dès le milieu du XIX<sup>ème</sup> (Aubin, 2008 : 105). Il faut, néanmoins, faire à cette époque et pour ces classes aisées, la distinction entre l'enseignement consacré aux filles et celui aux garçons. Les sciences classiques sont enseignées aux garçons ; l'enseignement des langues prend alors en compte une vision lexicale (politique, économie, diplomatie), grammaticale et syntaxique. En opposition à cela, les filles se doivent de connaître les arts d'agrément. L'enseignement des langues est majoritairement oral et a pour but de favoriser l'utilisation contrôlée d'une « *langue* » mélodieuse. L'apprentissage des filles étant vu comme inférieur à celui des garçons, les maîtres de langue étaient plus libres dans leurs méthodes. Il n'est donc pas étonnant de comprendre que « *c'est dans le système scolaire masculin que la mort de l'approche artistique et des méthodes musicales, qui n'étaient qu'au stade embryonnaire, fut la plus spectaculaire* » (Aubin, 2008 : 108). En effet, les sciences classiques, proposées aux garçons, font intervenir des calques didactiques de l'enseignement des langues mortes (latin) aux langues vivantes.

L'influence du structuralisme à travers Ferdinand de Saussure (1906-1911) et le Cercle linguistique de Prague (1929-1939 : R. Jakobson et N. S. Troubetzkoy) (Vaissière, 2006 : 19) a eu pour effet de faire évoluer cette approche où l'appareil phonatoire est vu comme un instrument de musique, par les pionniers de la phonétique, vers une dissection systématique, avec l'arrivée des moyens technologiques, d'une machine à sons.



**Figure 25:** La double réalité des unités sonores verbales (par l'auteur).

Le phonème, unité minimale linguistique de l'oral (Martinet, 1970 : 16), peut donc accepter une double réalité (Figure 25) : physique telle qu'elle a été définie initialement par D. Jones (1931 : 74) mais également psychologique, mise en emphase par Baudouin de Courtenay cité par Twaddell (1935 : 56), Hyman (1975 : 72) et repris par Randall (2007 : 39) :

*« c'est un concept abstrait, une construction mentale, un symbole qui existe dans l'esprit des locuteurs. Le défi de la psychologie et de la linguistique est d'apporter des explications sur le transfert de données brutes physiques vers des représentations abstraites symboliques de la parole ».*

Cette dernière vision est renforcée lorsque Krashen (1988 : 35), d'obédience chomskyenne, affirme que « la prononciation est beaucoup plus ancrée dans la personnalité de l'apprenant que n'importe quel autre aspect de la langue ». Par ailleurs, la réalité même du phonème peut être remise en doute, en tant que son du discours représenté dans l'esprit et que « l'on s'efforce d'atteindre » (Goody, 1994 : 276) au profit du phone : « le son concret employé dans toute émission du discours » (*Ibidem* : 277). Nous pouvons le dissocier en deux niveaux : physiophonique « représentant les sons qui sont émis en fait » et psychophonique « représentant que des phonèmes » (*Ibidem*).

Cette double réalité a trouvé un écho différent selon les époques et les différentes méthodes. Nous reviendrons brièvement sur les différentes méthodes appliquées depuis le XVI<sup>ème</sup> siècle :

– **la méthode traditionnelle** (Wachs, 2011 : 186) :

Principe : il s'agit d'une méthode grammaire/traduction qui vise la traduction de textes sacrés et la maîtrise de l'écrit littéraire.

Prononciation : il n'y a aucun intérêt porté à cette composante de la « *langue* ».

Critique : il n'y a pas de prise en compte des processus acquisitionnels.

– **la méthode directe** (Wachs, 2011 : 186-7 ; Lauret, 2007 : 83-8) :

Principe : l'apprentissage est considéré comme intuitif par imitation, à l'image des jeunes enfants, dans l'apprentissage de la L1. Elle est basée sur les premières recherches acquisitionnistes.

Prononciation : il est préconisé de laisser le temps aux apprenants dans leur écoute des nouveaux sons afin de favoriser l'imprégnation sonore.

Critique : l'apprentissage d'une langue étrangère n'est pas le même que l'apprentissage d'une L1 (langue première) chez l'enfant.

– **la méthode audio-orale et la méthode Structuro-Globale Audio Visuelle (SGAV)**

(Wachs, 2011 : 187-8 ; Guimbretière, 1994 : 46-7 ; Lauret, 2007 : 95-107) :

Principe : l'intérêt est porté sur la notion d'habitude en lien avec les travaux sur le béhaviorisme (Skinner, 1904-1990 dans Rolland, 2011 : 16-7). L'accent est tantôt porté sur :

- la méthode articulatoire : « *l'émission des sons implique une connaissance relativement poussée du fonctionnement de l'appareil phonatoire* » (Guimbretière, 1994 : 46) ;
- l'audition de modèle : « *l'intégration des notions est facilitée par un apprentissage inconscient [à travers des drills]* » (Guimbretière, 1994 : 47) ;
- la méthode des oppositions phonologiques : « *mémoriser les phonèmes par oppositions de type binaire en les faisant répéter sous forme de paires minimales* » (Guimbretière, 1994:47).

Prononciation : ces trois méthodes coexistent dans un continuum pédagogique dans l'approche audio-orale. Deux courants apparaissent : « *les tenants d'une approche consciente, raisonnée, en un mot intellectualisée, qui s'opposent aux tenants d'une pédagogie fondée sur l'apprentissage mécaniste qui s'organiserait au niveau de l'inconscient, et ou serait bannie toute fonction raisonnante* » (Guimbretière, 1994 : 47).

Critique : le système est proposé sous une forme réduite de discours à travers des unités isolées qui, par conséquent, réduit les possibilités de combinaisons des phonèmes entre eux. Les dialogues proposés dans des documents audio ou vidéo manquent de naturel.

- **la méthode verbo-tonale** (Wachs, 2011 : 186-7 ; Lauret, 2007 : 83-8) :

Principe : dans les années 50, Petar Guberina affirme que :

*« la stratégie verbo-tonale se caractérise par une rééducation de l'audition passant par un quadruple conditionnement au plan psychologique, corporel, psychosomatique et audio-phonatoire. [...] Les procédés correctifs sont disposés de manière à reconditionner l'audition par une action portant non pas sur l'articulation du sujet mais sur le modèle afin de procéder à une intégration inconsciente »* (Guimbretière, 1996 : 48).

Le but final n'est pas de faire acquérir de nouvelles unités mais de faire en sorte que les unités soient utilisées, de manière différente, par rapport à leurs usages dans la langue maternelle/première.

Prononciation : la méthode verbo-tonale intègre l'individu dans son apprentissage. Six principes généraux sont retenus : la motivation maximale, le recours au non-verbal, pas d'intellectualisation, le respect de la structure, la faute comme point de départ, la patience dans la maturation (Renard, 2002 : 233).

Critique : le refus d'intellectualisation, notamment chez l'adulte, peut empêcher la distanciation avec la production sonore en L1, par exemple. Il n'y a pas de transition entre les contextes facilitant proposés et la réalisation d'énoncés en parole spontanée.

- **Les approches communicative et actionnelle** (Wachs, 2011 : 189-90 ; Lauret, 2007 : 153) :

L'approche communicative, à la fin des années 70, révèle un changement de paradigme dans la didactique des langues. Comme le rappelle Rolland (2011 : 54) : *« les langues ne sont plus considérées comme des objets d'étude examinés indépendamment de l'usage qui en est fait. Le langage est considéré du point de vue de son utilisation [de ce fait] la phonologie est peu exploitée au profit d'activités de communication permettant d'utiliser la langue et non de l'étudier. »* Quand un apprentissage plus spécifique sur la prononciation est prévu, il s'intègre souvent dans une tradition béhavioriste. Les simulations, au centre de cette approche, favorisent l'émergence de processus d'évitement (évoqué dans Narcy-Combes, 2005 : chapitre 5 et mentionné dans Narcy-Combes, 2012b : 82). Dans le prolongement, l'approche actionnelle du Conseil de l'Europe (2001) propose un Cadre Européen Commun de Référence en Langue (CECRL) dont Macaire *et al.* (2010 : 15) rappellent qu'il a acquis une *« autorité incontestable »*. Dans ce dernier, la *« compétence phonologique »* (nous reviendrons plus tard sur la notion de *« compétence »*) est définie par (1) une connaissance de la perception et de la production et (2) une aptitude à percevoir et produire certaines unités et caractéristiques prosodiques (Tableau 3).

## 5.2.1.4 Compétence phonologique

- Elle suppose une connaissance de la perception et de la production et une aptitude à percevoir et à produire
- les unités sonores de la langue (phonèmes) et leur réalisation dans des contextes particuliers (allophones)
  - les traits phonétiques qui distinguent les phonèmes (traits distinctifs tels que, par exemple sonorité, nasalité, occlusion, labialité)
  - la composition phonétique des mots (structure syllabique, séquence des phonèmes, accentuation des mots, tons, assimilation, allongements)
  - la prosodie ou phonétique de la phrase :
    - accentuation et rythme de la phrase
    - intonation
    - réduction phonétique
    - réduction vocalique
    - formes faibles et fortes
    - assimilation
    - élision.

MAÎTRISE DU SYSTÈME PHONOLOGIQUE	
<b>C2</b>	Comme C1
<b>C1</b>	Peut varier l'intonation et placer l'accent phrastique correctement afin d'exprimer de fines nuances de sens.
<b>B2</b>	A acquis une prononciation et une intonation claires et naturelles.
<b>B1</b>	La prononciation est clairement intelligible même si un accent étranger est quelquefois perceptible et si des erreurs de prononciation proviennent occasionnellement.
<b>A2</b>	La prononciation est en général suffisamment claire pour être comprise malgré un net accent étranger mais l'interlocuteur devra parfois faire répéter.
<b>A1</b>	La prononciation d'un répertoire très limité d'expressions et de mots mémorisés est compréhensible avec quelque effort pour un locuteur natif habitué aux locuteurs du groupe linguistique de l'apprenant/utilisateur.

Les utilisateurs du Cadre de référence envisageront et expliciteront selon le cas

- les aptitudes phonologiques nouvelles exigées de l'apprenant
- l'importance relative des sons et de la prosodie
- si l'exactitude phonétique et l'aisance constituent un objectif d'apprentissage immédiat ou à plus long terme.

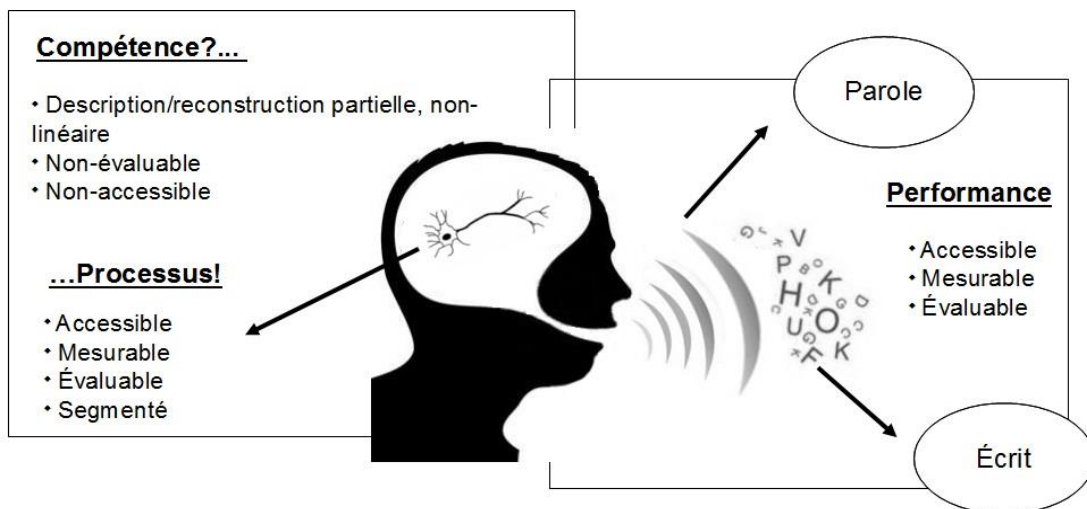
Tableau 3: La compétence phonologique par le CECRL (Conseil de l'Europe, 2001 : 91-2).

Il intègre donc d'une part, une connaissance explicite et d'autre part, son intégration implicite. Nous ne reviendrons pas en détails sur les problèmes épistémiques posés, dans la typologie en niveau (A1 → C2), de formulations vagues et de l'utilisation de notions du type « *répertoire très limité* », « *erreurs de prononciation* ». La « *compétence* » de l'apprenant étant évaluée sur le simple fait que le « *locuteur natif* » puisse comprendre « *avec quelque effort* » tout en étant « *habitué aux locuteurs du groupe linguistique de l'apprenant/utilisateur* » ou qu'il doive « *faire répéter* ». La maîtrise totale du système est caractérisée, selon le référentiel, par la maîtrise de la prosodie. Le cadre de Référence délègue la question de « *l'importance relative des sons et de la prosodie* » à ses utilisateurs. Il est clair que ni l'approche communicative, ni l'approche actionnelle à travers le CECRL ne prennent position sur la place, dans une perspective interactionniste, des unités phonétiques et phonologiques dans l'apprentissage d'une langue étrangère.



Dans la perspective émergentiste et socioconstructiviste qui est la nôtre, nous nous positionnerons dans une approche par tâches (Narcy-Combes, 2006 : 77-87) (voir partie 1, chapitre 2). Elle permet notamment de préciser la trichotomie « *compétence / performance / processus* » (Figure 26) :

« la « *compétence* » est un construit social ancré dans les représentations des individus par reconstruction a posteriori. Elle n'est, par conséquent, ni directement accessible ni évaluable. A contrario, les processus neurocognitifs sont eux mesurables à travers des techniques médicales (IRM(f), EEG, etc.) mais restent segmentés à cause des limites expérimentales que ces outils imposent. [...] Nous n'avons finalement accès qu'à des performances orale (parole/discours) et/ou écrite (textes) qui sont mesurables et évaluables. » (Miras & Narcy-Combes, 2014 : 19)



**Figure 26:** Représentation des rapports entre les notions de processus, compétence et performance (Miras & Narcy-Combes, 2014 : 19).

Nous rappellerons que l'approche par tâche nous semble répondre aux résultats actuels en didactique/acquisition des langues tout en permettant de prendre en compte la non-linéarité des apprentissages. C'est également une approche non-symboliste puisque reposant sur les performances en tant que processus neurophysiologiques. Néanmoins, le caractère particulier de la prononciation (intuitif même en L1) requiert un travail ancré dans des tâches d'entraînement. Il est donc important de commencer à repenser quel type de tâches d'entraînement proposer dans **la médiation de la perception et la production des sons en langue(s) étrangère(s)**. Ce terme sera préféré, compte tenu de notre approche, aux dénominations de « *phonétique corrective* », « *cours de prononciation* », « *cours de phonétique* », etc. Le but de notre perspective n'étant pas de « *corriger* », « *remédier à* » ou « *donner un cours de* » mais d'accompagner l'apprenant, s'il le souhaite, vers une prononciation dans laquelle il se sentira plus à l'aise en production mais également de favoriser la

perception/réception de lui-même et par autrui. Il nous reste donc à comprendre quels sont les processus qui sous-tendent l'acquisition d'un nouveau système phonético-phonologique afin d'atteindre cet objectif.

### 3.1.2 Acquisition phonético-phonologique

Le travail de médiation de la perception/production des sons aura donc pour but de tendre vers une limitation des variations entre ce que l'individu perçoit de sa production et ce qu'il produit réellement, voire avec ce que perçoit autrui et ce qu'il perçoit d'autrui. Nous nous focaliserons sur les problèmes liés à l'acquisition d'un nouveau système phonologique (la parole) même si nous mesurons l'importance de la prise en compte des phénomènes liés à la voix tels que la différence entre la voix aérienne et solide : respectivement « *le feed-back auditif [qui] s'effectue à la fois par conduction aérienne et par conduction osseuse* » et « *l'onde sonore [transmise] à l'oreille par conduction osseuse* » (Scotto Di Carlo, 1990 : 246). Le travail portera sur ce que nous entendons à travers notre filtre physiologique et cognitif et ce que nous produisons réellement. Cette distorsion n'est pas un phénomène nouveau puisque Troubetzkoy, d'obédience pragoise, nous propose dès 1939 :

*« le système phonologique d'une langue est semblable à un crible (phonological sieve) à travers lequel passe tout ce qui est dit. Seules restent dans le crible les marques phoniques pertinentes pour individualiser les phonèmes »* (Troubetzkoy, 1939 : 54 dans Rolland, 2011 : 34).

Notre réflexion, socioconstructiviste et émergentiste, sous-tend l'observation de processus physiologiques dont les ruptures sont déclenchées par des interactions sociales. Dans cette perspective, Renard (2002 : 218) fait référence au crible comme :

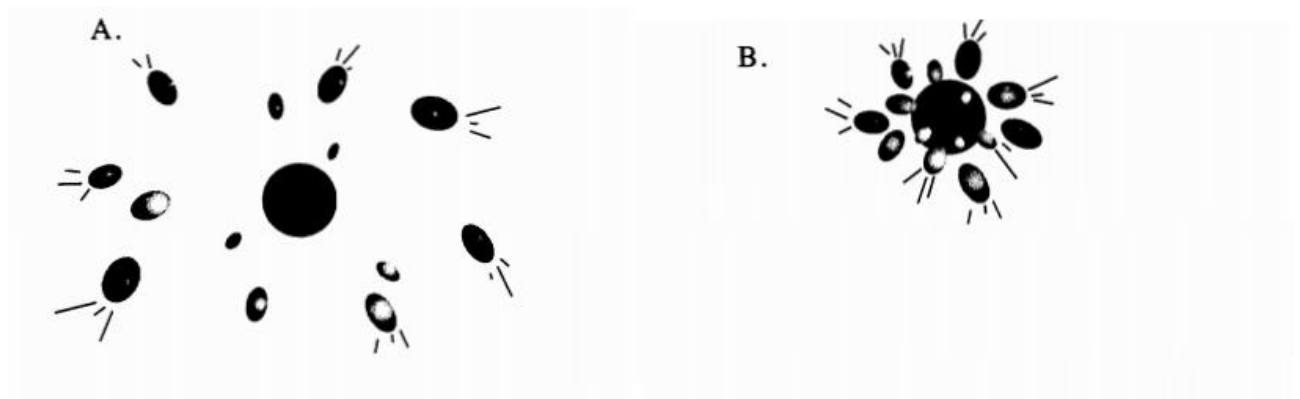
*« une structuration conditionnée du cerveau qui « filtre » la réalité linguistique en fonction de la référence au système phonologique maternel [alors que] l'articulation erronée n' [est] que la manifestation de cette imprégnation auditive préalable ».*

Nous préférons, par conséquent, la notion de nativisation phonologique, en référence aux travaux d'Andersen (1983), mettant en évidence la notion de processus plutôt que celui de crible dont les représentations s'inscrivent dans une perspective symboliste. Les termes de nativisation/ dénativisation sont préférés à ceux d'assimilation/accommodation de Piaget (1970) en raison de l'ambiguïté avec l'assimilation phonologique : « *deux consonnes en contact ne partageant pas le même trait s'assimilent, le trait devenant commun, en fonction de la consonne en position forte.* » (Lauret, 2007 : 74) ; par exemple : dans « *médecin* » le /d/ aura tendance à se dévoiser au profit d'un /t/ au contact de la consonne sourde /s/. Malgré l'obédience

environnementaliste pour la nativisation et constructiviste pour l'assimilation, Andersen (1983 : 10-11) rappelle que la nativisation est proche de l'assimilation.

Dans une perspective neurocognitive, Kuhl (2004) traite du phénomène de « *crible phonologique* » à travers une organisation de la perception autour d'un « *magnet effect* » (effet d'aimant perceptif) (Kuhl & Iverson, 1995 : 121-154) (Figure 27). Il se définit par :

« *le prototype a un effet d'attracteur sur les sons qui l'entourent. Il attire perceptivement les autres membres de la catégorie autour de lui [...] l'effet d'aimant perceptif implique la réduction de la distance perceptive entre le prototype et les sons périphériques.* »<sup>49</sup> (Kuhl & Iverson, 1995 : 123)



**Figure 27:** Modélisation de l'effet d'aimant perceptif. Les sons perceptivement proches (A) sont attirés par le son prototypique le plus proche (B) (Kuhl, 1993 dans Kuhl et Iverson, 1995 : 124).

La notion de « *crible phonologique* » en tant que processus absolu est remise en cause par le « *speech learning model (SLM)* » (modèle de l'apprentissage de la parole) de Flege (1995a) et le « *perceptual assimilation model (PAM)* » (modèle de l'assimilation perceptive) développé par Best *et al.* (1988 ; Best, 1993, 1995). Ces deux modèles partent du postulat que l'appropriation de nouveaux sons dépend de la relation perçue entre les éléments phonétiques de la/les langue(s) première(s) et la/les langue(s) étrangère(s) apprise(s). Guion *et al.* (2000) précisent que :

« **PAM** débute par l'observation du fait que certaines paires de sons d'une langue étrangère inconnue sont plus faciles à discriminer que d'autres paires. En réalité, certains contrastes étrangers sont plus faciles à discriminer, même pour des auditeurs qui ne les ont jamais entendus. D'autre contraste, par contre, sont plutôt difficiles à discriminer. »<sup>50</sup> (*ibidem*, 2712)

---

49 « *the prototype displays an attractor effect on the sounds around it. [It] perceptually pulls other members of the category toward itself [...] the magnet effect implies that the perceptual distance between outlying sounds and the prototype is reduced* »

50 « **PAM** starts with the observation that certain pairs of sounds from an unknown foreign language are easier to discriminate than other pairs are. In fact, certain foreign contrasts are easy to discriminate, even for listeners

« La *SLM* postule que les apprenants en L2 peuvent établir de nouvelles catégories phonétiques en L2 s'ils détectent les différences phonétiques entre un son en L2 et le son le plus proche dans sa L1. »<sup>51</sup> (2713).

Dans la perspective qui est la nôtre, ces trois notions ne sont pas contradictoires car la dénativisation/accommodation phonologique sera d'autant plus complexe si le nouveau son est perçu comme phonétiquement proche. Au contraire, la nativisation/assimilation sera plus faible si le son en LE est perçu comme distant des autres sons catégorisés dans la/les L1(s), le phénomène d'aimant perceptif étant plus limité dans ce dernier cas. Le traitement phonologique du son est un des processus permettant un traitement efficace et rapide de l'*input* verbal sonore. Nous précisons également que la notion d' « *articulatory settings* » (paramètres articulatoires) développée par Honikman (1964) permet de prendre du recul sur l'influence des phénomènes perceptifs dans ce qui est caractérisé comme « *l'accent étranger* ». Honikman (1964 : 73) considère que toutes les « *langues* » possèdent des caractéristiques articulatoires générales propres :

« [*articulatory setting* (les paramètres articulatoires)] sont les bases qui tentent de « persuader du » voire « déterminer le » caractère phonétique et le timbre spécifique d'une « langue » à travers les mécanismes et la posture générale orale requis pour favoriser la production confortable, économique et fluide de sons isolés constituant un ensemble reconnaissable qui constitue la prononciation établie d'une « langue ». »<sup>52</sup>

Elle propose un tableau récapitulatif des « *articulatory settings* » (paramètres articulatoires) du français et de l'anglais (Tableau 4). Cette théorie a été confirmée plus tard, de manière partielle ou totale, notamment par les travaux de Wilson (2006). Cette approche renforce l'idée que le processus de production des sons est fortement lié à la notion d'automatisme. Chaque « *langue* » présenterait des schémas articulatoires préférés qui conditionnerait, d'une certaine façon, les habitudes articulatoires chez les locuteurs de cette « *langue* ». L'apprentissage d'un nouveau système phonologique passerait par une déconstruction de ces habitudes pour s'adapter à celles de la langue visée. Il ne faut, néanmoins, pas nier des variations possibles au sein de ce système.

---

who have never heard them before. Other contrasts, on the other hand, are quite difficult to discriminate. »

51 « The *SLM* [...] hypothesizes that L2 learners can establish new L2 phonetic categories if they detect phonetic differences between an L2 sound and the nearest L1 sound »

52 « [*articulatory setting*] is the fundamental groundwork which pervades and, to an extent, determines the phonetic character and specific timbre of a language [...] the gross oral posture and mechanics [...] requisite as a framework for the comfortable, economic, and fluent merging and integrating of the isolated sounds into that harmonious, cognizable whole which constitutes the established pronunciation of a language. »

Ajustements des articulateurs	Anglais (RP)	Français (parlé à Paris)
<i>Mâchoires</i>	Assez fermées, mais relâchées (non serrées)	Modérément ouvertes
<i>Lèvres</i>	Neutres. Modérément actives	Arrondies. Très actives ; s'étirent et s'arrondissent vigoureusement
<i>État de tension de la cavité orale</i>	Relâchée	Joues contractées
<i>Articulation principale des consonnes</i>	Alvéolaire, avec la pointe de la langue	Dentale, avec la lame de la langue
<i>Langue : ancrage</i> (c'est-à-dire : position de base, quasi-permanente)	Vers le palais. Les éléments « ancrés » sont les parties latérales de la langue, qui demeurent en permanence à proximité de l'intérieur des gencives et de l'arcade dentaire supérieure	Vers la partie inférieure de la bouche. L'élément clef de l'« ancrage » est l'extrémité/lame, qui demeure positionnée près des incisives inférieures ou en contact avec elles
<i>Langue : extrémité</i>	Effilée	Non effilée
<i>Langue : corps</i>	Légèrement incurvé vers la partie inférieure de la bouche (concave par rapport au palais)	Incurvé en direction du palais (convexe par rapport au palais)
<i>Langue : dessous</i>	Incurvé vers la partie inférieure de la bouche (concave par rapport au palais)	Neutre

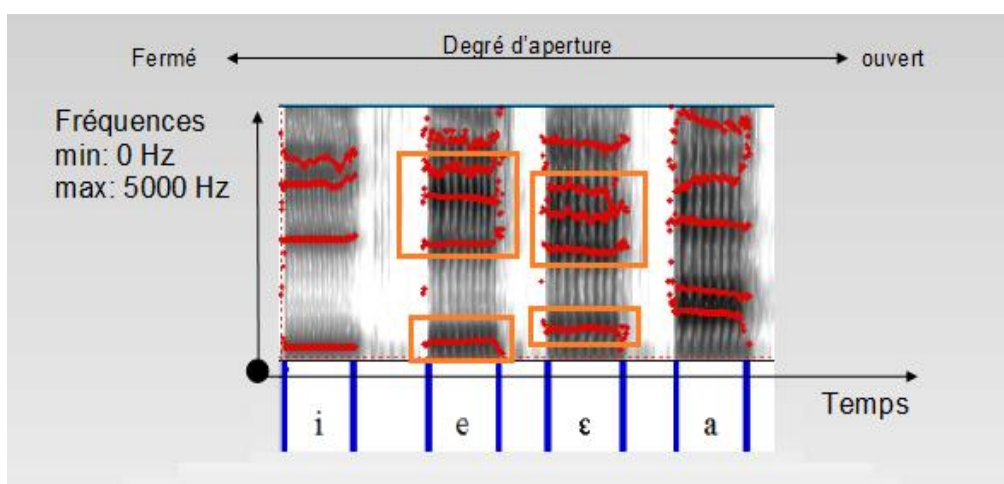
**Tableau 4:** Comparaison des ajustements articulatoires caractéristiques de l'anglais (RP) et du français parisien (traduit en français par Wilhelm, 2012 : 16 d'après Honikman, 1964 : 79).

Il nous semble de plus en plus clair que la didactique traditionnelle de la prononciation du français en contexte FLE/S a mis l'accent sur des approches favorisant la prise de conscience articulatoire malgré les nombreux travaux portant sur les phénomènes de perception dans l'acquisition d'une langue étrangère. Même si Honikman, et la notion d'« *articulatory settings* » (paramètres articulatoires), nous permettent de ne pas sous-estimer l'aspect articulatoire dans le développement langagier, il nous reste à voir en quoi un travail perceptif en didactique de la prononciation, et plus particulièrement pour le français, nous apparaît pertinent.

### 3.1.3 Cibles perceptives et point d'appui

La notion de « *paysage sonore* » développée par Lhote (1995 : 22) et qu'elle définit comme « *par paysage sonore d'une langue, nous entendons tout ce qui participe à la représentation mentale des caractéristiques sonores d'une langue donnée pour un locuteur de cette langue* » permet de resituer l'importance de la médiation de la perception des sons en langue(s) étrangère(s). Compte tenu des théories présentées ci-dessus, nous tenons à préciser

qu'un travail sur la perception seule ne permet pas d'induire une médiation pertinente de la production des unités sonores. Néanmoins, il nous semble pertinent de penser qu'un travail basé initialement sur la perception permet de prendre conscience, de manière implicite ou explicite, de comment obtenir les cibles perceptives phonologisées dans une « langue ». Comme nous l'avons montré (chapitre 1, partie 1, c) 3.), « *il existe de grandes possibilités de compensation entre les articulateurs, souvent méconnues* » (Vaissière, 2006 : 74). Les cibles articulatoires telles que présentées dans un trapèze vocalique ne seraient dès lors pertinentes que pour les grandes oppositions du type /y/-/u/ pour l'antériorité/postériorité, par exemple. Les oppositions des voyelles moyennes en français, par exemple /e/-/ɛ/, peuvent être acoustiquement réalisées par (1) une augmentation de F1 ou (2) un abaissement de F2 et F3 ou (3) par les deux phénomènes (Figure 28). Dans tous les cas, l'opposition perçue par un francophone natif sera /e/-/ɛ/.



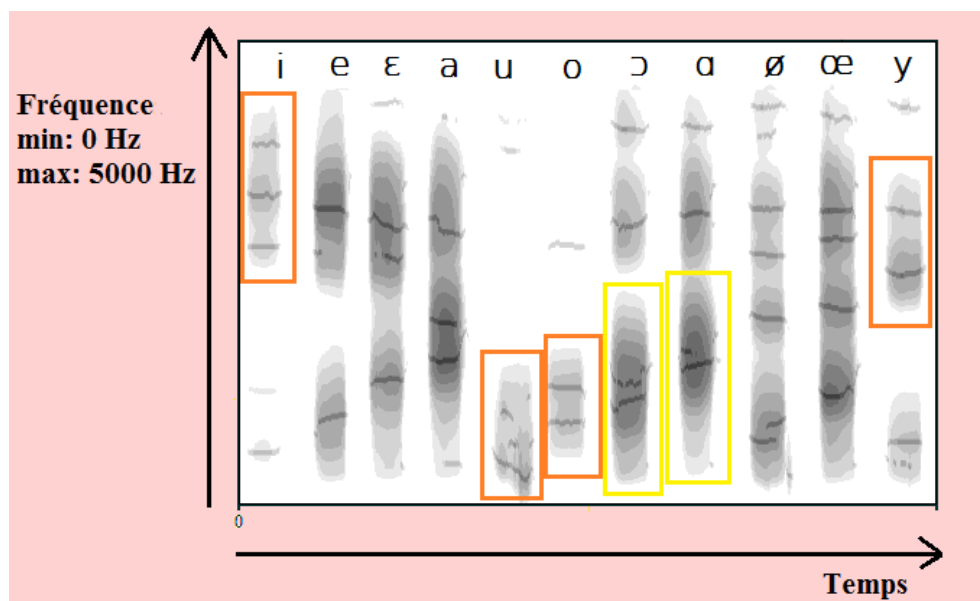
**Figure 28:** Représentation spectrographique des quatre voyelles antérieures du français standard. L'opposition /e/-/ɛ/ est réalisée par la double stratégie : une augmentation de F1 et un abaissement de F2 et F3. Calcul formantique automatique par Praat.

Vaissière (2009 : 31-2) propose une classification des voyelles de référence par des cibles acoustico-perceptives plutôt qu'articulatoires car selon elle, « *il est nécessaire de fixer des références pour la description phonétique des voyelles. De telles références pourraient correspondre à des critères acoustiques connus* »<sup>53</sup>. Concernant l'enseignement, elle ajoute que « *l'affichage en temps réel sur ordinateur (l'affichage de F1-F2 n'est pas suffisant) et la description des caractéristiques acoustiques des sons peut également aider dans l'enseignement des voyelles de référence en langue étrangère ainsi que l'enseignement de la prononciation à*

<sup>53</sup> « *it is necessary to establish fixed references for the phonetic description of vowels. Such references may correspond to well-defined acoustic criteria* »

*des apprenants mal- et normo-entendants (Lindbom & Sundberg, 1969) »<sup>54</sup>. Shiller et al. (2010 : 181) ont pu démontrer, dans le champ orthophonique, que « les enfants présentant un trouble phonologique ont, en moyenne, des difficultés significatives avec la connaissance perceptive des sons qu'ils ne prononcent pas correctement ».*

D'autre part, comme le rappelle Luzzati (2007 : 24), « tous ceux qui s'y sont confrontés savent combien l'objet « dialogue oral spontané » est insaisissable et fluctuant ». Dans cette perspective, Barth (2004) développe la notion de « point d'appui » selon lequel l'échec scolaire aurait principalement pour origine la difficulté de l'apprenant à conceptualiser ce qui est demandé. Le français a la particularité de ne présenter que des monophthongues plutôt stables même en parole continue, dont l'origine pourrait venir des sept voyelles cardinales primaires en français standard (nous ne prenons pas en compte le /a/ postérieur) et six voyelles focales. Nous pouvons supposer que ces voyelles proposent des cibles perceptives précises et pertinentes à l'acquisition d'un nouveau système phonologique à travers des caractéristiques acoustiques et phonétiques saillantes.



**Figure 29:** Cochléogramme des onze voyelles orales du français standard. Représentation automatique par Praat. En encadré orange, les renforcements d'énergie les plus saillants pour ce locuteur avec atténuation quasi-totale des autres fréquences et en encadré jaune, les renforcements d'énergie avec une atténuation moindre des autres fréquences.

<sup>54</sup> « real time computer displays (F1-F2 displays are not sufficient) and written description of the acoustic characteristics of the sounds may also help teaching reference vowels in second language teaching and the teaching of pronunciations to hearing-impaired and normal-listening students (Lindbom & Sundberg, 1969) »

La représentation par un cochléogramme (Figure 29) permet d'avoir une idée de l'importance de certaines propriétés acoustiques de certains sons vocaliques du français. Un cochléogramme est un « *outil de décomposition spectrale de l'oreille. L'excitation de la membrane basilaire dans l'oreille interne est modélisée au cours du temps, en réponse à un échantillon de parole. Le cochléogramme accentue les basses fréquences qui sont consacrées à la prosodie, contrairement à un spectrogramme classique.* » (Blanc, 2005 : en ligne, III.2.2.). Même si cette représentation n'est qu'une modélisation, elle nous permet de visualiser, notamment pour les voyelles [i], [y], [u], [o], [ɔ] et [ɑ], une prégnance de certaines fréquences, respectivement des hautes fréquences, moyennes et basses fréquences, au niveau perceptif. Ces caractéristiques sont facilement visualisables par l'utilisation de spectrogrammes en temps réel favorisant un travail qui permet de prendre conscience, de manière intuitive ou explicite, du lien perception-production d'un son. Les logiciels les plus connus actuellement sont :

- analyse en différé : Praat (Boersma & Weenink, 2013), Winsnoori (Laprie, 2009)
- analyse en temps réel : Winpitch (Martin, 2005), RTGram (UCL, 2010)

Ces caractéristiques vocaliques du français nous semblent une piste pertinente dans la médiation de la perception et la production des sons du français standard. Néanmoins, il reste à mener des recherches systématiques et empiriques sur leurs potentialités acquisitionnelles et de l'influence de la visualisation spectrographique des difficultés phonético-phonologiques. Nous noterons que Pillot-Loiseau *et al.* (2010) ont mesuré l'impact, dans l'enseignement/apprentissage du français en contexte FLE/S, d'autres types de données acoustiques (sortie nasale, niveau laryngé). La formation des enseignants de langue pourrait également être repensée vers une connaissance plus complète de la phonation et des processus d'acquisition des sons d'une langue étrangère afin de leur donner les outils utiles à la gestion des ruptures dans la prononciation de l'apprenant.

Nous avons pu voir que, jusqu'à présent, la didactique traditionnelle de la prononciation porte majoritairement son attention sur des caractéristiques articulatoires et à travers des activités de type béhavioriste. Cette approche tend à perdurer actuellement du fait que la prononciation n'a pas de légitimité assumée dans les approches communicatives et actionnelles. Néanmoins, de nombreux travaux montrent l'importance de la perception et notamment en français dont les caractéristiques acoustiques, au niveau segmental, nous semblent intéressantes. Cependant, ces approches articulatoire et perceptive ne sont pas en opposition. Il faut donc questionner les liens entre ces deux processus.



### 3.2 De l'articulatoire et du perceptif

Dans cette section, nous traiterons de trois principaux éléments qui ont permis de faire avancer la problématique du lien entre perception et production dans intégration de sons d'un nouveau système phonologique. Nous définirons et développerons l'effet McGurk, la théorie motrice de la perception de la parole et enfin les neurones miroirs.

#### 3.2.1 Effet McGurk

L'effet McGurk porte le nom de celui qui, en collaboration avec MacDonald (1976), a démontré l'impact de la visualisation de la labialité sur la perception de certains sons consonantiques. A cette époque, la perception de la parole était considérée comme un phénomène purement auditif. McGurk et MacDonald ont présenté à des sujets normo-entendants des vidéos d'une femme prononçant, de manière labiale, un [ga] mais dont le son était substitué par un [ba]. La plupart des sujets adultes discriminaient alors un [da], réponse qu'ils nommèrent « *responses fused* » (réponses fusionnées). En inversant les modalités visuelles et auditives, les auteurs ont obtenu des réponses combinées (« *combination* ») perçues tantôt comme [gabga], [bagba], [baga] ou [gaba] (Tableau 5) (McGurk & MacDonald, 1976 : 746). Les données montrent également que les adultes sont plus largement sujets à subir l'influence de l'*input* visuel comparés aux enfants entre 3 et 8 ans.

**Table 2** Percentage of responses in each category in the auditory visual condition

Stimuli		Subjects	Responses		Responses Fused	Combination	Other
Auditory	Visual		Auditory	Visual			
ba-ba	ga-ga	3-5 yr (n=21)	19	0	81	0	0
		7-8 yr (n=28)	36	0	64	0	0
		18-40 yr (n=54)	2	0	98	0	0
ga-ga	ba-ba	3-5 yr (n=21)	57	10	0	19	14
		7-8 yr (n=28)	36	21	11	32	0
		18-40 yr (n=54)	11	31	0	54	4
pa-pa	ka-ka	3-5 yr (n=21)	24	0	52	0	24
		7-8 yr (n=28)	50	0	50	0	0
		18-40 yr (n=54)	6	7	81	0	6
ka-ka	pa-pa	3-5 yr (n=21)	62	9	0	5	24
		7-8 yr (n=28)	68	0	0	32	0
		18-40 yr (n=54)	13	37	0	44	6

**Tableau 5:** Pourcentage de réponses pour chaque catégorie dans les conditions auditivo-visuelles et mise en évidence de l'effet McGurk (McGurk & MacDonald, 1976 : 747).

A l'époque, les auteurs précisent que ces résultats doivent être confirmés par de futures études de manière statistique. Ils tiennent, néanmoins à préciser, que « *nous avons nous-mêmes expérimenté ces effets, sur des centaines d'essais, auxquels nous ne nous sommes jamais*

*habitué dans le temps et ce malgré la connaissance objective de l'illusion impliquée.* »<sup>55</sup> (McGurk & MacDonald, 1976 : 747). Ces résultats furent plus tard confirmés et repris par de nombreuses études telles que Norrix & Green (1996), Massaro (1987, 1998) ou Hardison (1996), entre autres. La découverte de l'effet McGurk fut un grand pas en avant dans la compréhension des phénomènes de perception de la parole puisqu'ils permettaient de montrer que les théories basées sur l'audition seule n'étaient pas adéquates. Il fallait, à partir de ce fait, théoriser une nouvelle approche, prenant en compte les phénomènes en jeu dans l'effet McGurk.

### 3.2.2 Théorie motrice de la perception de la parole

La théorie motrice de la perception a été formulée initialement par Liberman (1967, 1985) mais elle est la finalité d'un ensemble de travaux menés, entre autres par Liberman, Cooper & Delattre (1952, 1954, etc.), ainsi que d'autres chercheurs du Laboratoire Haskins. Bien que la théorie motrice de la perception ait beaucoup évolué depuis les premières mentions qui y sont faites :

*« chaque version a déclaré que les objets de la perception de la parole sont plus articulatoires qu'acoustiques ou auditifs. Plus précisément, l'hypothèse était que les faits articulatoires reconstruits par les auditeurs humains sont les commandes neuromotrices des articulateurs plutôt que des réseaux plus périphériques tels que les mouvements articulatoires en action.* »<sup>56</sup> (Diehl et al., 2004 : 150).

Dans cette théorie, l'auditeur interpréterait le signal de la parole par reconstruction des différents mouvements articulatoires à l'origine du signal. L'espace des consonnes posséderait des frontières naturelles (passage d'un /d/ à un /t/ par exemple) tandis que l'articulation des voyelles serait plus un continuum. Ce point fut contredit, notamment pour les consonnes fricatives, glides et liquides, par Fujisaki et Kawashima (1971). Ils ont également montré que mêmes les voyelles pouvaient être perçues de manière catégorielle dans des conditions défavorables (Nguyen, 2005 : 4).

Selon Liberman & Mattingly (1985), le lien qui lie perception et production n'est pas une association apprise de par le fait que ce qu'un individu entend est ce qu'il peut produire. Ce lien est, selon les deux auteurs, inné : *« le lien est spécifié de manière innée et ne nécessite qu'un*

---

<sup>55</sup> *« we ourselves have experienced these effects on many hundreds of trials ; they do not habituate over time, despite objective knowledge of the illusion involved »*

<sup>56</sup> *« every version has claimed that the objects of speech perception are articulatory events rather than acoustic or auditory events. More specifically, it was hypothesized that the articulatory events recovered by human listeners are neuromotor commands to the articulators [...] rather than more peripheral events such as actual articulatory movements or gestures »*

développement épigénétique pour le mettre en action »<sup>57</sup> (Liberman & Mattingly, 1985 : 3). Ce traitement particulier permettrait, en outre, un traitement plus rapide des informations acoustiques de la parole plutôt que de passer par un traitement coûteux de toutes les informations phonétiquement pertinentes.

En 1980, la « *direct realist theory of speech perception* » (théorie directe et réaliste de la perception de la parole) (DRT) développée par Carol Fowler (1981, 1984 entre autres) vient en contradiction avec la théorie motrice de la perception. Le principal désaccord entre ces deux approches vient du fait que la DRT nie l'idée d'un mécanisme particulier dédié au traitement de la parole humaine. Selon Fowler, cette particularité est partagée par tous les systèmes sensoriels, de plus, « *ils utilisent la structure dans le médium, qui a été légitimement causée par des événements environnementaux, comme informations sur les événements.* »<sup>58</sup> (Fowler, 1996 : 1732 dans Diehl, Lotto & Holt, 2004 : 152). Les paradigmes de ces deux approches sont synthétisés dans Tableau 6.

	Mécanismes spécifiques	Mécanismes généraux
Gestuel	Théorie motrice (« <i>Motor theory</i> »)	Réalisme direct (« <i>Direct realism</i> »)
Non-gestuel	Spécialisations éclectiques (« <i>Eclectic specializations</i> »)	Approche générale (« <i>General approach</i> »)

**Tableau 6:** Taxonomie des principales approches théoriques de la perception de la parole (traduit par l'auteur d'après Diehl, Lotto & Holt, 2004 : 1558).

Pour terminer, nous rappellerons que selon Nguyen (2005 : 5) :

« [les travaux de Fujisaki et Kawashima] entraînent en contradiction avec la théorie motrice en donnant à penser que la perception catégorielle n'était pas un mode de traitement propre à une certaine classe de sons. Ils ont permis d'établir qu'un même son pouvait être perçu de façon plutôt catégorielle, ou plutôt continue, en fonction de la situation expérimentale. Fujisaki et Kawashima ont interprété leurs résultats dans le cadre d'un modèle qui attribuait un rôle central à la mémoire auditive à court terme (Pisoni, 1975) »

Bien que l'effet McGurk et la théorie motrice de la perception aient bénéficié des avancées technologiques avec la mesure de données physiologiques (Green, K.P. ; 1996 : 2591-2603), par exemple « *mismatch negativity* » (négativité de discordance) (MMN) qui représente le traitement des différences acoustiques dans des stimuli auditifs (Näätänen *et al.*, 1978 : 313–29), il faudra attendre la démocratisation de l'IRM(f) pour en découvrir plus sur les liens perception-production.

<sup>57</sup> « *the link is innately specified, requiring only epigenetic development to bring it into play* »

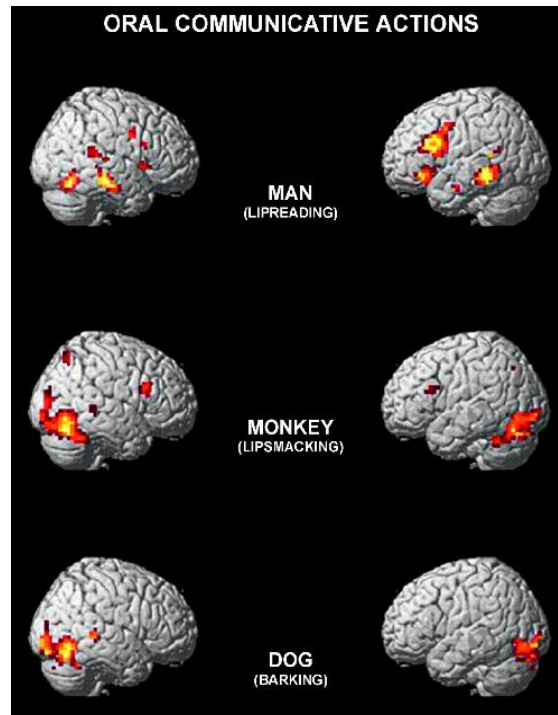
<sup>58</sup> « *they use structure in the media that has been lawfully caused by events in the environment as information for the events* »

### 3.2.3 Les neurones miroirs

Comme de nombreuses découvertes, l'existence de neurones miroirs a été découverte, par hasard, par Rizzolatti *et al.* (2006) en remarquant qu'une zone cérébrale était activée lorsqu'un singe faisait un mouvement mais également lorsqu'il regardait un autre singe ou un humain réaliser ce même mouvement. Les zones F5 et AIP (aire intra-pariétale antérieure) furent, entre autres, isolées dans cette activation spécifique. Rizzolatti *et al.* (2006 : 45) résumant leurs actions dans l'acte de prendre une tasse :

*« dans l'exemple de la tasse à café, les affordances visuelles offertes à notre système moteur concernent l'anse, le corps central, le bord supérieur, etc. Dès que nous voyons la tasse, ces affordances activent sélectivement des groupes de neurones de l'aire AIP. L'information visuelle ainsi parcellisée est alors transmise aux neurones visuo-moteurs de F5 ; ces derniers, toutefois, ne codent pas des affordances particulières, mais les actes moteurs qui leur sont congruents. »*

Il existerait donc une certaine « *phonologie* » du geste permettant de traiter un mouvement complexe de manière simple et efficace. En ce qui concerne la parole, Buccino *et al.* (2004) ont mis en évidence que le gyrus frontal inférieur (aire de Broca) s'activait fortement lorsqu'un individu voit un pair bouger les lèvres ; ce dernier diminuait à la vue d'un singe claquant les lèvres et s'annulait à la vue d'un chien qui aboie (Figure 30). Rizzolatti & Sinigaglia (2006 : 149) soulignent le fait que « *ces expériences ne font que confirmer le rôle décisif de la connaissance motrice pour la compréhension de la signification des actes d'autrui* ». Les auteurs tiennent à signaler, néanmoins, que ce mode de traitement n'est pas le seul existant et coexisterait avec un traitement plus « *intellectif* » (*Ibidem*). Néanmoins, l'utilisation du système de neurones miroirs permet d'impliquer le sujet à la première personne comme s'il exécutait l'action lui-même. Les neurones miroirs pourraient également être à l'origine de la prise en charge de l'empathie (Rizzolatti & Sinigaglia, 2006 : 183-203).



**Figure 30:** Aires corticales activées durant l'observation d'actes communicatifs oraux (d'après Buccino et al., 2004 dans Rizzolatti & Sinigaglia, 2006 : figure 5.10).

D'autres travaux, portant sur l'imitation, en tant que (1) capacité d'un individu à reproduire un acte en voyant quelqu'un le réaliser ou en tant que (2) capacité à apprendre un nouveau type d'action et le reproduire fidèlement, ont montré que « *la découverte des neurones miroirs suggère [...] un mécanisme de transformation direct des informations visuelles en actes moteurs potentiels* » (Rizzolatti & Sinigaglia, 2006 : 154). Les auteurs ajoutent qu' : « *il apparaît évident qu'une transformation de l'information visuelle en réponse motrice appropriée se produit dans le système des neurones miroirs* » (Rizzolatti & Sinigaglia, 2006 : 159). Ces phénomènes ont notamment été montrés sur l'apprentissage de comment jouer certains accords sur une guitare par des non-musiciens à travers l'imitation (Jacoboni et al. 1999, 2001).

Plus précisément pour la parole, Rizzolatti et Buccino (2005) ont mis en évidence l'existence de neurones miroirs « *échos* » s'activant « *en présence de sons produits moyennant des gestes analogues exécutés par d'autres* » (Rizzolatti & Sinigaglia, 2006 : 179). La présence de ces neurones miroirs avait été mise en avant dans une étude clinique menée par Fadiga et al. (2002). Cette dernière montrait que l'écoute de mots et pseudo-mots contenant un double « *R* » provoquait une augmentation de l'amplitude des potentiels évoqués moteurs (PEM) des muscles de la langue contrairement à l'écoute de sons bitonaux et de mots ou pseudo-mots contenant un double « *F* ». Ces résultats étaient expliqués par le fait que le phonème /r/ nécessite une implication de la langue plus importante dans sa réalisation que le phonème /f/ (Figure 31).

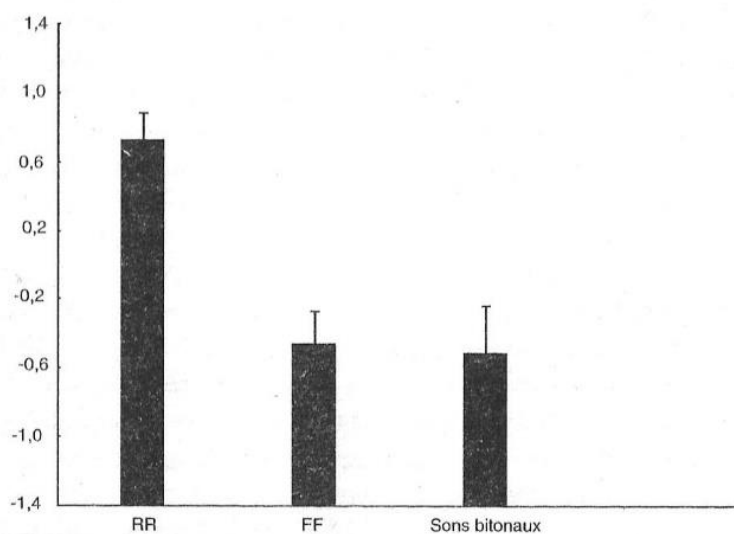


Fig. 6.4. Potentiels moteurs évoqués enregistrés par les muscles de la langue durant l'écoute de matériaux verbaux et de sons bitonaux. Les données concernent tous les sujets. RR se réfère à des stimuli verbaux contenant des consonnes fricatives linguopalatales, tandis que FF se réfère à des matériaux verbaux contenant des consonnes fricatives labiodentales. (Adapté de Fadiga *et al.*, 2002.)

**Figure 31:** Potentiels évoqués moteurs enregistrés par les muscles de la langue durant l'écoute de matériaux verbaux et de sons bitonaux. (Rizzolatti et Sinigaglia, 2006 : 180).

A travers l'ensemble de ces approches et phénomènes (effet McGurk, théorie motrice de la perception, neurones miroirs), nous pouvons constater que la traditionnelle dichotomie entre production et perception n'est plus pertinente. Ces deux processus sont caractérisés par des limites biologiques et physiologiques dont la potentialité est dirigée par l'environnement. Les individus ont intégré un programme processuel liant les unités produites à celles perçues. Ces connexions restent dans le domaine du vivant et par conséquent ne sont pas totalement fiables. Le système phonologique permet la stabilisation de ces connexions. La didactique de la prononciation devrait donc, selon les avancées scientifiques, prendre en compte le corps en tant qu'interface entre les représentations cognitives et l'environnement. Les méthodes les plus répandues actuellement, en didactique du français en contexte FLE/S, ne semblent pas intégrer cette perspective. Il nous reste donc à voir si d'autres approches existent et si la musique peut permettre ce type de travail.

### 3.3 Didactique de la parole et de la musique

Dans cette section, nous passerons en revue les méthodes en marge de la didactique de la prononciation qui prennent en compte d'autres paramètres qu'un enseignement purement réflexif. Nous verrons, en outre, que les approches qui utilisent la musique dans l'enseignement/apprentissage de la prononciation en contexte FLE/S fait partie de cet ensemble de méthodes. Nous terminerons en montrant que la didactique de la musique peut également apporter des éléments constructifs dans la médiation de la perception et la production des sons en français comme langue seconde ou étrangère.

#### 3.3.1 Les méthodes en marge de la didactique de la prononciation

Nous avons vu que, depuis de nombreuses années, coexistent différentes méthodes, théories, approches liées à la didactique de la prononciation, et plus précisément en contexte FLE/S. Ces méthodes suivent généralement, de manière plus ou moins synchronique, les paradigmes de la recherche scientifique. Néanmoins, certaines approches semblent se maintenir dans les pratiques pédagogiques, et ce en dehors des pratiques dites traditionnelles. Ces dernières reposent sur des corrélations avec le yoga, le théâtre, la chanson ou la musique, par exemple. La **suggestopédie**, développée par Georgi Lozanov dans les années 1970, met en avant l'importance de la création d'un environnement relaxant et agréable permettant de limiter les phénomènes d'anxiété, de blocages psychologiques, etc. Cette méthode repose sur des exercices liés au souffle, à la respiration, la multisensorialité et la chanson (Lauret, 2007 : 118). Le **Silent Way**, créée par Caleb Gattegno, va à contre-courant des autres méthodes, en prônant l'absence de la voix de l'enseignant en tant que modèle. L'enseignant reste le plus silencieux possible ; il ne transmet pas de connaissance ni de modèle mais intervient seulement pour attirer l'attention sur des éléments qui nécessitent une attention particulière. Les sons sont différenciés, de manière visuelle, à travers une palette de couleur et leurs représentations graphémiques (Lauret, 2007 : 119). Ces deux approches nous semblent intéressantes par le fait qu'elles mettent l'accent sur l'apprenant et son rapport au corps. Néanmoins, elles ne sont que des outils, parmi d'autres, de par leurs limites didactiques en termes d'acquisition d'un nouveau système phonologique. Dans cette perspective, nous citerons, l'exemple du **Bonhomme sonore** (Guimbretière & Kaneman-Pougatch, 1991 : 70) (annexe 7) qui organise sur un axe haut/bas et gauche/droite les phonèmes du français. Le principe étant de faire conceptualiser à l'apprenant la notion de point d'articulation des voyelles. Il faudra alors associer un geste/une partie du corps avec la production sonore (Figure 32).



**[i]**  
Ce qui distingue essentiellement [y] de [i], c'est la labialité. Le son [i], non labial le plus tendu des sons sort au sommet du crâne, de chaque côté de la tête, comme si vous aviez à cet endroit des cornes. Sentez-les vibrer quand vous prononcez [i]. Et maintenant, pour prononcer [y], ramenez le son au milieu de votre front et faites-le sortir par cette petite bouche gourmande.



**[u]**  
Comme [y], le [u] est labial. Ce qui distingue alors ces deux sons, c'est l'acuité. Le son [u] qui est grave trouve son chemin le long des jambes et finit par sortir par une petite bouche imaginaire qui, elle, se trouve sous la plante des pieds.

**Figure 32:** Représentation corporelle des sons vocaliques [i] et [u] selon le bonhomme sonore (Guimbretière & Kaneman-Pougatch, 1991 : 16).

Le /u/ est donné comme un phonème d'acuité grave, avec un placement de la langue en position arrière et donc comme terrien. Il est demandé alors à l'apprenant de conceptualiser le /u/ par une vibration le long des jambes s'exprimant par les pieds tandis que le phonème /y/ a une acuité aiguë ce qui le rend aérien, la vibration est ressentie dans le haut du corps. Il serait donc possible de faire intervenir une implication corporelle afin d'améliorer la réalisation des phonèmes. D'autres propositions existent telles qu'une représentation du trapèze vocalique sous une forme de **lyre** dans un manuel de langue intitulé « *La Clé des champs* » (Geoffroy *et al.*, 1991 : 16) (annexe 8). L'ensemble de ces propositions repose sur les représentations liées à la hauteur spectrale des voyelles. Cette conception est particulièrement pertinente pour les voyelles focales dont le renforcement d'énergie rend la perception de la hauteur spectrale déterminante.

Ces différentes approches sont un premier pas vers la conception de l'individu dans son intégralité et dans le fait de donner corps au son. Elles permettent, notamment, de déconstruire l'espace de « *cours* » souvent fixe : les étudiants sont assis sur une chaise et bloqués par une table devant eux. Elles montrent également que d'autres approches de l'enseignement/apprentissage sont possibles. Il nous reste à voir si, plus spécifiquement, certaines méthodes « *en marge* » intègrent la musique instrumentale comme support ou point d'appui d'apprentissage.

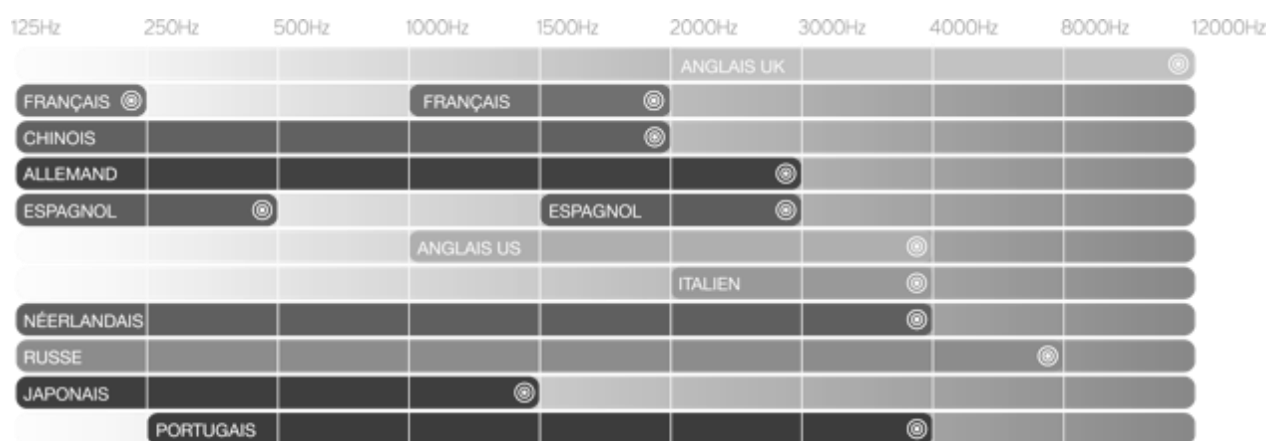


### 3.3.2 La musique en didactique des langues : le cas du FLE/S

De manière globale, la didactique de la prononciation fait plus intervenir la chanson que la musique instrumentale en tant que support didactique (pour la chanson se référer, entre autres, à : Lauret, 2007 : 125-30 ; Guimbretière, 1994 : 84). Pour l'anglais, nous citerons brièvement Carolyn Graham (2000) et ses **Jazz chants** qui reposent sur l'analogie entre les temps fort-faible du jazz et de l'anglais. Chaque énoncé du quotidien peut intervenir dans un schéma rythmique jazz qui permet à l'apprenant d'anticiper le rythme de la langue anglaise et ainsi acquérir naturellement la musicalité de la « *langue* ». Il peut aussi améliorer la prononciation des phonèmes eux-mêmes dont la réalisation peut varier suivant que la syllabe soit accentuée ou non (cf. réduction et neutralisation des sons vocaliques dans des syllabes non-accentuées). L'utilisation de la musique rend plus prégnante ces notions linguistiques. Pour le français, Régine Llorca (1998, 1999) et ses **Ritnimots** a utilisé sa double formation en danse et en linguistique afin de porter une réflexion sur les liens entre expression corporelle et prononciation en passant notamment par le rythme. Elle préconise, par exemple, pour l'intégration de l'accentuation de syllabe en fin de groupe rythmique de placer ce genre de phrase dans une hauteur tonale plus élevée puis descendre tout le long de la phrase puis brusquement sur la syllabe accentuée. Cette action d'élargir la palette de hauteur tonale pour un groupe rythmique permet de renforcer la chute brutale en fin d'énoncé et donc dérouler les syllabes de façon naturelle, spontanée et automatique. Il est important d'accompagner cet exercice par des gestes qui vont venir impliquer le corps dans la production et donc aider l'assimilation. L'important dans l'assimilation du rythme langagier à travers la musique consistera donc en une implication multimodale afin de sensibiliser l'apprenant à une nouvelle approche de la chaîne du parler en impliquant son corps, sa gestuelle et sa capacité d'anticipation musicale. D'autres travaux sur la voix chantée/parlée ont permis de mettre en évidence des ponts entre ces deux performances (Pillot, 2004 ; Cornaz *et al.*, 2010 ; Cornaz & Caussade, 2014 ; Cornaz & Fonio, 2014 ; entre autres).

L'utilisation de la musique instrumentale se retrouve également dans l'**enseignement intégral** des pédagogies Montessori, Freinet ou Dewey pour lesquelles, selon Pamula (2008 : 133-140), l'apprentissage du français dans une structure comme celle-ci permettrait d'intégrer les compétences phonologiques au sein d'une dynamique communicative à travers la musique et les arts plastiques, par exemple. Plus récemment, la méthode **SpeedLingua** a remis au goût du jour les théories développées par Tomatis, dans les années 1950, et selon lesquelles il est possible de transformer la phonation (production orale) par une remédiation auditive à travers

des outils de traitement du signal sonore. Malgré le fait que ce postulat ne puisse être validé sur le plan empirique et que Tomatis ait fait l'objet de nombreuses controverses (Abgrall, 2001 : 145-9), la méthode SpeedLingua tend à se démocratiser dans le monde du FLE/S par des utilisations médiatisées à l'Alliance Française de New York ou l'Armée de l'Air française<sup>59</sup>, entre autres. La méthode *SpeedLingua* propose d'intervenir sur l'écoute, la compréhension et la prononciation en stimulant l'oreille interne avec les fréquences qui ne font pas partie des fréquences dominantes de la L1 de l'apprenant (Figure 33). Chaque session se déroule en trois étapes : la phase d'écoute (15 minutes : l'apprenant écoute une musique aux fréquences dominantes de la langue cible), la phase de prononciation (la voix de l'apprenant est modulée de manière informatisée vers les fréquences dominantes de la langue cible) et celle dite de « résultats ».



SpeedLingua Les fréquences dominantes de chaque langue

Si l'on modifie l'audition, on modifie inconsciemment et naturellement la phonation.

**Figure 33:** *Fréquences dominantes des langues selon la méthode SpeedLingua et d'après le principe Tomatis*<sup>60</sup>.

La notion de fréquence dominante des langues n'a jamais pu être démontrée de manière formelle car, comme le rappelle Lauret (2007 : 116), « le spectre est globalement similaire pour toutes les langues même si de nombreuses différences statistiquement significatives ont été observées et qu'il est largement dominé par les caractéristiques individuelles du mécanisme vocal ». De plus, nous rappellerons que chez des normo-apprenants, les problèmes de prononciation (perception) se trouvent au niveau cognitif plutôt que dans la physiologie du conduit auditif (cf. mention de « l'oreille interne »). Nous noterons également qu'un travail de mémoire de Master

59 Site de la méthode SpeedLingua, témoignages :

[http://www.speedlingua.com/Speedlingua/Temoignages\\_f.html](http://www.speedlingua.com/Speedlingua/Temoignages_f.html) (consulté le 24/04/13)

60 [http://www.speedlingua.com/Speedlingua/Procédé\\_f\\_files/FREQ\\_FRweb.png](http://www.speedlingua.com/Speedlingua/Procédé_f_files/FREQ_FRweb.png) (consulté le 24/04/13)

réalisé par Porée (2012) n'a pas pu démontrer d'effet significatif sur la prononciation d'apprenants utilisant la solution *SpeedLingua* et que de manière générale « *la phase musicale a été considérée comme superflue, les apprenants ne souhaitaient pas s'y soumettre et l'abrégeaient autant qu'ils le pouvaient, malgré notre insistance* » (Porée, 2012 : 84). L'auteur admet volontiers que « *la phase musicale et l'utilisation de l'oreille électronique sont des arguments vendeurs auxquels certains apprenants ont peut-être été sensibles* » (Porée, 2012 : 113). Selon nous, les effets positifs qui peuvent être observés avec la méthode *SpeedLingua* proviennent des exercices traditionnels systématiques proposés dans la phase 2 et que sans étude comportant un groupe témoin, il est impossible de dire si une méthode propose réellement des avantages acquisitionnels.

Nous ne reviendrons pas ici sur l'ensemble des pratiques ancrées dans la doxa et dont les origines scientifiques sont quelque peu problématiques. Nous prendrons, cependant pour exemple, un document publié par la Commission Scolaire des Affluents regroupant près de 67 établissements scolaires au Québec et qui met en évidence les différentes utilisations de la musique en classe. En s'appuyant sur les travaux de S. Dubé (2001) et de Vos *et al.* (1999), la commission met en avant six grands effets<sup>61</sup> :

- 1) La musique détend et fait baisser le niveau de stress qui inhibe l'apprentissage,
- 2) La musique agit directement sur le corps,
- 3) La musique peut favoriser certains apprentissages complexes,
- 4) La musique agit sur les émotions,
- 5) La musique est un langage universel,
- 6) La musique agit de façon très puissante sur la mémorisation.

Même si certaines de ces affirmations peuvent être validées en partie par la recherche scientifique, Bernardi *et al.* (2006) ont démontré un impact du tempo d'une musique sur le rythme cardiaque et que cet impact était plus important chez les musiciens que les non-musiciens ; il n'est pas pour autant possible de généraliser en disant que « *la musique détend* ». De la même manière, le stress peut également stimuler certains apprentissages: Shors (2004 : 143) affirme que « *tout d'abord, il n'est pas vraisemblable de penser qu'il n'y ait qu'un ou deux liens entre le stress et l'apprentissage. Ensuite, ils ne sont pas exclusivement négatifs.* »<sup>62</sup>. Au-delà de ces propositions didactiques basées sur des raccourcis scientifiquement critiquables, nous souhaitons montrer que d'autres initiatives semblent plus prometteuses comme le

---

61 [http://www.csaffluents.qc.ca/im/PDF2005/ens\\_outils/utilisation\\_musique\\_classe.pdf](http://www.csaffluents.qc.ca/im/PDF2005/ens_outils/utilisation_musique_classe.pdf) (consulté le 25/04/13)

62 « *first, it is unlikely that there are one or even two defining relationships between stress and learning. Second, they are not exclusively negative.* »

« *European Music Portfolio : a creative way into languages* »<sup>63</sup> (Ludke, *et al.*, 2012) qui a été financé par la Commission Européenne. Ce projet, regroupant plus de quatorze chercheurs, propose un cadre didactique (manuel de l'enseignant, fiches pédagogiques, etc.) pour soutenir l'apprentissage **de et par** la musique et les « *langues* » à travers une approche commune. Il intègre les travaux les plus récents dans le domaine. Il permet notamment de nous questionner, au-delà des utilisations déjà effectives de la musique en didactique de la prononciation, de ce que la didactique de la musique peut nous apporter.

#### 3.3.3 Ce que nous apprend la didactique de la musique

Les apports de la musique instrumentale et plus particulièrement de la didactique de la musique à la didactique des langues sont nombreux mais particulièrement pertinents en ce qui concerne l'approche du monde sensible. Corral Fulla (2008 : 143) dénote cinq grands points communs entre ces deux didactiques :

- la réflexion portée sur ce qu'on apprend.
- la participation du corps et du geste à la production.
- l'expressivité/émotivité implicite dans le parler et le jouer d'un instrument.
- l'entraînement répétitif : visant à atteindre une flexibilité et maîtrise manuelle/vocale, ce qui implique le développement des muscles participant à la production dans les deux disciplines.
- l'assimilation dans les deux cas de structures conceptuelles.

Les cas de la didactique de l'oral/prononciation et de la musique partagent un même médium qui est le son et dont les caractéristiques posent des problématiques communes. Le son est un élément extrêmement dépendant du temps et de l'espace. Kivy affirme que « *lire et la performance ont plus en commun que ne laisse paraître le sens commun* »<sup>64</sup> (Kivy, 2006 : 1). Selon lui, la musique, le théâtre et la lecture se situent sur un continuum de performance. Pour chacun d'entre eux, il existe l'œuvre (« *a work* »), sa matérialité (« *a score* ») et un exemplaire (« *token of the type* »). Le son varie de manière constante et constitue une partie ineffable, définie par Jankélévitch comme « *quelque chose de caché voire d'inexprimable* » (Corbel, 2005 : 76). En musique, cet ineffable prend corps dans l'interprète en opposition à l'exécutant et dont la dichotomie, selon Stravinsky, réside dans le fait que « *l'exécutant se soumet au texte tandis*

---

63 « *Le portfolio européen de musique : une méthode créative vers les langues* »

64 « *reading and performance have more in common than common sense suspects* »

que l'on attend de l'interprète, qu'il fasse une traduction du texte, qu'il fasse de la perfection de cette traduction matérielle, une complaisance amoureuse » (Corbel, 2005 : 78). Cette problématique pose la question de ce qui est enseignable dans la didactique de la prononciation et comment l'enseigner. L'héritage du behaviorisme a favorisé des pratiques visant l'exécutant à travers des exercices mécaniques et répétitifs (repris par la SGAV) et les méthodes alternatives ont plutôt favorisé l'interprète. Actuellement, aucune méthode en didactique de la prononciation en contexte FLE/S ne propose réellement d'articulation entre ces deux approches, qui sont selon nous, complémentaires. Ces deux postures font émerger des problématiques liées à l'individu et notamment chez l'adulte. Malinen (2000 : 134-5) rappelle l'existence d'une « *protective belt* » (périmètre de protection) (Malinen, 2000 : 134-135) autour des connaissances, construits, savoirs acquis. L'un des enjeux de l'apprentissage est donc d'amener l'apprenant vers « *a cracking point* » (un point de rupture) (Huhtanen, 2008 : 59). Cet aspect nous rappelle le caractère non-linéaire de l'apprentissage ponctué par des ruptures (cf. théorie du chaos ; Larsen-Freeman, 1997) dont le déroulement ne peut être prévisible. De la même manière, le caractère non-saisissable de l'oral est un facteur très important dans la pratique orale. En musique, Reitan (2008) a étudié les facteurs de bien-être et de stress dans les cours d'écoute musicale. Elle met en évidence que généralement les cours oraux ou audio sont vecteurs de stress mais étrangement de bien-être à la fois. L'élément déterminant étant que « *l'entraînement audial met en difficulté l'estime de soi* »<sup>65</sup> (Reitan, 2008 : 86). Par conséquent, réfléchir en didactique des langues sur l'influence du stress dans la participation orale en cours de prononciation, par exemple, pourrait faire intervenir la notion d'estime de soi. Les facteurs majeurs de stress sont (*Ibidem* : 87) : la peur de faire des erreurs, la pression à réussir, la peur de ne pas maîtriser l'exercice, une basse estime de soi. Réduire le stress consisterait donc à maîtriser au mieux le sujet, créer un climat de confiance dans la classe, disposer d'un niveau homogène, et favoriser une faible pression pour la réussite du premier coup (Reitan, 2008 : 86).

Nous remarquerons qu'en anglais, la prise en compte de la perception et de la production est marquée respectivement par la dichotomie « *aural* » et « *oral* » surtout en musique. En français, le terme « *oral* » est utilisé en tant qu'hyperonyme généraliste intégrant les phénomènes de compréhension, perception et production des unités orales. Néanmoins, il est possible de retrouver, en musique et en français, une telle prise en compte avec le terme d'oralité comme « *une action de transmission d'informations verbales ou musicales* » (Terrien, 2012 : 35) et celui d'auralité en tant que « *relation immédiate de l'énergie humaine du son dès que l'oreille –*

---

65 « *aural training [...] challenges the person's self-esteem* »

*auris – est maitre du jeu, c'est-à-dire aussi du je-ici-maintenant* » (Savouret, 2002 : 10 dans Terrien, 2012 : 38) ou encore « *l'auralité induit [...] une relation intrapersonnelle. Elle est d'abord la marque d'une réflexion sur ce que l'individu entend de la production sonore, et l'analyse qu'il développe de sa perception* » (Terrien, 2012 : 45). Dans une perspective vygotkienne, l'oralité et l'auralité posent la question du double rapport intra- et interpsychique. Terrien (2012 : 45-8) propose un va-et-vient entre ces rapports en ce qui concerne le chant. Selon lui, la situation d'apprentissage du chant propose un travail de l'inter vers l'intra donc de l'oralité vers l'auralité. Les apprenants vivent et perçoivent des éléments à travers le groupe et/ou l'enseignant (inter) qu'ils vont devoir mémoriser, analyser et intérioriser (intra). A l'inverse, dans l'interprétation, la relation va de l'intra vers l'inter soit de l'auralité vers l'oralité : « *savoir s'écouter, savoir écouter les autres pour s'adapter, s'apprend dans le cadre d'une pratique vocale basée sur l'oreille* » (Terrien, 2012 : 47). Nous noterons, bien sûr, que l'interaction orale met en œuvre de manière synchronisée l'oralité et l'auralité, cependant cette dichotomie permet de prendre une distance épistémique nécessaire entre les deux processus qui y sont respectivement associés. Le terme « *oral* » est un terme trompeur et, selon nous, trop généraliste pouvant favoriser un comportement pédagogique principalement axé sur la production et minimisant ou mettant de côté les phénomènes de perception de soi. Le caractère pragmatique des interactions verbales fait prédominer un traitement du sens limitant la prise de conscience des unités sonores réellement perçues. Travailler sur la prononciation d'un individu, c'est avant tout lui faire prendre conscience de **son** utilisation de **son** appareil phonatoire et les liens entre **ses** perceptions et **ses** productions. La musique portant sur une intentionnalité émotionnelle et s'intègre de manière plus explicite dans le monde du sensible, sa didactique a dû s'adapter à ces caractéristiques prégnantes. Selon nous, la didactique de la musique permet de questionner certaines problématiques de la didactique de la prononciation qui pourraient rester muettes dans une conception tournée principalement vers les interactions.

## **Conclusion générale de la première partie :**

Il est clair que notre positionnement ne vise à pas proclamer la perception comme prédominante par rapport à la production ni même qu'un travail sur la perception entraînerait forcément des réorganisations dans la production. Nous avons cependant vu que tout tend à croire que ces deux phénomènes sont concomitants et interconnectés (section 3.2). Ce lien semble exister depuis leur origine commune que nous avons appelé, d'après les travaux de Mithen (2006), le « H.M.M.M. » (section 1.1) ; même si en synchronie, les travaux en neurocognition tendent plutôt vers un traitement partiellement modulaire (Peretz & Coltheart, 2003) mais non encapsulé (Chobert & Besson, 2012), de la musique et la parole (section 1.2). Cette modularité serait, par ailleurs, plus le produit du développement cérébral plutôt que le reflet de spécialisations déterminées de manière innée comme le rappelle Patel (2008 : 357). Néanmoins, nous avons vu que l'entraînement musical peut améliorer les capacités auditives dans d'autres domaines, comme la parole par exemple (Kraus et Chandrasekaran, 2010 ; Chobert & Besson, 2012). Nous pensons donc que l'utilisation de la musique instrumentale dans la médiation de la perception et production des sons du français en contexte FLE/S, peut permettre de sensibiliser les individus sur les caractéristiques acoustiques et phonétiques du discours et donc limiter et/ou prendre du recul sur les effets du processus de phonologisation mis en évidence par la nativisation/assimilation phonologique (section 3.1). Ce travail apparaîtrait comme pertinent dans l'acquisition des voyelles du français standard dans un contexte FLE/S puisqu'un certain nombre d'entre elles possèdent des caractéristiques acoustiques (stables, cardinales et focales) qui peuvent être autant de traits saillants comme point d'appui à l'apprentissage (section 1.3). L'utilisation de la musique instrumentale permettrait également de centrer l'individu vers une réflexion intrapsychique avant de l'intégrer dans une directionnalité interpsychique (section 3.3). L'ensemble de ces processus ascendants et descendants prennent sens dans une approche intégrée des théories émergentiste (section 2.2) et socioconstructiviste (section 2.3) qui permettent de légitimer un travail pluri-intertransdisciplinaire (section 2.1). Il nous reste toutefois à mesurer un tel impact à travers des protocoles expérimentaux au cœur de ce travail de thèse.

### **Synthèse de positionnement**

Ce cadre théorique met en avant une approche plurielle des liens qui unissent la musique et la parole. Il fait partie intégrante de ce travail de thèse puisque, outre le fait d'apporter un éclairage théorique, il permet de justifier notre approche empirique qui dépasse le champ disciplinaire de la didactique des langues uniquement. Il a nécessité, d'une part, une formation importante en phonétique-phonologie, disciplines dans lesquelles nous avons puisé les méthodologies de recueil et d'analyse des données puisque la didactique des langues ne propose pas d'outils empiriques propres. Cette formation a également été indispensable étant donné que la formation en master FLES que j'ai suivie ne m'a pas permis de me former dans les disciplines de référence telles que la linguistique fondamentale. Ce cadre théorique a également nécessité le recours à un cadre nouveau (toujours en co-construction) qui permet de prendre en compte les processus ascendants et descendants qui impactent les processus langagiers. Le caractère pluri-inter-transdisciplinaire m'a permis également en tant que chercheur de me positionner et de justifier ce travail de thèse face à des domaines parfois cloisonnés (linguistique fondamentale – linguistique appliquée – didactique des langues) pourtant fortement interdépendants : de la description à l'intervention. Le but majeur de cet éclairage théorique était de questionner les liens entre musique et parole à plusieurs niveaux afin de montrer la complexité de ces liens que ce soit en diachronie et en synchronie.

Ce cadrage théorique a été également l'occasion de porter une réflexion sur la place et les objectifs possibles du travail de médiation de la prononciation dans une approche actuelle telle que l'approche par tâches dans laquelle nous nous positionnons. Il a permis également de reprendre des terminologies symbolistes comme le « crible phonologique » au regard de notre approche à la fois émergentiste et socioconstructiviste.

Pour terminer, nous avons pu montrer que la prise en compte du caractère complexe des processus du vivant nécessite la prise en compte de prismes disciplinaires et paradigmes expérimentaux pluriels afin d'appréhender au mieux l'ensemble des pressions qui s'exerce sur les processus questionnés.



## **Deuxième partie : les liens musique-parole à l'épreuve de l'expérimentation**

Cette deuxième partie « *les liens musique-parole à l'épreuve de l'expérimentation* » aura pour de questionner le cadre théorique dressé dans la partie 1 de ce travail doctoral à travers trois études expérimentales : une enquête psychosociale, une expérimentale psychoacoustique et une étude en didactique interventionniste et quasi-expérimentale. Nous les traiterons de manière indépendante afin de mettre en avant leurs méthodologies, recueil et analyse des données ainsi que leurs conclusions importantes. Une discussion générale permettra de mettre en relation ces trois études dans la perspective de la problématique générale de ce travail de thèse. Nous montrerons, en outre, les perspectives globales, les apports au(x) domaine(s), les limites et enfin les perspectives didactique de cette thèse.

### **4. Etude 1 : enquête psychosociale**

#### **Résumé**

Cette étude psychosociale a pour but d'amener une première réflexion dans les liens entre musique et parole et ce dans une perspective de didactique des langues. Elle vise, à travers plusieurs questionnaires, à montrer des tendances fortes parmi les représentations de divers panels sélectionnés en rapport à notre problématique. Nous tenterons principalement de voir si la relation au son chez les musiciens peut apporter une valeur acquisitionnelle dans la médiation de la perception et de la production des sons en langue étrangère chez des locuteurs plurilingues et ce sur un plan intra- et interpsychique (émotions, interactions, cognition, etc.). Pour ce faire, nous avons mis en place trois questionnaires à destination de trois panels différents : les musiciens, les locuteurs-plurilingues et les locuteurs-signeurs. Les musiciens ont été divisés en deux groupes en fonction qu'ils aient fait du solfège ou non (respectivement MAS et MSS). Ils étaient questionnés seulement sur leur rapport à la musique. Le panel de locuteurs-plurilingues a été réparti en deux groupes en fonction du fait qu'ils pratiquent ou non au moins une langue étrangère en contexte naturel (respectivement LPR et LPNR). Ils étaient questionnés seulement sur leur rapport au langage verbal. Pour terminer, un groupe de locuteurs-signeurs (LS) qui n'ont pas de trouble du langage et qui pratiquent à la fois une langue des signes et une langue orale a répondu à cette enquête.

Au total, nous avons obtenu 1293 répondants sur l'ensemble des cinq panels. Etant donné le profil socio-culturel des répondants, les résultats proposés ont été pris dans le cadre d'individus

plutôt jeunes adultes, avec des représentations occidentales et un haut niveau métacognitif. Les musiciens ayant fait du solfège pratiquent une plus grande variété d'instruments, depuis plus longtemps et depuis un âge plus précoce. Les musiciens n'ayant pas fait de solfège ont principalement appris la musique de manière autonome et pratiquent dans des groupes amateurs. Les locuteurs-plurilingues parlent français dans leur famille et/ou dans la vie de tous les jours. Ils ont appris, de manière générale, leur première langue étrangère entre 0 et 15 ans même si les LPNR semblent avoir commencé légèrement plus tardivement que les LPR. Les langues étrangères les plus apprises sont l'anglais, l'espagnol, l'allemand et l'italien. Les locuteurs-signeurs de notre panel sont majoritairement des signeurs tardifs et sont entrés dans le langage par l'oral. Ils ont en majorité suivi des cours formels de langue des signes et l'utilisent dans un contexte professionnel et/ou scolaire et/ou amical. Ils ont appris leur première langue étrangère dans la même tranche d'âge que les panels de locuteurs-plurilingues. Concernant le rapport entre musiciens et langage musical, nous avons pu montrer peu de différences dans les tendances des répondants en fonction de la variable étudiée, soit la pratique ou non du solfège. Ils montrent néanmoins des profils musicaux différents (début, type de pratique, instruments joués, etc.). Nous avons constaté, d'après les résultats obtenus, que pour eux, un « bon » musicien n'est pas seulement celui qui reproduit les intentions du compositeur mais surtout un individu qui apporte sa personnalité et sa vision de l'œuvre dans son jeu : l'interprète. Changer de musique, selon eux, c'est changer d'état d'esprit mais pas de personnalité. Jouer une partition/tablatrice semble imposer aux musiciens une prise de conscience des différentes composantes de l'œuvre écrite contrairement à lorsqu'ils doivent simplement « jouer de la musique » sans œuvre écrite de référence. L'apprentissage du solfège augmenterait cette prise de conscience métacognitive. Cette dernière semble limiter légèrement les musiciens avec solfège dans leur capacité à improviser. Dans les deux cas, il y a une très bonne prise de conscience sur les différences acoustiques et perceptives entre ce que le musicien joue et ce que les spectateurs perçoivent. Cependant, ils déclarent minoritairement le fait qu'avoir un retour positif des spectateurs est important dans ce qu'est « bien » jouer.

A titre de comparaison transdisciplinaire, nous n'avons également pas pu montrer de fortes différences dans les tendances en fonction de la pratique ou non en contexte naturel d'au moins une langue étrangère. Pour eux, pour « bien » parler une langue étrangère, il est plutôt important de parler comme un natif mais également de transmettre des émotions. De ce fait, nous avons pu constater que lorsque les locuteurs-plurilingues changent de langue, ils n'ont pas l'impression de changer de personnalité mais pensent être la même personne tout en

agissant différemment avec les autres. Ce qui importe c'est de garder leur identité personnelle mais aussi d'intégrer les éléments culturels de la LE. Cette ambiguïté se retrouve lorsque les deux panels montrent une tendance modérée contre le fait qu'ils pensent aux différentes caractéristiques de la langue lorsqu'ils parlent leur(s) langue(s) première(s). Nous remarquerons que cette tendance est plus forte chez les LPNR. À l'inverse, les deux panels sont plutôt favorables envers le fait de penser aux caractéristiques de la langue lorsqu'ils parlent une langue étrangère. Concernant les formes sonore et écrite du discours, il ne serait pas plus difficile de lire un texte que de parler spontanément en langue étrangère. Le plus important pour eux, pour « bien » parler, est d'être à l'aise dans la production en facilitant la compréhension, même s'il y a des erreurs. La qualité de la forme n'était pas leur préoccupation principale.

Nous avons constaté que la pratique d'au moins une langue des signes ne modifie pas les représentations des individus sur leur rapport au langage oral. Ce point peut être expliqué par le fait que la majorité des sondés sont des locuteurs-signeurs tardifs. Néanmoins, nous avons pu mettre en évidence que la modalité physique du signe rend la forme écrite plus éloignée de la première. Contrairement aux individus ne pratiquant pas de LE en contexte naturel qui pensent qu'il est plus facile de lire un texte en LE que de parler spontanément, les locuteurs-signeurs pensent qu'il est plus difficile de signer un texte que de signer spontanément. La modalité signée imposerait donc chez des apprenants tardifs des contraintes spécifiques à la modalité physique du signe.

Pour conclure, selon tous ces éléments et en comparant les panels de musiciens et de locuteurs-plurilingues, nous pouvons dire que les rapports individu-son-écrit entre la parole et les locuteurs-plurilingues sont largement influencés par le caractère pragmatique des « langues ». Les musiciens ont un rapport plus introspectif en lien avec l'intentionnalité (explicite) émotionnelle que le jeu musical implique. Cet aspect permet d'entrevoir des applications en didactique des langues notamment pour une plus grande prise de conscience des individus plurilingues dans leur performance verbale. L'importance peut être d'amener les apprenants à se focaliser non plus sur la validation de l'interaction mais sur une prise de recul sur leur manière de prononcer, et de l'impact de différentes variables sur ces prononciations.

## 4.1 Méthodologie et questions de recherche

Malgré l'existence physique (Jones, 1931) indéniable des sons dans la remédiation de l'accent dans l'apprentissage d'une langue étrangère, ces derniers font également intervenir d'autres réalités (cf. section 1,3). Elles peuvent correspondre aux phonèmes comme réalité psychologique tels que défini par Randall (2007: 39) « *c'est un concept abstrait, une construction mentale, un symbole qui existe dans l'esprit des locuteurs. Le défi de la psychologie et de la linguistique est d'apporter des explications sur le transfert de données brutes physiques vers des représentations abstraites symboliques de la parole* » ; mais également à des rapports graphophonématiques dont B. Lauret (2007 : 78-82) propose une étude condensée pour le français ; ou bien encore à standard une dimension individuelle en faisant référence « *au plus profond du centre de la personnalité de l'apprenant* » (Krashen, 1988 : 35). Cette approche psychologique est renforcée lorsque Krashen (1988 : 35) affirme que « *la prononciation est beaucoup plus ancrée dans la personnalité de l'apprenant que n'importe quel autre aspect de la langue* ». Dans cette perspective, les acteurs de la médiation de la perception et la production des sons en langue étrangère chez l'adulte seront amenés à prendre conscience que certaines pratiques mettent en place « *une régression mentale et psychologique, voire une infantilisation (Kris, 1952, cité par D. Singleton, 1989) accompagnée de la mise en place d'une nouvelle identité : ce défi à l'intégrité identité de base pose moins de problèmes à l'enfant* » (Rolland, 2011 : 102). Nous tenterons, à travers cette étude, de voir si les représentations sur les pratiques musicales entre autres, peuvent nous aider dans cette prise de conscience de l'impact émotionnel de la construction d'une nouvelle identité sonore vocale.

Notre étude quantitative ne s'intéressera pas aux représentations sur la « *forme* » des éléments linguistique (phonèmes, sons, etc.) mais aux représentations des individus sur les pratiques qui visent la mise en son d'un langage dans une perspective inter- et intrapsychique dont la dichotomie a été développée par Vygotski (1934) (cf. section 2.3.1). Nous irons plus loin en nous intéressant aux représentations des liens forme écrite/sonore dont les travaux de Dehaene (2007, entre autres) sont les plus marquants en France. Nous considérerons la notion de représentation comme quelque chose qui « *nous guide dans la façon de nommer et définir ensemble les différents aspects de notre réalité de tous les jours, dans la façon de les interpréter, statuer sur eux et, le cas échéant, prendre une position à leur égard et la défendre* » (Jodelet, 2003 : 47). Nous tenterons d'apporter un éclairage multimodal en comparant ces représentations sur plusieurs types de langages : langage verbal oral, des signes et musical. Deux variables

supplémentaires seront présentes : chez les musiciens, l'apprentissage de la musique par le solfège (ou non) et chez les locuteurs, la pratique régulière d'au moins une langue étrangère en contexte naturel. L'ensemble de ces données nous permettra d'apporter des réponses aux questionnements suivants : (i) le type de langage verbal ou musical favorise-t-il l'émergence de représentations différentes sur le traitement de l'objet sonore et sa forme écrite ? (ii) des pratiques différentes (solfège, pratique régulière) favorisent-elles le renforcement de certaines représentations sur les formes écrite et sonore ? (iii) la modalité de langage (langue des signes) influe-t-elle sur les représentations des locuteurs ?

L'ensemble de ce travail visera à porter un regard pluri-inter-transdisciplinaire (Claverie, 2010 ; cf. section 2.1.1) à visée didactique afin de mesurer l'influence des propriétés intrinsèques et extrinsèques d'un langage donné sur les représentations de ceux qui lui donnent corps. Nous tenterons, en outre, de voir si la relation au son chez les musiciens peut apporter une valeur acquisitionnelle dans la médiation de la perception et de la production des sons en langue étrangère chez des locuteurs plurilingues et ce sur un plan intra- et interpsychique (émotions, interactions, cognition, etc.). Nous finirons en proposant des pistes didactiques sur l'enseignement/apprentissage de la prononciation en contexte FLE.

#### 4.1.1 Matériel et méthodes

Dans cette sous-section, nous présenterons le déroulement général de l'enquête depuis sa mise en place jusqu'au recueil et à l'analyse des données.

##### 4.1.1.1 Enquête

L'enquête a été créée sur la base d'un formulaire GoogleDocument©. Elle comporte cinq questionnaires différents (Tableau 7) :

Questionnaires	Panel de sujets concernés		Spécificité(s) des enquêtés	Abréviation
	1	Musicien	Sans solfège	MSS
	2		Avec solfège	MAS
	3	Locuteur-plurilingue	Pratique régulière LE	LPR
	4		Pratique non-régulière LE	LPNR
5	Locuteur-signeur	LSF et langue orale	LS	

**Tableau 7:** Caractéristiques des 5 questionnaires de l'enquête quantitative selon les sujets auxquels ils s'adressent.

Chaque questionnaire (disponible en intégralité en annexes 9, 10 et 11) se décompose en 4 sous parties (les différents concepts seront définis dans les parties spécifiques à chaque panel) :

– **Les informations générales**

Différentes informations sont recueillies: l'âge, la nationalité, le genre, le pays actuel où les enquêtés vivent, le(s) pays d'études, le niveau d'étude, leur situation professionnelle actuelle.

– **Les informations spécifiques :**

– Musiciens :

Différentes informations sont recueillies : les instruments pratiqués, l'instrument le plus pratiqué et depuis combien de temps, le type de formation en musique, l'âge de début de la pratique musicale, les différentes expériences de pratique musicale, la prise ou non de cours de solfège.

– Locuteurs-plurilingues :

Différentes informations sont recueillies: la/les langue(s) de la famille, la/les langue(s) de tous les jours, la/les langue(s) étrangère(s) apprise(s), l'âge de la première langue étrangère apprise, la pratique ou non de la langue étrangère (LE) en contexte naturel.

– Locuteurs-signeurs :

Différentes informations sont recueillies: la/les langue(s) de la famille, la/les langue(s) de tous les jours, la/les langue(s) étrangère(s) apprise(s), l'âge de la première langue étrangère, la pratique de la LE en contexte naturel ou non, l'âge du premier contact avec langue orale/langue des signes, le contexte de pratique de la langue des signes, la prise ou non de cours formels de langue des signes.

– **Le questionnaire à échelle d'attitudes**

Même si le questionnaire à échelle d'attitudes n'est pas « *l'expression d'un déterminant réel de la conduite sociale, [nous prenons une position psycho-sociologique et par conséquent, nous cherchons] des opinions fortes et personnelles, celles qui sont à chaque instant capables de se traduire en actes, en réactions, en sentiments puissants, socialement significatifs* » (Mucchielli, 1967 : 27). Nous avons choisi le système d'auto-notation défini par Mucchielli (1967: 27-28) ou échelle de Likert (Legrain, 2007: 36). Nous avons pour cela proposé treize affirmations évaluables sur une échelle d'attitude à cinq gradients : « tout à fait d'accord », « plutôt d'accord' », « moyennement d'accord », « plutôt pas d'accord » et « pas du tout d'accord ».

– **Questions à choix multiples**

Nous avons également mis en place cinq questions à choix multiples (voir annexes) acceptant une ou plusieurs réponses afin d'apporter des précisions sur certains points car la question à choix multiple « *est située entre la question ouverte et la question fermée. Elle offre le choix entre plusieurs réponses et introduit ainsi une nuance dans le commentaire* » (Legrain, 2007 : 33).

#### **4.1.1.2 Recueil des données**

Les données ont été récupérées à travers l'envoi d'un courriel (annexe 12) diffusés à de nombreux organismes et établissements dont la liste étudiante de l'université Paris 3 Sorbonne-Nouvelle, l'université Stendhal-Grenoble 3, le Conservatoire National Supérieur de Musique et de Danse de Paris (CNSMDP), les étudiants en Master d'interprétariat de LSF des université de Lumière-Lyon2 et Charles-de-Gaulle Lille 3 ainsi qu'à des listes d'orchestres amateurs et des annonces ont été postées sur des sites de musiciens amateurs : [www.zinkinf.com](http://www.zinkinf.com) et [www.easyzik.com](http://www.easyzik.com). La présente liste n'est pas exhaustive.

#### **4.1.1.3 Analyse des données**

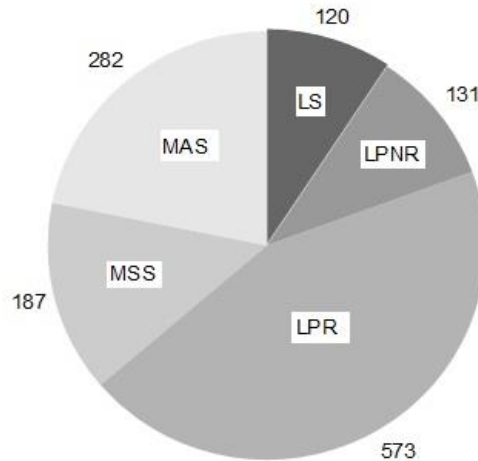
Les données ont été collectées automatiquement par Googledocument© dans un document tableur. Les analyses ont été faites soit à la main en calculant chaque occurrence soit en utilisant la formule « *NB.SI* » qui permet de calculer automatiquement le nombre d'occurrences pour un label donné. Chaque calcul a été contrôlé en comparant le nombre total de réponses et le nombre d'occurrences mesurées. Toutes les données sont en corrélation à 100%. Dans les tableaux proposés, les abréviations NSP et NT correspondent respectivement aux non-réponses et aux réponses non-traitables. L'ensemble des tableaux d'analyse sont disponibles en annexes de 13 à 25.

#### **4.1.2 Profil socio-culturel des sondés**

Cette sous-section présentera le profil socio-culturel général des sondés mais également plus précisément celui de chaque panel : les musiciens, les locuteurs-plurilingues et les locuteurs-signeurs.

#### 4.1.2.1 Profil général des sondés

L'ensemble de l'enquête regroupe 1293 répondants répartis dans les proportions du diagramme ci-dessous (Figure 34).



**Figure 34:** Répartition quantitative des cinq panels de l'enquête. Les sigles signifient respectivement musiciens sans cours de solfège (MSS), musiciens avec cours de solfège (MAS), locuteurs-plurilingues avec pratique régulière en contexte naturel (LPR), locuteurs-plurilingues sans pratique régulière en contexte naturel (LPNR) et locuteurs-signeurs (LS).

Il a été précisé que chaque répondant devait choisir un seul questionnaire parmi les cinq en fonction de celui qui correspondait le plus à son parcours. De plus, les musiciens ont été interrogés sur leur rapport au langage musical et les locuteurs-plurilingues au langage verbal uniquement. Nous avons ainsi limité le biais issu du fait que certains musiciens puissent être plurilingues et certains locuteurs-plurilingues, des musiciens. Aucun moyen technique ne permet la vérification de cette donnée. Les données seront présentées au dixième près afin de pouvoir représenter certains pourcentages inférieurs à 1%.

Sexe	Femme	Homme	NSP				
	<b>63%</b>	34%	3%				
Age	moins de 15	15 à 25	26 à 35	36 à 45	46 à 55	56 et +	NSP
	0,1%	<b>60,3%</b>	22,4%	7,7%	5,7%	3,6%	0,2%
Niveau Étude	Primaire	Collège	Lycée	Ens. Sup.	NSP		
	0,1%	1,1%	21,5%	<b>77,2%</b>	0,2%		
Situation professionnelle	Etudiant	Ens/Cherch/ing	Spécialiste	Autre	Sans activité	NT	
	<b>55,7%</b>	12,1%	4,5%	19,0%	8,2%	0,5%	

**Tableau 8:** Profil socioculturel et professionnel des sondés ( $n = 1293$ ). Les résultats remarquables sont mis en évidence par un fond foncé et en gras.



	Nationalité	Pays actuel de résidence	Pays principal des études
<b>Europe</b>	<b>90,3%</b>	<b>95,9%</b>	<b>92,0%</b>
<b>Amérique Nord</b>	1,5%	1,8%	2,0%
<b>Amérique Centrale</b>	0,6%	0,2%	0,7%
<b>Amérique Sud</b>	0,9%	0,2%	0,8%
<b>Afrique</b>	2,2%	1,0%	2,4%
<b>Asie</b>	1,1%	0,5%	0,8%
<b>Moyen-Orient</b>	1,1%	0,3%	0,8%
<b>Océanie</b>	0,1%	0,1%	0,5%
<b>NT</b>	0,1%	0,1%	0,2%
<b>Bi-Nationalité</b>	2,0%	-	-

**Tableau 9:** Profil culturel des sondés. ( $n = 1293$ ). Les résultats remarquables sont mis en évidence par un fond foncé et en gras. NT signifie non-traitable.

Nous présenterons ci-dessous les grandes tendances du profil socio-culturel des sondés en reprenant les données présentées dans le Tableau 8 et le Tableau 9 :

- **sexe** : une majorité de femmes compose le panel ; néanmoins, la littérature scientifique ne démontre pas, de manière formelle, que le genre influe de manière importante sur le traitement cognitif.

Eclairage théorique : il semblerait, selon de récentes découvertes, que les caractéristiques individuelles seraient plus importantes que celles du genre à ce niveau de traitement (Crepault, 1997 ; Daumas, 2005 ; Vidal, 2005).

3% des personnes n'ont pas répondu à la question sur leur genre en raison d'un problème technique sur le questionnaire. Cependant, puisque la variable du genre n'est ici pas pertinente, nous avons décidé de conserver ces réponses.

- **Age** : une majorité de personnes est âgée entre 15 et 25 ans.

Eclairage théorique : cette donnée est importante car comme le rappelle Rolland, (2011: 100) « les adultes ayant des capacités cognitives plus développées auront une supériorité sur ce qui touche à la compréhension, la conceptualisation, car ils s'appuient sur des capacités d'apprentissage inductif (allant de l'usage à la règle) ». Nous supposons donc que notre panel, bien que majoritairement composé d'adolescents ou jeunes adultes, sera apte à répondre à des questions d'ordre méta-réflexif sur leurs pratiques liées aux langages étudiés.

- **Métacognition** : la plupart du panel a fait/ fait encore des études supérieures ce qui implique que les résultats proposés seront valables dans le cadre d'individus ayant un niveau métacognitif important étant donné le recul imposé par des études de type universitaire.

Eclairage théorique : cette tendance a été montrée, en outre, par Doudin & Martin (1992 : 19) : « *les performances scolaires élevées sont associées à des compétences métacognitives efficaces [...] la maîtrise de savoirs et de savoir-faire métacognitifs permet de bénéficier de l'instruction et facilite le développement cognitif et les apprentissages notionnels* ». Ce point renforce le précédent vers des capacités méta-réflexive et métacognitive de notre panel.

- **Géographie** : l'ensemble des résultats devra être pris dans une perspective majoritairement occidentale avec environ 90% de personnes ayant une nationalité occidentale et/ou vivant dans un pays occidental et/ou ayant fait ses études dans un pays d'occident. Ce biais implique de prendre du recul sur de possibles généralisations qui ne pourront être faites avec le monde oriental où d'autres systèmes musicaux et langagiers peuvent modifier de manière radicale les représentations.

Eclairage théorique : cette influence est notamment mise en valeur pour la parole dans l'hypothèse Sapir-Whorf (HSW) : « *le fait est que la réalité est, dans une grande mesure, inconsciemment construite à partir des habitudes langagières du groupe. Deux langues ne sont jamais suffisamment semblables pour être considérées comme représentant la même réalité sociale. Les mondes où vivent des sociétés différentes sont des mondes distincts, pas simplement le même monde avec d'autres étiquettes* »<sup>66</sup> (traduit par Détrie, Siblot, Vérine 2001 : 138 ; d'après Mandelbaum 1949 : 162). Nous supposons donc que notre panel montrera des tendances qui ne seront peut-être pas généralisable à d'autres cultures, modes de pensées, rapports aux langages étudiés.

En conclusion, nous pouvons penser que le grand nombre de réponses permet une représentativité suffisante pour nos conclusions même si nous devons prendre en considération les différences d'effectif entre les sous-groupes. Nous noterons que la marge d'erreur pour une population mère très grande (considérée comme infinie), pour un taux de confiance de 95% et une proportion des éléments de la population mère considérée comme inconnue ( $p=0.05$ ), est comprise entre 4.09% et 8.94% en fonction de la taille d'échantillon des panels de notre étude (Tableau 10).

---

<sup>66</sup> « *the fact of the matter is that the "real world" is to a large extent unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with the different labels attached* ».

Panel	n	Marge d'erreur
LPR	573	4,09%
LPNR	131	8,56%
LS	120	8,95%
MAS	282	5,84%
MSS	187	7,17%

**Tableau 10** : Marge d'erreur calculée en fonction de la taille d'échantillon du panel. La formule utilisée est :  $e = \frac{t\sqrt{p(1-p)}}{n}$ . La population mère est considérée comme infinie car très grande.  $t$  correspond au coefficient de marge, calculé à partir d'un taux de confiance de 95% il est égal à 1,96.  $p$  correspond à la proportion des éléments de la population-mère et qui est égal à 0.05 car considéré comme inconnu.  $n$  correspond à la taille de l'échantillon<sup>67</sup>.

Cependant, les résultats proposés devront être pris dans le cadre d'un profil socio-culturel spécifique d'individus plutôt jeunes adultes, avec des représentations occidentales et un haut niveau métacognitif. Nous ne reprendrons pas dans les analyses des trois profils (musiciens, locuteurs-plurilingues, locuteurs-signeurs) les données spécifiques socio-culturelles pour chacun de ces groupes.

#### 4.1.2.2 Profil des musiciens

L'enquête regroupe 282 musiciens ayant fait du solfège (MAS) et 187 n'ayant pas fait de solfège (MSS). Les profils ont été dissociés à la réponse à la question : « *avez-vous déjà suivi des cours de solfège régulièrement ?* ». Le solfège est ici défini comme un enseignement de la musique dont la valeur « *en Occident, est issue de la volonté particulièrement sensible au siècle des Lumières, de rationaliser les pratiques musicales* » (Tripier-Mondancin, 2008 : 19). Il est « *considéré comme un outil pédagogique pour maîtriser la pratique de la musique occidentale savante* » (*ibidem*: 20). Ainsi, le solfège se cristalliserait autour d'une culture et de pratiques définies : rythme, déchiffrage, etc. favorisant ou non une approche métacognitive ou intuitive de la musique.

Le Tableau 11 et le Tableau 12 présentent l'ensemble des données correspondant au profil des musiciens sondés. Les résultats seront commentés ci-après.

<sup>67</sup> Formule proposée sur <http://icp.ge.ch/sem/cms-spip/spip.php?article1641> (consulté le 10/09/14)

Instruments pratiqués	TT	MAS	MSS
Les bois	8,0%	<b>11,0%</b>	3,6%
Les cordes frottées	8,5%	<b>11,2%</b>	4,4%
Cuivres	3,9%	5,0%	2,4%
Percussions	<b>11,1%</b>	<b>10,2%</b>	<b>12,4%</b>
Voix	<b>17,1%</b>	<b>15,8%</b>	<b>18,9%</b>
Clavier	<b>20,6%</b>	<b>25,6%</b>	<b>13,3%</b>
Cordes pincées	<b>29,5%</b>	<b>20,6%</b>	<b>42,6%</b>
Direction	0,1%	0,2%	0,0%
Ordinateur	0,6%	0,4%	0,9%
Harmonica	0,5%	0,0%	1,2%
Didgeridoo	0,1%	0,0%	0,3%
n (effectif)	469	282	187

**Tableau 11:** Instruments joués régulièrement par les musiciens. Les résultats sont donnés pour chaque panel (MSS et MAS) et le total des deux groupes (TT). Les résultats marquants sont mis en évidence par un fond foncé et en gras.

Durée de pratique (en années)	TT	0 à 5	6 à 10	11 à 20	21 à 30	31 et plus	NSP/NE
		19%	<b>26%</b>	<b>35%</b>	11%	7%	2%
	MAS	13%	<b>21%</b>	<b>44%</b>	11%	9%	2%
	MSS	<b>29%</b>	<b>34%</b>	<b>22%</b>	10%	4%	1%
Age de début de pratique (en années)	TT	0 à 5	5 à 10	10 à 15	15 et plus	NE	
		11,1%	<b>42,2%</b>	22,2%	24,1%	0,4%	
	MAS	<b>14,9%</b>	<b>58,2%</b>	<b>16,0%</b>	10,3%	0,7%	
	MSS	5,3%	<b>18,2%</b>	<b>31,6%</b>	<b>44,9%</b>	0,0%	
Type de formation	TT	Apprentissage en autonomie	Cours particuliers	Conservatoires municip., nat. et équival.	Conservatoires supérieurs	Autre	
		36,9%	24,1%	31,9%	6,3%	0,8%	
	MAS	12%	<b>26%</b>	<b>51%</b>	11%	1%	
	MSS	<b>74,0%</b>	<b>21,4%</b>	4,2%	0,0%	0,5%	
Type de pratique	TT	Orchestre	Formation (semi) pro	groupe amateur	Solo	Autre	Aucun
		29,6%	21,9%	<b>45,6%</b>	1,0%	0,5%	1,5%
	MAS	<b>36,2%</b>	<b>24,3%</b>	<b>37,1%</b>	0,9%	0,6%	0,9%
	MSS	15,7%	16,9%	<b>63,2%</b>	1,1%	0,4%	2,7%

**Tableau 12:** Profil spécifique des musiciens. Les résultats sont donnés pour chaque panel (MSS et MAS) et le total des deux groupes (TT). Les résultats marquants sont mis en évidence par un fond foncé et en gras.

- **Instruments pratiqués** : la grande majorité des sondés pratiquent régulièrement soit des cordes pincées (29,5%), des claviers (20,6%), des voix (17,1%) ou des percussions (11,1%). Nous remarquons ce profil pour l'ensemble des musiciens MAS et MSS ; cependant, nous remarquons également une forte représentation chez les musiciens ayant fait du solfège d'autres familles d'instruments tel que les bois (11%) et les cordes frottées (11,2%).

- **Durée de pratique instrumentale** : la plupart des musiciens ont une pratique d'une durée de 6 à 20 ans (61%). Cependant, les musiciens n'ayant pas fait de solfège (MSS) montrent également une forte représentation entre 0 et 5 ans de pratique (29%).
- **Age de début de pratique musicale** : l'ensemble des musiciens a majoritairement commencé entre l'âge de 5 et 10 ans (42,2%). Les MAS ont principalement commencé avant l'âge de 15 ans (89,1%) avec une importante part entre 5 et 10 ans (58,2%). Les MSS ont, quant à eux, principalement commencé entre l'âge de 5 ans et plus (94,7%) dont 44,9% après 15 ans.
- **Type de formation musicale** : la majorité des musiciens a suivi des cours particuliers (24,1%) ou dans des conservatoires municipaux, nationaux ou équivalents (31,9%) ou par un apprentissage en autonomie (36,9%). Les MAS ont principalement suivi des cours particuliers et/ou dans des conservatoires municipaux, nationaux ou équivalents (77%) tandis que les MSS ont principalement appris en autonomie et/ou avec des cours particuliers (95,4%)
- **Type de pratique musicale** : la majorité des musiciens a joué dans des groupes amateurs (45,6%), dans des orchestres (29,6%) et/ou dans des formations semi-professionnelles (21,9%). Les MAS correspondent à cette répartition tandis que les MSS jouent surtout dans des groupes amateurs (63,2%)

**Conclusion** : nous pouvons donc remarquer que les musiciens ayant fait du solfège pratiquent une plus grande variété d'instruments, depuis plus longtemps et depuis un âge plus précoce. Les musiciens n'ayant pas fait de solfège ont principalement appris la musique de manière autonome et pratiquent dans des groupes amateurs.

#### 4.1.2.3 Profil des locuteurs-plurilingues

L'enquête regroupe 573 locuteurs-plurilingues ayant une pratique régulière en contexte naturel (LPR) et 131 n'ayant pas une pratique régulièrement en contexte naturel (LPNR). Les profils ont été dissociés à la question : « *Pratiquez-vous très régulièrement au moins une de ces langues étrangères dans un contexte naturel?* ». La notion de « *contexte naturel* » ayant été développé par « *parler avec des locuteurs natifs (réel ou virtuel), voyage/séjour dans un des pays, lire des livres, regarder la TV* ». Cette notion a été reprise par Tyne (2012) d'après les travaux notamment de Noyau (1988) et Regan (2002) par exemple.

Le Tableau 13 et le Tableau 14 présentent l'ensemble des données correspondant au profil des locuteurs-plurilingues sondés. Les résultats seront commentés ci-après.

	Langues de la famille			Langues de tous les jours			LE apprises		
	TT	LMR	LMNR	TT	LMR	LMNR	TT	LMR	LMNR
<b>Anglais</b>	4,3%	5,2%	0,0%	8,2%	9,8%	1,5%	<b>93,9%</b>	<b>93,2%</b>	<b>96,9%</b>
<b>Espagnol</b>	5,5%	6,6%	0,8%	2,6%	3,1%	0,0%	<b>60,5%</b>	<b>58,6%</b>	<b>68,7%</b>
<b>Allemand</b>	1,7%	1,7%	1,5%	1,3%	1,6%	0,0%	<b>32,0%</b>	<b>33,7%</b>	<b>24,4%</b>
<b>Italien</b>	2,3%	2,6%	0,8%	1,4%	1,7%	0,0%	<b>25,1%</b>	<b>25,7%</b>	<b>22,9%</b>
<b>Portugais</b>	2,8%	2,8%	3,1%	0,1%	0,0%	0,8%	8,1%	8,9%	4,6%
<b>Russe</b>	1,4%	1,6%	0,8%	0,1%	0,2%	0,0%	7,5%	8,4%	3,8%
<b>Chinois</b>	1,8%	2,1%	0,8%	0,7%	0,7%	0,8%	6,4%	7,2%	3,1%
<b>Japonais</b>	0,7%	0,9%	0,0%	0,4%	0,5%	0,0%	8,5%	9,8%	3,1%
<b>Français</b>	<b>77,0%</b>	<b>74,7%</b>	<b>87,0%</b>	<b>85,5%</b>	<b>82,9%</b>	<b>96,9%</b>	<b>17,8%</b>	<b>21,1%</b>	3,1%
<b>Nb langues</b>	X	54	17	X	23	7	X	72	26
<b>Nb réponses</b>	796	659	137	742	608	134	2102	1775	327

**Tableau 13:** Profil spécifique des locuteurs-plurilingues. Pourcentages calculés sur un total de 704 répondants pour le total (TT), 573 locuteurs-plurilingues ayant une pratique régulièrement en contexte naturel (LPR) et 131 n'ayant pas une pratique régulièrement en contexte naturel (LPNR). Le pourcentage correspond au pourcentage de sondés pratiquant cette langue. Les résultats marquants sont mis en évidence par un fond foncé.

Age d'apprentissage de la première langue étrangère apprise (en années)	TT	0 à 5	5 à 10	10 à 15	15 et plus
		<b>15,5%</b>	<b>36,5%</b>	<b>45,0%</b>	3,0%
LMR	<b>18,2%</b>	<b>36,8%</b>	<b>42,1%</b>	3,0%	
LMNR	3,8%	<b>35,1%</b>	<b>58,0%</b>	3,1%	

**Tableau 14:** Âge d'apprentissage de la première langue étrangère apprise. Pourcentages calculés sur un total de 704 répondants pour le total (TT), 573 locuteurs-plurilingues ayant une pratique régulièrement en contexte naturel (LPR) et 131 n'ayant pas une pratique régulièrement en contexte naturel (LPNR). Les résultats marquants sont mis en évidence par un fond foncé.

- **Age d'apprentissage de la première langue étrangère:** la plupart des sondés ont appris leur première langue étrangère entre l'âge de 0 et 15 ans (97%). Les locuteurs-plurilingues ayant une pratique régulière en contexte naturel ont commencé, de manière globale, plus précocement avec 18,2% de sondés ayant appris leur première LE entre 0 et 5 ans contre 3,8% pour les LPNR.
- **Langue(s) de la famille :** la plupart des sondés (77%) parlent le français avec leur famille. Le nombre de langues mentionnées pour les LPR (54) est plus important que celui des LPNR (17).
- **Langue(s) de tous les jours :** la plupart des sondés (85,5%) parlent le français dans la vie de tous les jours. Le nombre de langues mentionnées pour les LPR (23) est plus important que celui des LPNR (7).
- **Langues étrangères apprises :** les langues étrangères les plus apprises correspondent dans les deux groupes : anglais (93,9%), espagnol (60,5%), allemand (32%), italien (25,1%). Le français est également représenté dans le groupe LPR avec 21,1% des sondés qui l'ont appris comme langue étrangère.

**Conclusion** : nous pouvons donc remarquer que les locuteurs-plurilingues pratiquant ou non une langue étrangère régulièrement dans un contexte naturel parlent français dans leur famille et/ou dans la vie de tous les jours. Ils ont appris, de manière générale, leur première langue étrangère entre 0 et 15 ans même si les LPNR semblent avoir commencé légèrement plus tardivement que les LPR. Les langues étrangères les plus apprises sont l'anglais, l'espagnol, l'allemand et l'italien et correspondent aux langues étrangères les plus enseignées dans le cursus scolaire français.

#### 4.1.2.4 Profil des locuteurs-signeurs

L'enquête regroupe 120 locuteurs-signeurs ayant une pratique d'au moins une langue orale et une langue des signes.

Le Tableau 15 présente l'ensemble des données correspondant au profil des locuteurs-signeurs sondés. Les résultats seront commentés ci-après.

Age d'apprentissage de la 1ere LE	0 à 5	5 à 10	10 à 15	15 et plus		
	6%	<b>26%</b>	<b>65%</b>	3%		
Age d'apprentissage de la langue des signes	0 à 5	5 à 10	10 à 15	15 et plus		
	5%	7%	13%	<b>76%</b>		
Age d'apprentissage de la langue orale	0 à 5	5 à 10	10 à 15	15 et plus		
	<b>95%</b>	2%	0%	3%		
Cours formels de langue des signes	Oui	Non				
	<b>89%</b>	11%				
Pratique LE régulière contexte naturel	Oui	Non				
	<b>60%</b>	40%				
Contexte utilisation Langue des signes	Amical	Scolaire	Pro	Familial	NSP	Non-pertinent
	<b>23%</b>	<b>27%</b>	<b>36%</b>	13%	1%	1%

**Tableau 15:** Profil spécifique des locuteurs-signeurs. Pourcentages calculés sur un total de 120 locuteurs-signeurs à l'exception du contexte de pratique de la langue des signes dont le pourcentage est calculé sur la base totale des réponses soit 183 à la question. Les résultats importants sont mis en évidence par un fond foncé.

- **Age d'apprentissage de la 1ere langue étrangère** : la majorité des sondés ont appris leur première langue étrangère entre l'âge de 5 et 15 ans (91%).
- **Age d'apprentissage de la 1ere langue des signes** : la majorité des sondés ont appris leur première langue des signes après 15 ans (76%).
- **Age d'apprentissage de la 1ere langue orale** : la majorité des sondés ont appris leur première langue orale entre l'âge de 0 et 5 ans (95%).
- **Type d'apprentissage de la langue des signes** : 89% des sondés ont pris des cours formels de langue des signes.
- **Pratique d'au moins une langue étrangère** : 60% des sondés pratiquent au moins une langue étrangère dans un contexte naturel.
- **Contexte d'usage de la langue des signes** : la langue des signes est utilisée par les sondés à 36% dans le milieu professionnel, à 27% dans le milieu scolaire, à 23% dans le milieu amical et à 13% dans le milieu familial.

**Conclusion** : nous pouvons donc remarquer que les locuteurs-signeurs de notre panel sont majoritairement des signeurs tardifs et sont entrés dans le langage par l'oral. Ils ont en majorité suivi des cours formels de langue des signes et l'utilisent dans un contexte professionnel et/ou scolaire et/ou amical. Ils ont appris leur première langue étrangère dans la même tranche d'âge que les panels de locuteurs-plurilingues.

## 4.2 Résultats

Après avoir déterminé le profil socio-culturel des cinq panels repartis en trois groupes : musiciens (MAS, MSS), locuteurs-plurilingues (LPR, LPNR) et locuteurs-signeurs (LS), nous allons traiter les résultats par panel des questionnaires à échelle d'attitudes, d'une part et à choix multiples, d'autre part. Les résultats seront présentés, de manière séparée, pour chaque panel. Nous précisons que les résultats présentés reflètent les représentations qui émergent des réponses données par les sondés et non de pratiques réelles. Les résultats seront arrondis à l'unité en raison du choix limité de réponses qui ne nécessite pas une plus grande précision.

### 4.2.1 Questionnaires à échelle d'attitudes

Le questionnaire à échelle d'attitude des cinq panels se compose de treize phrases affirmatives évaluées sur une échelle à cinq attitudes (voir section 4.1.1.1). Les phrases affirmatives des cinq panels ont été créées de manière à ce qu'une grande majorité d'entre elles correspondent à la thématique (Tableau 16).

Panel	Musiciens (MAS/MSS)	Locuteurs-plurilingues (LPR/LPNR)	Locuteurs-signeurs (LS)
Question	La partition/tablature m'aide à mieux jouer.	La lecture m'aide à mieux prononcer.	La lecture m'aide à mieux signer.

**Tableau 16:** Exemple de correspondances entre phrases affirmatives dans le questionnaire à échelle d'attitude des 5 panels

Cependant, toutes les questions ne sont pas en correspondance afin de préciser dans l'analyse la spécificité de certains points pour chaque domaine. Dans cette partie, les résultats seront donnés en supprimant les réponses correspondant à « moyennement d'accord ». Une tendance sera remarquable pour une forte représentation dans les extrêmes (« tout à fait d'accord » + « plutôt d'accord » ou « pas du tout d'accord » + « pas d'accord ») associée à une faible représentation dans l'extrême contraire ou une forte représentation dans les « moyennement d'accord ».

Nous présenterons les résultats de chaque panel indépendamment : les musiciens, les locuteurs-plurilingues et les locuteurs-signeurs.



### 4.2.1.1 Les Musiciens

Les 13 phrases affirmatives des panels MAS et MSS sont indiquées dans le Tableau 17.

1	Un bon musicien doit seulement reproduire les intentions du compositeur.	7	Ma qualité de jeu est meilleure lorsque j'improvise.
2	Un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète.	8	Ma qualité de jeu est meilleure lorsque je joue une partition.
3	Quand je joue de la musique, je pense aux différentes techniques que nécessite ce type de musique.	9	J'ai l'impression que ce que j'entends quand je joue n'est pas la même chose que ce qu'entendent les spectateurs.
4	Avant de déchiffrer une partition/tablatrice, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	10	Lorsque je joue une partition/tablatrice dans un style, je ressens des différences physiques par rapport à lorsque je joue une partition dans un style différent.
5	Avant de jouer une partition/tablatrice que je connais, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	11	La partition/tablatrice m'aide à mieux jouer.
6	Il est plus difficile de jouer une partition/tablatrice que d'improviser.	12	La timidité influence la qualité du jeu.
13	Quand je joue, je ne pense pas en même temps.		

**Tableau 17:** Les treize phrases affirmatives du questionnaire à échelle d'attitude des musiciens (MAS et MSS).

Nous analyserons les résultats pour chaque sous-panel de musiciens :

- Musiciens avec solfège (MAS)

Les résultats sont présentés sous forme de tableau présentant les phrases affirmatives et le pourcentage de réponses pour les cinq gradients de réponse (Tableau 18).

	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Un bon musicien doit seulement reproduire les intentions du compositeur.	3%	11%	<b>34%</b>	<b>27%</b>	<b>25%</b>
Un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète.	<b>57%</b>	<b>37%</b>	5%	1%	0%
Quand je joue de la musique, je pense aux différentes techniques que nécessite ce type de musique.	20%	<b>35%</b>	<b>22%</b>	17%	6%
Avant de déchiffrer une partition/tablatrice, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	<b>41%</b>	<b>40%</b>	8%	8%	2%
Avant de jouer une partition/tablatrice que je connais, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	<b>32%</b>	<b>35%</b>	17%	11%	5%
Il est plus difficile de jouer une partition/tablatrice que d'improviser.	6%	15%	<b>26%</b>	<b>29%</b>	<b>25%</b>
Ma qualité de jeu est meilleure lorsque j'improviser.	9%	16%	<b>30%</b>	<b>26%</b>	19%
Ma qualité de jeu est meilleure lorsque je joue une partition.	<b>23%</b>	<b>27%</b>	<b>31%</b>	15%	5%
J'ai l'impression que ce que j'entends quand je joue n'est pas la même chose que ce qu'entendent les spectateurs.	<b>30%</b>	<b>35%</b>	13%	17%	4%
Lorsque je joue une partition/tablatrice dans un style, je ressens des différences physiques par rapport à lorsque je joue une partition dans un style différent.	<b>41%</b>	<b>35%</b>	16%	5%	3%
La partition/tablatrice m'aide à mieux jouer.	11%	<b>27%</b>	<b>34%</b>	18%	10%
La timidité influence la qualité du jeu.	<b>39%</b>	<b>37%</b>	12%	7%	4%
Quand je joue, je ne pense pas en même temps.	16%	17%	<b>21%</b>	<b>23%</b>	<b>22%</b>

**Tableau 18:** Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel MAS. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% = ou > X > 20\%$ .

Nous analyserons ces données en cinq catégories thématiques :

**Rôle du musicien :** à travers les réponses données par le panel de musiciens ayant fait du solfège, un « *bon musicien* » est celui qui apporte sa personnalité dans l'œuvre qu'il interprète (94%) mais ne doit pas reproduire les intentions du compositeur (52%).

**Intrapsychique :** en général, les musiciens MAS du panel pensent aux différentes techniques que nécessite un type de musique (55%) et pensent aux différents éléments d'une partition/tablatrice (tempo, rythme, tonalité, etc.) avant de la déchiffrer (81%) mais également avant de jouer une partition/tablatrice qu'ils connaissent (67%). En général, les sondés MAS n'ont plutôt pas l'impression de penser en même temps qu'ils jouent (50%) mais ce résultat n'est pas remarquable puisque 33 % ont un avis contraire.

**Influence de l'écrit** : selon le panel, il n'est pas plus difficile de jouer une partition que d'improviser (54%) et la qualité de jeu n'est pas meilleure lorsqu'ils improvisent (45%). Cependant, la qualité de jeu est perçue par les sondés comme meilleure lorsqu'ils jouent une partition (50%). Toutefois, la partition ne serait pas spécialement une aide afin de mieux jouer. 76% des sondés ressentent une différence physique lorsqu'ils changent de style de musique.

**Interpsychique**: les sondés présentent une forte impression sur la différence entre ce qu'ils entendent de leur jeu musical et ce qu'entendent les spectateurs (76%).

**Émotion** : la timidité serait une variable importante dans la qualité de jeu (76%).

- Musiciens sans solfège (MSS)

Les résultats sont présentés sous forme de tableau présentant les phrases affirmatives et le pourcentage de réponses pour les cinq gradients de réponse (Tableau 19).

	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Un bon musicien doit seulement reproduire les intentions du compositeur.	1%	11%	<b>34%</b>	<b>30%</b>	<b>25%</b>
Un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète.	<b>62%</b>	<b>35%</b>	3%	0%	0%
Quand je joue de la musique, je pense aux différentes techniques que nécessite ce type de musique.	14%	<b>34%</b>	<b>27%</b>	18%	7%
Avant de déchiffrer une partition/tablatrice, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	<b>27%</b>	<b>31%</b>	16%	17%	9%
Avant de jouer une partition/tablatrice que je connais, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	<b>24%</b>	<b>33%</b>	16%	16%	11%
Il est plus difficile de jouer une partition/tablatrice que d'improviser.	12%	18%	<b>30%</b>	<b>22%</b>	17%
Ma qualité de jeu est meilleure lorsque j'improviser.	14%	<b>27%</b>	<b>29%</b>	18%	11%
Ma qualité de jeu est meilleure lorsque je joue une partition.	9%	<b>24%</b>	<b>30%</b>	<b>20%</b>	17%
J'ai l'impression que ce que j'entends quand je joue n'est pas la même chose que ce qu'entendent les spectateurs.	14%	<b>40%</b>	<b>22%</b>	15%	9%
Lorsque je joue une partition/tablatrice dans un style, je ressens des différences physiques par rapport à lorsque je joue une partition dans un style différent.	<b>30%</b>	<b>43%</b>	16%	8%	3%
La partition/tablatrice m'aide à mieux jouer.	14%	<b>25%</b>	<b>21%</b>	<b>25%</b>	15%
La timidité influence la qualité du jeu.	<b>42%</b>	<b>35%</b>	12%	6%	5%
Quand je joue, je ne pense pas en même temps.	14%	<b>27%</b>	<b>20%</b>	<b>21%</b>	18%

**Tableau 19:** Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel MSS. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% =$  ou  $> X > 20\%$ .

Nous analyserons ces données en cinq catégories thématiques :

**Rôle du musicien** : à travers les réponses données par le panel de musiciens n'ayant pas fait de solfège, un « *bon musicien* » est celui qui apporte sa personnalité dans l'œuvre qu'il interprète (97%) mais ne doit pas reproduire les intentions du compositeur (55%).

**Intrapsychique** : en général, les musiciens MSS du panel pensent aux différentes techniques que nécessite un type de musique (48%) et pensent aux différents éléments d'une partition/tablatrice (tempo, rythme, tonalité, etc.) avant de la déchiffrer (58%) mais également avant de jouer une partition/tablatrice qu'ils connaissent (57%). Les sondés MSS ne montrent pas de tendance précise à savoir s'ils pensent en même temps qu'ils jouent.

**Influence de l'écrit** : Selon le panel, il n'y a pas d'avis tranché sur le fait que jouer une partition puisse être plus facile que d'improviser cependant une légère tendance montre que la qualité de jeu serait meilleure lorsqu'ils improvisent (41%). De la même manière, il n'y a pas de tendance importante sur le fait que la qualité de jeu serait perçue par les sondés comme meilleure lorsqu'ils jouent une partition, ni sur le fait que la partition serait une aide afin de mieux jouer. Toutefois, 73% des sondés ressentent une différence physique lorsqu'ils changent de style de musique.

**Interpsychique**: les sondés présentent une impression modérée sur la différence entre ce qu'ils entendent de leur jeu musical et ce qu'entendent les spectateurs (54%).

**Émotion** : la timidité serait une variable importante dans la qualité de jeu (77%).

### 4.2.1.2 Les locuteurs-plurilingues

Les treize phrases affirmatives des panels LPR et LPNR sont indiquées dans le Tableau 20.

1	Parler comme un natif est le plus important pour bien parler une langue étrangère	7	Ma prononciation est meilleure quand je parle spontanément à l'oral.
2	Transmettre des émotions est le plus important pour bien parler une langue étrangère	8	Ma prononciation est meilleure quand je lis un texte.
3	Quand je parle dans ma langue maternelle, je pense aux différentes caractéristiques de la langue	9	J'ai l'impression que ce que j'entends quand je parle n'est pas la même chose que ce qu'entendent mes interlocuteurs.
4	Avant d'utiliser une langue étrangère à l'oral, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	10	Lorsque je parle/lis un texte en langue étrangère, je ressens des différences physiques par rapport à lorsque je parle/lis un texte en langue maternelle.
5	Avant de lire un texte dans une langue étrangère, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	11	La lecture m'aide à mieux prononcer.
6	Il est plus difficile de lire un texte que de parler spontanément en langue étrangère	12	La timidité influence la qualité de l'oral.
13	Quand je parle en langue étrangère, je ne pense pas en même temps.		

**Tableau 20:** Les 13 phrases affirmatives du questionnaire à échelle d'attitude des locuteurs-plurilingues (LPR/LPNR).

Nous analyserons les résultats pour chaque sous-panel de locuteurs plurilingues :

- Locuteurs-plurilingues avec pratique régulière (LPR)

Les résultats sont présentés sous forme de tableau présentant les phrases affirmatives et le pourcentage de réponses pour les cinq gradients de réponse (Tableau 21).

	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Parler comme un natif est le plus important pour bien parler une langue étrangère	<b>22%</b>	<b>36%</b>	<b>26%</b>	11%	5%
Transmettre des émotions est le plus important pour bien parler une langue étrangère	18%	<b>50%</b>	<b>25%</b>	6%	2%
Quand je parle dans ma langue maternelle, je pense aux différentes caractéristiques de la langue	16%	19%	<b>21%</b>	<b>26%</b>	19%
Avant d'utiliser une langue étrangère à l'oral, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	17%	<b>34%</b>	<b>20%</b>	18%	11%
Avant de lire un texte dans une langue étrangère, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	14%	<b>29%</b>	20%	<b>22%</b>	16%
Il est plus difficile de lire un texte que de parler spontanément en langue étrangère	9%	14%	15%	<b>28%</b>	<b>34%</b>
Ma prononciation est meilleure quand je parle spontanément à l'oral.	20%	<b>28%</b>	<b>24%</b>	20%	10%
Ma prononciation est meilleure quand je lis un texte.	12%	<b>33%</b>	<b>27%</b>	19%	9%
J'ai l'impression que ce que j'entends quand je parle n'est pas la même chose que ce qu'entendent mes interlocuteurs.	9%	<b>27%</b>	<b>23%</b>	<b>25%</b>	16%
Lorsque je parle/lis un texte en langue étrangère, je ressens des différences physiques par rapport à lorsque je parle/lis un texte en langue maternelle.	<b>29%</b>	<b>36%</b>	17%	12%	7%
La lecture m'aide à mieux prononcer.	<b>24%</b>	<b>35%</b>	<b>23%</b>	13%	5%
La timidité influence la qualité de l'oral.	<b>54%</b>	<b>30%</b>	8%	5%	3%
Quand je parle en langue étrangère, je ne pense pas en même temps.	10%	18%	<b>26%</b>	<b>24%</b>	<b>22%</b>

**Tableau 21:** Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel LPR. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% = \text{ou} > X > 20\%$ .

Nous analyserons ces données en cinq catégories thématiques :

**Rôle du locuteur :** à travers les réponses données par le panel de locuteurs-plurilingues ayant une pratique régulière d'au moins une langue étrangère en contexte naturel, nous pouvons constater qu'il serait important de parler comme un natif pour bien parler une langue étrangère (58%) mais qu'il faut également être capable de transmettre des émotions (68%).

**Intrapsychique :** les réponses des LPR ne montrent pas de tendance importante à savoir s'ils pensent aux différentes caractéristiques de la langue quand ils parlent leur langue maternelle, ou quand ils lisent un texte en langue étrangère. Ils montrent néanmoins une légère tendance sur le fait de penser aux différents éléments qui composent une langue étrangère avant de l'utiliser à l'oral (51%). 46% des personnes n'ont pas l'impression de penser pendant qu'ils parlent une langue étrangère.

**Influence de l'écrit** : 62% des LPR ne pensent pas qu'il soit plus difficile de lire un texte que de parler spontanément à l'oral une langue étrangère néanmoins 48% pensent que leur prononciation est meilleure quand ils parlent à l'oral et 45% quand ils lisent un texte. 59% des LPR pensent que la lecture les aide à mieux prononcer. Toutefois, 65% des sondés ressentent une différence physique lorsqu'ils parlent/lisent un texte en langue étrangère comparée aux mêmes actes en langue maternelle.

**Interpsychique** : les LPR ne montrent pas de tendance sur la prise de conscience d'une différence entre leur production et la perception par les interlocuteurs.

**Émotion** : la timidité influencerait l'oral pour 84% des sondés.

- Locuteurs-plurilingues avec pratique non-régulière (LPNR)

Les résultats sont présentés sous forme de tableau présentant les phrases affirmatives et le pourcentage de réponses pour les cinq gradients de réponse (Tableau 22).

	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Parler comme un natif est le plus important pour bien parler une langue étrangère	20%	<b>44%</b>	<b>27%</b>	8%	2%
Transmettre des émotions est le plus important pour bien parler une langue étrangère	15%	<b>37%</b>	<b>34%</b>	12%	3%
Quand je parle dans ma langue maternelle, je pense aux différentes caractéristiques de la langue	13%	19%	13%	<b>23%</b>	<b>32%</b>
Avant d'utiliser une langue étrangère à l'oral, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	16%	<b>44%</b>	<b>21%</b>	13%	5%
Avant de lire un texte dans une langue étrangère, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	11%	<b>27%</b>	<b>23%</b>	<b>28%</b>	10%
Il est plus difficile de lire un texte que de parler spontanément en langue étrangère	10%	17%	18%	<b>22%</b>	<b>34%</b>
Ma prononciation est meilleure quand je parle spontanément à l'oral.	18%	<b>32%</b>	<b>23%</b>	20%	7%
Ma prononciation est meilleure quand je lis un texte.	12%	<b>31%</b>	<b>29%</b>	19%	8%
J'ai l'impression que ce que j'entends quand je parle n'est pas la même chose que ce qu'entendent mes interlocuteurs.	12%	<b>32%</b>	<b>30%</b>	19%	7%
Lorsque je parle/lis un texte en langue étrangère, je ressens des différences physiques par rapport à lorsque je parle/lis un texte en langue maternelle.	<b>30%</b>	<b>34%</b>	18%	15%	4%
La lecture m'aide à mieux prononcer.	<b>21%</b>	<b>42%</b>	18%	12%	6%
La timidité influence la qualité de l'oral.	<b>64%</b>	<b>21%</b>	7%	6%	2%
Quand je parle en langue étrangère, je ne pense pas en même temps.	5%	<b>21%</b>	<b>26%</b>	<b>28%</b>	20%

**Tableau 22:** Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel LPNR. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% = \text{ou} > X > 20\%$ .

Nous analyserons ces données en cinq catégories thématiques :

**Rôle du locuteur** : à travers les réponses données par le panel de locuteurs-plurilingues n'ayant pas de pratique régulière d'au moins une langue étrangère en contexte naturel, nous pouvons constater qu'il serait important de parler comme un natif pour bien parler une langue étrangère (64%) mais qu'il faut également être capable de transmettre des émotions (52%).

**Intrapsychique** : les réponses des LPNR montrent une tendance remarquable contre le fait qu'ils penseraient aux différentes caractéristiques de la langue quand ils parlent leur langue maternelle (55%). A l'opposé, ils penseraient aux différentes caractéristiques de la langue quand ils parlent une langue étrangère (60%), mais nous ne pouvons mettre en avant de tendance en ce qui concerne la lecture en langue étrangère. Il n'y a pas de tendance importante sur l'impression de penser pendant qu'ils parlent une langue étrangère.

**Influence de l'écrit** : 56% des LPNR ne pensent pas qu'il est plus difficile de lire un texte que de parler spontanément à l'oral une langue étrangère néanmoins 50% pensent que leur prononciation est meilleure quand ils parlent à l'oral et 43% quand ils lisent un texte. 63% des LPNR pensent que la lecture les aide à mieux prononcer. Toutefois, 64% des sondés ressentent une différence physique lorsqu'ils parlent/lisent un texte en langue étrangère comparée aux mêmes actes en langue maternelle.

**Interpsychique** : les LPNR montrent une légère tendance (44%) sur la prise de conscience d'une différence entre leur production et la perception par les interlocuteurs.

**Émotion** : la timidité influencerait l'oral pour 85% des sondés.



### 4.2.1.3 Les locuteurs-signeurs

Les treize phrases affirmatives du panel LS sont indiquées dans le Tableau 23.

1	Communiquer (LS ou LO) comme un natif est le plus important pour bien communiquer	7	Je communique mieux quand je signe spontanément que quand je parle à l'oral spontanément.
2	Transmettre des émotions est le plus important pour bien communiquer (LS ou LO)	8	Ma prononciation est meilleure quand je lis un texte à l'oral que quand je parle spontanément à l'oral.
3	Quand je signe, je pense aux différentes caractéristiques de la langue.	9	J'ai l'impression que ce que je signe n'est pas la même chose que ce que perçoivent mes interlocuteurs.
4	Avant de signer je pense aux différentes composantes: gestes, mouvements, etc.	10	Il est plus facile de lire des phrases d'un texte en signant que de les lire à l'oral.
5	Avant de signer des phrases d'un texte, je pense aux différents éléments qui les composent: gestes, mouvements, etc.	11	La lecture m'aide à mieux signer.
6	Il est plus difficile de signer des phrases d'un texte que de signer spontanément.	12	La timidité influence la qualité de communication lorsque je signe.
13	Quand je signe, j'oralise dans ma tête en même temps.		

**Tableau 23:** Les 13 phrases affirmatives du questionnaire à échelle d'attitude des locuteurs-signeurs (LS).

Les résultats sont présentés sous forme de tableau présentant les phrases affirmatives et le pourcentage de réponses pour les cinq gradients de réponse (Tableau 24).

	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Communiquer (LS ou LO) comme un natif est le plus important pour bien communiquer	<b>33%</b>	<b>33%</b>	<b>22%</b>	9%	3%
Transmettre des émotions est le plus important pour bien communiquer (LS ou LO)	<b>43%</b>	<b>42%</b>	13%	2%	0%
Quand je signe, je pense aux différentes caractéristiques de la langue.	<b>24%</b>	<b>27%</b>	<b>22%</b>	18%	9%
Avant de signer je pense aux différentes composantes: gestes, mouvements, etc.	<b>20%</b>	19%	<b>28%</b>	15%	18%
Avant de signer des phrases d'un texte, je pense aux différents éléments qui les composent: gestes, mouvements, etc.	<b>23%</b>	<b>43%</b>	11%	12%	12%
Il est plus difficile de signer des phrases d'un texte que de signer spontanément.	<b>36%</b>	<b>43%</b>	14%	5%	2%
Je communique mieux quand je signe spontanément que quand je parle à l'oral spontanément.	7%	10%	<b>22%</b>	<b>31%</b>	<b>31%</b>
Ma prononciation est meilleure quand je lis un texte à l'oral que quand je parle spontanément à l'oral.	8%	<b>25%</b>	<b>25%</b>	<b>20%</b>	<b>22%</b>
J'ai l'impression que ce que je signe n'est pas la même chose que ce que perçoivent mes interlocuteurs.	3%	16%	<b>28%</b>	<b>29%</b>	<b>23%</b>
Il est plus facile de lire des phrases d'un texte en signant que de les lire à l'oral.	1%	5%	13%	<b>25%</b>	<b>57%</b>
La lecture m'aide à mieux signer.	1%	6%	<b>20%</b>	<b>26%</b>	<b>48%</b>
La timidité influence la qualité de communication lorsque je signe.	<b>23%</b>	<b>38%</b>	13%	13%	13%
Quand je signe, j'oralise dans ma tête en même temps.	<b>20%</b>	16%	17%	19%	<b>28%</b>

**Tableau 24:** Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel LS. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% = \text{ou} > X > 20\%$ .

Nous analyserons ces données en cinq catégories thématiques :

**Rôle du locuteur :** à travers les réponses données par le panel de locuteurs-signeurs, nous pouvons constater qu'il serait important de parler comme un natif pour bien parler une langue orale ou une langue des signes (66%) mais qu'il faut également être capable de transmettre des émotions (85%).

**Intrapsychique :** les réponses des LS montrent une légère tendance au fait qu'ils puissent penser aux différentes caractéristiques de la langue quand ils signent (51%) et quand ils signent des phrases d'un texte (66%). Par contre, il n'y a pas de tendance précise sur le fait de penser aux différentes composantes de la langue avant de signer. Il n'y a pas de tendance remarquable sur l'impression de penser pendant qu'ils parlent une langue étrangère.

**Influence de l'écrit** : 79% des LS pensent qu'il est plus difficile de signer un texte que de signer spontanément. 62% des sondés ne pensent pas qu'ils communiquent mieux quand ils signent spontanément que quand ils parlent à l'oral spontanément. Cependant, une très légère tendance montrant que les LS penseraient que leur prononciation est meilleure quand ils lisent un texte à l'oral que quand ils parlent spontanément à l'oral. Nous constatons une forte tendance (82%) contre le fait qu'il soit plus facile de lire des phrases d'un texte en signant que de les lire à l'orale. La lecture n'aiderait pas à mieux signer pour 74% d'entre eux.

**Interpsychique** : les LS montrent une légère tendance (46%) contre le fait d'une différence entre leur production et la perception par les interlocuteurs.

**Émotion** : la timidité influencerait la qualité de communication en langue des signes pour 61% des sondés.

#### 4.2.2 Questionnaires à choix multiples

Le questionnaire à choix multiples des cinq panels se compose de cinq questions acceptant une ou plusieurs réponses (voir 4.1.1.1). Les cinq questions et les choix possibles ne sont pas tous similaires mais sont tous correspondants en adaptant les termes en fonction des objets d'étude (Tableau 25).

Panel	Musiciens (MAS/MSS)	Locuteurs-plurilingues (LPR/LPNR)	Locuteurs-signeurs (LS)
Question	Quand je change de genre de musique, je souhaite	Quand je change de langue, je souhaite	Quand je change de langue (LO/LS), je souhaite
Réponse	Intégrer les éléments de chaque style.	Intégrer les éléments culturels de la LE.	Intégrer les éléments culturels de la LO/LS.

**Tableau 25:** Exemple de correspondances entre les questions et réponses dans le questionnaire à choix multiples des 5 panels. LO correspondant à langue orale, LS à langue des signes et LE à langue étrangère.

Nous présenterons les résultats de chaque panel indépendamment : les musiciens, les locuteurs-plurilingues et les locuteurs-signeurs.

### 4.2.2.1 Les Musiciens

Nous analyserons les résultats aux questionnaires à choix multiple de manière dissociée pour les musiciens avec solfège (MAS) et les musiciens sans solfège (MSS) :

- Musiciens avec solfège (MAS)

Les résultats sont présentés sous forme de tableau présentant les questions, les choix multiples et le pourcentage de réponses pour chaque choix (Tableau 26).

Lorsque je change de style de musique (classique, baroque, moderne, folk, etc..) j'ai l'impression:	D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	<b>45%</b>	<b>38%</b>	17%
Lorsque je change de style de musique (classique, baroque, moderne, folk, etc..) j'ai l'impression:	D'être dans le même état d'esprit.		De changer d'état d'esprit.
	16%		<b>84%</b>
Je pense que la partition et l'œuvre jouée:	Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	<b>65%</b>	10%	25%
Selon moi, le plus important pour bien jouer c'est:	Retour positif des spectateurs	A l'aise même si erreurs	La qualité du son
	19%	<b>41%</b>	<b>40%</b>
Quand je change de genre de musique	Intégrer les éléments de chaque style	Garder mon identité personnelle	Garder mon identité culturelle
	<b>41%</b>	<b>45%</b>	14%

**Tableau 26:** Résultats en pourcentages de réponses au questionnaire à choix multiples du panel MAS. En gras surligné gris : pourcentages  $X = ou > 50\%$  ; en gras : pourcentages  $30\% = ou < X$ .

Nous analyserons ces données en deux catégories thématiques :

**Intra- et interpsychique :** les musiciens ayant fait du solfège ont répondu à 45% que lorsqu'ils changent de style de musique, ils ont l'impression d'agir différemment avec les autres et/ou 38% d'être la même personne mais peu (17%) pensent changer de personnalité. Dans le même temps, 84% d'entre eux pensent changer d'état d'esprit en changeant de style de musique. Quand les MAS changent de genre de musique, ils pensent à 41% intégrer les éléments de chaque style et/ou à 45% de garder leur identité personnelle. Le plus important pour bien jouer, selon eux, est être à l'aise même s'il y a des erreurs (41%) et la qualité du son (40%) mais pas le retour positif des spectateurs (19%).

**Influence écriture** : Selon les MAS, la partition et l'œuvre jouée sont à 65% naturellement liées.

- Musiciens sans solfège (MSS)

Les résultats sont présentés sous forme de tableau présentant les questions, les choix multiples et le pourcentage de réponses pour chaque choix (Tableau 27).

Lorsque je change de style de musique (classique, baroque, moderne, folk, etc..) j'ai l'impression:	D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	<b>39%</b>	<b>39%</b>	22%
Lorsque je change de style de musique (classique, baroque, moderne, folk, etc..) j'ai l'impression:	D'être dans le même état d'esprit.		De changer d'état d'esprit.
	19%		<b>81%</b>
Je pense que la partition et l'œuvre jouée:	Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	<b>51%</b>	15%	<b>34%</b>
Selon moi, le plus important pour bien jouer c'est:	Retour positif des spectateurs	A l'aise même si erreurs	La qualité du son
	19%	<b>42%</b>	<b>39%</b>
Quand je change de genre de musique	Intégrer les éléments de chaque style	Garder mon identité personnelle	Garder mon identité culturelle
	<b>37%</b>	<b>42%</b>	21%

**Tableau 27:** Résultats en pourcentages de réponses au questionnaire à choix multiples du panel MSS. En gras surligné gris : pourcentages  $X = ou > 50\%$  ; en gras : pourcentages  $30\% = ou < X$ .

Nous analyserons ces données en deux catégories thématiques :

**Intra- et interpsychique** : les musiciens n'ayant pas fait du solfège ont répondu à 39% que lorsqu'ils changent de style de musique, ils ont l'impression d'agir différemment avec les autres et/ou 39% d'être la même personne mais peu (22%) pensent changer de personnalité. Dans le même temps 81% d'entre eux pensent changer d'état d'esprit en changeant de style de musique.

Quand ils changent de genre de musique, ils pensent à 37% intégrer les éléments de chaque style et/ou à 42% de garder leur identité personnelle. Le plus important pour bien jouer, selon eux, c'est être à l'aise même s'il y a des erreurs (42%) et la qualité du son (39%) mais pas le retour positif des spectateurs (19%).

**Influence de l'écriture** : selon les MSS, la partition et l'œuvre jouée sont à 51% naturellement liées et/ou à 34% fonctionnent différemment.

### 4.2.2.2 Les locuteurs-plurilingues

Nous analyserons les résultats aux questionnaires à choix multiple de manière dissociée pour les locuteurs-plurilingues avec pratique régulière (LPR) et les locuteurs-plurilingues avec pratique non-régulière (LPNR) :

- Locuteurs-plurilingues avec pratique régulière (LPR)

Les résultats sont présentés sous forme de tableau présentant les questions, les choix multiples et le pourcentage de réponses pour chaque choix (Tableau 28).

Lorsque je change de langue, j'ai l'impression:	D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	<b>37%</b>	<b>43%</b>	20%
Lorsque je change de langue, j'ai l'impression:	D'être dans le même état d'esprit.		De changer d'état d'esprit.
	<b>47%</b>		<b>53%</b>
Je pense que l'écriture et l'oral	Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	<b>38%</b>	4%	<b>58%</b>
Selon moi, le plus important pour bien parler c'est:	Être compris et à l'aise	A l'aise même si erreurs	La qualité de la forme
	<b>47%</b>	<b>30%</b>	23%
Quand je change de langue:	Intégrer les éléments culturels de la LE	Garder mon identité personnelle	Garder mon identité culturelle
	<b>40%</b>	<b>40%</b>	20%

**Tableau 28:** Résultats en pourcentages de réponses au questionnaire à choix multiples du panel LPR. En gras surligné gris : pourcentages  $X = ou > 50\%$  ; en gras : pourcentages  $30\% = ou < X$ .

Nous analyserons ces données en deux catégories thématiques :

**Intra- et interpsychique :** les locuteurs-plurilingues avec pratique régulière ont répondu à 37% que lorsqu'ils changent de langue, ils ont l'impression d'agir différemment avec les autres et/ou 43% d'être la même personne mais peu (20%) pensent changer de personnalité. Néanmoins, il n'y a pas de tendance marquante sur le fait qu'ils soient dans le même d'état d'esprit ou non quand ils changent de langue.

Quand les LPR changent de langue, ils pensent à 40% intégrer les éléments culturels de la LE et/ou à 40% de garder leur identité personnelle. Le plus important pour bien parler, selon eux, est être à l'aise même s'il y a des erreurs (30%) et compréhensible (42%) mais pas la qualité de la forme (23%).

**Influence écriture :** Selon les LPR, l'écriture et l'oral fonctionnent différemment à 58% et sont naturellement liés à 38%.

- Locuteurs-plurilingues avec pratique non-régulière (LPNR)

Les résultats sont présentés sous forme de tableau présentant les questions, les choix multiples et le pourcentage de réponses pour chaque choix (Tableau 29).

Lorsque je change de langue, j'ai l'impression:	D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	<b>36%</b>	<b>50%</b>	15%
Lorsque je change de langue, j'ai l'impression:	D'être dans le même état d'esprit.		De changer d'état d'esprit.
	<b>52%</b>		<b>48%</b>
Je pense que l'écriture et l'oral	Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	<b>40%</b>	4%	<b>56%</b>
Selon moi, le plus important pour bien parler c'est:	Être compris et à l'aise	A l'aise même si erreurs	La qualité de la forme
	<b>57%</b>	28%	16%
Quand je change de langue:	Intégrer les éléments culturels de la LE	Garder mon identité personnelle	Garder mon identité culturelle
	<b>33%</b>	<b>45%</b>	22%

**Tableau 29:** Résultats en pourcentages de réponses au questionnaire à choix multiples du panel LPNR. En gras surligné gris : pourcentages  $X = ou > 50%$  ; en gras : pourcentages  $30% = ou < X$ .

Nous analyserons ces données en deux catégories thématiques :

**Intra- et interpsychique :** les locuteurs-plurilingues avec pratique non-régulière ont répondu à 36% que lorsqu'ils changent de langue, ils ont l'impression d'agir différemment avec les autres et/ou 50% d'être la même personne mais peu (15%) pensent changer de personnalité. Néanmoins, il n'y a pas de tendance marquante sur le fait qu'ils soient dans le même d'état d'esprit ou non quand ils changent de langue.

Quand ils changent de langue, ils pensent à 33% intégrer les éléments culturels de la LE et/ou à 45% de garder leur identité personnelle. Le plus important pour bien parler, selon eux, c'est d'être à l'aise et compréhensible (57%) puis d'être à l'aise même s'il y a des erreurs (28%) et enfin la qualité de la forme (16%).

**Influence écriture :** selon les LPNR, l'écriture et l'oral fonctionnent différemment à 56% et sont naturellement liés à 40%.

### 4.2.2.3 Locuteurs-signeurs

Les résultats sont présentés sous forme de tableau présentant les questions, les choix multiples et le pourcentage de réponses pour chaque choix (Tableau 30).

Lorsque je change de langue (orale/ des signes), j'ai l'impression:	D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	<b>25%</b>	<b>62%</b>	13%
Lorsque je change de langue (orale/ des signes), j'ai l'impression:	D'être dans le même état d'esprit.		De changer d'état d'esprit.
	<b>55%</b>		<b>45%</b>
Je pense que l'écriture et l'oral	Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	<b>27%</b>	7%	<b>66%</b>
Selon moi, le plus important pour bien communiquer c'est:	Être compris et à l'aise	A l'aise même si erreurs	La qualité de la forme
	<b>49%</b>	38%	13%
Quand je change de langue (LO/LS):	Intégrer les éléments culturels de la LO/LS	Garder mon identité personnelle	Garder mon identité culturelle
	<b>46%</b>	<b>38%</b>	16%

**Tableau 30:** Résultats en pourcentages de réponses au questionnaire à choix multiples du panel LS. En gras surligné gris : pourcentages  $X =$  ou  $> 50\%$  ; en gras : pourcentages  $30\% =$  ou  $< X$ .

Nous analyserons ces données en deux catégories thématiques :

**Intra- et interpsychique :** les locuteurs-signeurs ont répondu à 62% que lorsqu'ils changent de langue, ils ont l'impression d'être la même personne et/ou 25% d'agir différemment avec les autres mais peu (13%) pensent changer de personnalité. Néanmoins, il n'y a pas de tendance importante sur le fait qu'ils soient dans le même d'état d'esprit ou non quand ils changent de langue.

Quand ils changent de langue, ils pensent à 46% intégrer les éléments culturels de la LO/LS et/ou à 38% de garder leur identité personnelle. Le plus important pour bien parler, selon eux, c'est d'être à l'aise et compréhensible (49%) puis d'être à l'aise même s'il y a des erreurs (38%) et enfin la qualité de la forme (13%).

**Influence écriture :** selon les LS, l'écriture et l'oral fonctionnent différemment à 66% et sont naturellement liés à 27%.



### **4.3 Analyses et discussions**

L'analyse des résultats sera présentée de manière thématique afin de mettre en relation les résultats des différents panels et ainsi mettre en évidence l'influence de certaines variables telles que l'apprentissage du solfège pour la musique, la fréquence et le type d'utilisation d'une langue étrangère avec les locuteurs-plurilingues et la modalité de langage avec les locuteurs-signeurs. Nous étudierons, dans un premier temps, les liens entre musiciens et langage musical, puis ceux entre locuteurs-plurilingues et langage verbal. A travers ces premières correspondances, nous verrons les liens entre langages musical et verbal. Pour finir, nous proposerons une ouverture sur l'influence de la modalité de langage sur les représentations des individus.

#### **4.3.1 Musiciens et langage musical**

L'analyse des liens entre musiciens et langage musical sera basée sur les résultats des questionnaires à échelle d'attitude et à choix multiples des panels MAS et MSS. Ils seront présentés sous forme de tableaux (Tableau 31 et Tableau 32) puis à travers une analyse complète.

De manière générale, nous pouvons affirmer que sur les points étudiés, à savoir l'influence de l'apprentissage du solfège sur les représentations sur le langage musical, très peu de différences peuvent être remarquées. Néanmoins, nous pouvons dire que, sur le plan socio-culturel, les musiciens qui font du solfège ont en général une formation plus traditionnelle, de type occidentale, ont commencé depuis un âge plus précoce et jouent une palette d'instruments plus variée que ceux qui ne font pas de solfège. Cette dernière peut être expliquée, entre autres, par des quotas par classe d'instrument dans les conservatoires mais également de la doxa qui favorise un apprentissage plus « *facile* » de la guitare, par exemple, que celui du violon. La question d'accès aux instruments en termes de prix, de connaissance générale est aussi un facteur pouvant expliquer cette différence. Les influences de l'apprentissage plus précoce de la musique a été mises en valeur par les travaux liés à la période de l'âge critique en montrant notamment que plus un individu commence tôt la musique, plus il est fréquent (voire indispensable) qu'il ait l'oreille absolue modifiant de ce fait les processus neurophysiologiques et la conception cognitive de l'objet musical (Levitin & Zatorre, 2003) (voir 1.2.2).

	MAS					MSS				
	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Un bon musicien doit seulement reproduire les intentions du compositeur.	3%	11%	<b>34%</b>	<b>27%</b>	<b>25%</b>	1%	11%	<b>34%</b>	<b>30%</b>	<b>25%</b>
Un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète.	<b>57%</b>	<b>37%</b>	5%	1%	0%	<b>62%</b>	<b>35%</b>	3%	0%	0%
Quand je joue de la musique, je pense aux différentes techniques que nécessite ce type de musique.	20%	<b>35%</b>	<b>22%</b>	17%	6%	14%	<b>34%</b>	<b>27%</b>	18%	7%
Avant de déchiffrer une partition/tablatrice, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	<b>41%</b>	<b>40%</b>	8%	8%	2%	<b>27%</b>	<b>31%</b>	16%	17%	9%
Avant de jouer une partition/tablatrice que je connais, je pense aux différents éléments de la partition/tablatrice: tempo, rythme, tonalité, etc.	<b>32%</b>	<b>35%</b>	17%	11%	5%	<b>24%</b>	<b>33%</b>	16%	16%	11%
Il est plus difficile de jouer une partition/tablatrice que d'improviser.	6%	15%	<b>26%</b>	<b>29%</b>	<b>25%</b>	12%	18%	<b>30%</b>	<b>22%</b>	17%
Ma qualité de jeu est meilleure lorsque j'improviser.	9%	16%	<b>30%</b>	<b>26%</b>	19%	14%	<b>27%</b>	<b>29%</b>	18%	11%
Ma qualité de jeu est meilleure lorsque je joue une partition.	<b>23%</b>	<b>27%</b>	<b>31%</b>	15%	5%	9%	<b>24%</b>	<b>30%</b>	<b>20%</b>	17%
J'ai l'impression que ce que j'entends quand je joue n'est pas la même chose que ce qu'entendent les spectateurs.	<b>30%</b>	<b>35%</b>	13%	17%	4%	14%	<b>40%</b>	<b>22%</b>	15%	9%
Lorsque je joue une partition/tablatrice dans un style, je ressens des différences physiques par rapport à lorsque je joue une partition dans un style différent.	<b>41%</b>	<b>35%</b>	16%	5%	3%	<b>30%</b>	<b>43%</b>	16%	8%	3%
La partition/tablatrice m'aide à mieux jouer.	11%	<b>27%</b>	<b>34%</b>	18%	10%	14%	<b>25%</b>	<b>21%</b>	<b>25%</b>	15%
La timidité influence la qualité du jeu.	<b>39%</b>	<b>37%</b>	12%	7%	4%	<b>42%</b>	<b>35%</b>	12%	6%	5%
Quand je joue, je ne pense pas en même temps.	16%	17%	<b>21%</b>	<b>23%</b>	<b>22%</b>	14%	<b>27%</b>	<b>20%</b>	<b>21%</b>	18%

**Tableau 31:** Résultats synthétiques en pourcentages de réponses au questionnaire à échelle d'attitude des panels MAS et MSS. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% =$  ou  $> X > 20\%$ .

Lorsque je change de style de musique (classique, baroque, moderne, folk, etc..) j'ai l'impression:		D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	MAS	<b>45%</b>	<b>38%</b>	17%
	MSS	<b>39%</b>	<b>39%</b>	22%
Lorsque je change de style de musique (classique, baroque, moderne, folk, etc..) j'ai l'impression:		D'être dans le même état d'esprit.		De changer d'état d'esprit.
	MAS	16%		<b>84%</b>
	MSS	19%		<b>81%</b>
Je pense que la partition et l'œuvre jouée:		Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	MAS	<b>65%</b>	10%	25%
	MSS	<b>51%</b>	15%	<b>34%</b>
Selon moi, le plus important pour bien jouer c'est:		Retour positif des spectateurs	A l'aise même si erreurs	La qualité du son
	MAS	19%	<b>41%</b>	<b>40%</b>
	MSS	19%	<b>42%</b>	<b>39%</b>
Quand je change de genre de musique		Intégrer les éléments de chaque style	Garder mon identité personnelle	Garder mon identité culturelle
	MAS	<b>41%</b>	<b>45%</b>	14%
	MSS	<b>37%</b>	<b>42%</b>	21%

**Tableau 32:** Résultats synthétiques en pourcentages de réponses au questionnaire à choix multiples des panels MAS et MSS. En gras surligné gris : pourcentages  $X =$  ou  $> 50%$  ; en gras : pourcentages  $30% =$  ou  $< X$ .

Malgré ces différences dans le profil socio-culturel des deux panels de musiciens, il est intéressant de constater qu'il semblerait que la variable du solfège, et celle de l'âge d'apprentissage ne modifient pas de manière considérable les représentations liées au langage musical. En effet, dans les deux cas, un « *bon* » musicien n'est pas seulement celui qui reproduit les intentions du compositeur, comme le souhaitait Stravinsky (voir 3.3.3), mais surtout un individu qui apporte sa personnalité et sa vision de l'œuvre dans son jeu. Nous pouvons ajouter qu'il ne semble pas y avoir de profil particulier sur la question à savoir si une position en faveur de la reproduction des intentions du compositeur serait en opposition formelle à une position en faveur d'apporter sa personnalité à l'œuvre (Tableau 33).

	Degré de différence				
	0	1	2	3	4
MSS	4%	24%	33%	19%	20%
MAS	8%	26%	28%	18%	21%

**Tableau 33:** Degré de différence entre les réponses au questionnaire à échelle d'attitude pour les affirmations : « un bon musicien doit seulement reproduire les intentions du compositeur » et « un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète ». Le résultat correspond à la valeur absolue de la différence entre les deux attitudes d'un même sondé. Les données ont été calculées sur la base de « tout à fait d'accord » = 1 ; « plutôt d'accord » = 2 ; « moyenne d'accord » = 3 etc. Les formules utilisées ont été « SI », « NB.SI » et « ABS ». Un degré 2 de différence dans le tableau correspond par exemple à « tout à fait d'accord » - « moyennement d'accord ». 0 correspond à un degré similaire pour les deux affirmations et 4 correspond à deux réponses totalement opposées.

De ce fait, nous avons pu constater que lorsque les musiciens (MAS et MSS) changent de style de musique, ils n'ont pas l'impression de changer de personnalité mais pensent changer d'état d'esprit. Ils pensent ainsi agir différemment avec les autres (musiciens ou spectateurs), tout en ayant la forte conviction d'être la même personne. Il leur importe de garder leur identité personnelle tout en intégrant les éléments spécifiques de chaque style. Ce changement d'esprit peut être expliqué lorsque les sondés montrent une forte tendance sur le fait qu'ils pensent aux différents éléments de la partition/tablatrice avant de déchiffrer ou de jouer une partition/tablatrice qu'ils connaissent déjà. Néanmoins, cette tendance semble légèrement plus faible lorsqu'il consiste à simplement « jouer de la musique ». Jouer une partition/tablatrice imposerait chez l'individu une prise de conscience des différentes composantes de l'œuvre écrite puis par extension à l'œuvre jouée même si pour cette dernière la tendance est atténuée. Cependant, nous remarquons que face à la partition, les MAS montrent une tendance plus importante à penser aux caractéristiques de la partition. L'apprentissage du solfège serait donc un catalyseur en faveur d'une prise de conscience métacognitive, face à la partition, des différents éléments qui la composent. Dans les deux panels, nous n'avons pas de tendance remarquable sur le fait qu'ils penseraient en même temps qu'ils jouent.

Concernant les formes écrite et sonore de l'œuvre, une légère tendance apparaît, dans les deux panels, sur le fait qu'il ne soit pas plus difficile de jouer une partition que d'improviser. Cependant les réponses des MSS montrent une légère tendance, contrairement à celle des MAS, sur le fait que la qualité de jeu serait meilleure lors de l'improvisation. A l'inverse, pour les musiciens ayant fait du solfège, la qualité de jeu serait meilleure lorsqu'ils jouent une partition.

Dans les deux cas, la partition/tablatrice n'aiderait pas à mieux jouer. Cette différence ne peut pas être expliquée en terme de conception des relations forme écrite/sonore car les deux panels montrent une forte tendance en faveur du fait que la partition et l'œuvre jouée sont naturellement liées. Néanmoins, les MSS montrent une légère tendance à penser qu'elles pourraient fonctionner différemment malgré leur lien naturel.

Dans les deux cas, il y a une très bonne prise de conscience sur les différences acoustiques et perceptives entre ce que le musicien joue et ce que les spectateurs perçoivent. Toutefois, et ce malgré cette prise de conscience, avoir un retour positif des spectateurs n'apparaît pas comme la principale réponse à savoir ce qui est le plus important pour « *bien* » jouer. Le plus important pour eux étant d'être à l'aise même s'il y a des erreurs et la qualité du son. Cette qualité pouvant être largement influencée par la timidité comme une très large majorité des sondés le pense. Au niveau individuel, les musiciens perçoivent très majoritairement une différence physique lorsqu'ils changent de style de musique.

#### **4.3.2 Locuteurs et langage verbal**

L'analyse des liens entre locuteurs et langage verbal sera basée sur les résultats des questionnaires à échelle d'attitude et à choix multiples des panels LPR et LPNR. Ils seront présentés sous forme de tableaux (Tableau 34 et Tableau 35) suivis d'une analyse complète.

	LPR					LPNR				
	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Parler comme un natif est le plus important pour bien parler une langue étrangère	<b>22%</b>	<b>36%</b>	<b>26%</b>	11%	5%	20%	<b>44%</b>	<b>27%</b>	8%	2%
Transmettre des émotions est le plus important pour bien parler une langue étrangère	18%	<b>50%</b>	<b>25%</b>	6%	2%	15%	<b>37%</b>	<b>34%</b>	12%	3%
Quand je parle dans ma langue maternelle, je pense aux différentes caractéristiques de la langue	16%	19%	<b>21%</b>	<b>26%</b>	19%	13%	19%	13%	<b>23%</b>	<b>32%</b>
Avant d'utiliser une langue étrangère à l'oral, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	17%	<b>34%</b>	<b>20%</b>	18%	11%	16%	<b>44%</b>	<b>21%</b>	13%	5%
Avant de lire un texte dans une langue étrangère, je pense aux différents éléments qui la composent: prononciation, intentions, etc.	14%	<b>29%</b>	20%	<b>22%</b>	16%	11%	<b>27%</b>	<b>23%</b>	<b>28%</b>	10%
Il est plus difficile de lire un texte que de parler spontanément en langue étrangère	9%	14%	15%	<b>28%</b>	<b>34%</b>	10%	17%	18%	<b>22%</b>	<b>34%</b>
Ma prononciation est meilleure quand je parle spontanément à l'oral.	20%	<b>28%</b>	<b>24%</b>	20%	10%	18%	<b>32%</b>	<b>23%</b>	20%	7%
Ma prononciation est meilleure quand je lis un texte.	12%	<b>33%</b>	<b>27%</b>	19%	9%	12%	<b>31%</b>	<b>29%</b>	19%	8%
J'ai l'impression que ce que j'entends quand je parle n'est pas la même chose que ce qu'entendent mes interlocuteurs.	9%	<b>27%</b>	<b>23%</b>	<b>25%</b>	16%	12%	<b>32%</b>	<b>30%</b>	19%	7%
Lorsque je parle/lis un texte en langue étrangère, je ressens des différences physiques par rapport à lorsque je parle/lis un texte en langue maternelle.	<b>29%</b>	<b>36%</b>	17%	12%	7%	<b>30%</b>	<b>34%</b>	18%	15%	4%
La lecture m'aide à mieux prononcer.	<b>24%</b>	<b>35%</b>	<b>23%</b>	13%	5%	<b>21%</b>	<b>42%</b>	18%	12%	6%
La timidité influence la qualité de l'oral.	<b>54%</b>	<b>30%</b>	8%	5%	3%	<b>64%</b>	<b>21%</b>	7%	6%	2%
Quand je parle en langue étrangère, je ne pense pas en même temps.	10%	18%	<b>26%</b>	<b>24%</b>	<b>22%</b>	5%	<b>21%</b>	<b>26%</b>	<b>28%</b>	20%

**Tableau 34:** Résultats synthétiques en pourcentages de réponses au questionnaire à échelle d'attitude des panels LPR et LPNR. En gras et surligné gris : pourcentages  $X > 30\%$  ; en gras : pourcentages  $30\% = \text{ou} > X > 20\%$ .

Lorsque je change de langue, j'ai l'impression:		D'agir différemment avec les autres	D'être la même personne	De changer de personnalité
	LPR	<b>37%</b>	<b>43%</b>	20%
	LPNR	<b>36%</b>	<b>50%</b>	15%
Lorsque je change de langue, j'ai l'impression:		D'être dans le même état d'esprit.		De changer d'état d'esprit.
	LPR	<b>47%</b>		<b>53%</b>
	LPNR	<b>52%</b>		<b>48%</b>
Je pense que l'écriture et l'oral		Sont naturellement liés	Fonctionnent de la même manière	Fonctionnent différemment
	LPR	<b>38%</b>	4%	<b>58%</b>
	LPNR	<b>40%</b>	4%	<b>56%</b>
Selon moi, le plus important pour bien parler c'est:		Être compris et à l'aise	A l'aise même si erreurs	La qualité de la forme
	LPR	<b>47%</b>	<b>30%</b>	23%
	LPNR	<b>57%</b>	28%	16%
Quand je change de langue:		Intégrer les éléments culturels de la LE	Garder mon identité personnelle	Garder mon identité culturelle
	LPR	<b>40%</b>	<b>40%</b>	20%
	LPNR	<b>33%</b>	<b>45%</b>	22%

**Tableau 35:** Résultats synthétiques en pourcentages de réponses au questionnaire à choix multiples des panels LPR et LPNR. En gras surligné gris : pourcentages  $X =$  ou  $> 50\%$  ; en gras : pourcentages  $30\% =$  ou  $< X$ .

De manière générale, nous pouvons affirmer que sur les points étudiés, à savoir l'influence de la pratique régulière d'au moins une langue étrangère en contexte naturel sur les représentations sur le langage verbal, très peu de différences peuvent être remarquées entre les deux panels.

Sur le plan socio-culturel, les deux panels sont homogènes avec un apprentissage de la première LE entre 0 et 15 ans même s'il semblerait que les LPNR aient commencé proportionnellement plus tardivement que les LPR. Ils parlent majoritairement le français avec leur famille et/ou dans la vie de tous les jours et ont appris comme langue étrangère principalement l'anglais, l'espagnol, l'allemand et l'italien ce qui correspond aux langues étrangères les plus enseignées en France<sup>68</sup>. Au sein de cette homogénéité, nous constatons qu'il semblerait que la variable de la pratique en contexte naturel ne modifie pas de manière considérable les représentations liées au langage verbal.

En effet, dans les deux cas, pour « bien » parler une langue étrangère, il est plutôt important de parler comme un natif mais également de transmettre des émotions. De ce fait, nous avons pu constater que lorsque les locuteurs-plurilingues changent de langue, ils n'ont pas l'impression de changer de personnalité mais pensent être la même personne tout en agissant

68 Rapport d'information n°63 délivré au Sénat par Legendre, J. en 2003 au nom de la commission des affaires culturelles sur l'enseignement des langues étrangères en France.

différemment avec les autres. Néanmoins, il est intéressant de remarquer qu'il n'y a pas de tendance importante pour les LPR et les LPNR sur le fait qu'ils changent d'état d'esprit ou non quand ils changent de langue. Il leur importe peu de garder leur identité culturelle ; ce qui importe c'est de garder leur identité personnelle mais aussi d'intégrer les éléments culturels de la LE. Cette ambiguïté se retrouve lorsque les deux panels montrent une tendance modérée contre le fait qu'ils pensent aux différentes caractéristiques de la langue lorsqu'ils parlent leur(s) langue(s) première(s). Nous remarquerons que cette tendance est plus forte chez les LPNR. À l'inverse, les deux panels sont plutôt favorables au fait de penser aux caractéristiques de la langue lorsqu'ils parlent une langue étrangère. Cependant, il n'y a pas de tendance remarquable au fait de penser aux caractéristiques de la langue avant de lire un texte en langue étrangère. Dans les deux panels, nous n'avons pas de tendance remarquable sur le fait de penser en même temps qu'ils jouent.

Concernant les formes sonore et écrite du discours, il ne serait pas plus difficile de lire un texte que de parler spontanément en langue étrangère. Une tendance modérée montre que la prononciation serait meilleure quand ils parlent spontanément à l'oral mais également quand ils lisent un texte. Cependant, nous pouvons constater (Tableau 36) qu'une majorité des sondés (environ 60%) chez les LPNR et le LPR ont répondu avec 1 ou 2 degrés de différence dans leurs réponses. Il semblerait, par conséquent, que ceux qui trouvent que leur prononciation est meilleure lorsqu'ils parlent, pensent qu'elle est également meilleure à l'écrit.

	Degré de différence				
	0	1	2	3	4
LPNR	17%	31%	32%	11%	8%
LPR	19%	30%	31%	12%	8%

**Tableau 36:** Degré de différence entre les réponses au questionnaire à échelle d'attitude pour les affirmations : « Ma prononciation est meilleure quand je parle spontanément à l'oral » et « Ma prononciation est meilleure quand je lis un texte ». Le résultat correspond à la valeur absolue de la différence entre les deux attitudes d'un même sondé. Les données ont été calculées sur la base de « tout à fait d'accord » = 1 ; « Plutôt d'accord » = 2 ; « moyenne d'accord » = 3 etc. Les formules utilisées ont été « SI », « NB.SI » et « ABS ». Un degré 2 de différence dans le tableau correspond par exemple à « tout à fait d'accord » - « moyennement d'accord ». 0 correspond à un degré similaire pour les deux affirmations et 4 correspond à deux réponses totalement opposées.

Toutefois, une large majorité pense que la lecture aide à mieux prononcer. Malgré ce constat, nous pouvons voir qu'une grande majorité des LPR et LPNR pensent que l'écriture et l'oral fonctionnent de manière différente même si une majorité modérée pense qu'ils sont également naturellement liés.



Dans les deux cas, il y n'y pas de résultats remarquables sur la prise de conscience sur les différences acoustiques et perceptives entre ce que le locuteur produit et ce que les interlocuteurs perçoivent. Le plus important pour eux, pour « *bien* » parler, étant d'être à l'aise dans la production en facilitant la compréhension et ce même s'il y a des erreurs. La qualité de la forme n'était pas leur préoccupation principale. Cette qualité pouvant être largement influencée par la timidité comme une très large majorité des sondés le pense. Au niveau individuel, les locuteurs-plurilingues perçoivent très fortement une différence physique lorsqu'ils changent de langue (L1/LE(s)).

### 4.3.3 Langages musical et verbal

Nous avons pu voir à travers l'analyse des données, à l'intérieur de chaque panel, les musiciens (MAS et MSS) d'un côté et les locuteurs-plurilingues (LPR et LPNR), que les représentations sur leurs rapports aux formes sonores et écrites de leurs langages (musique et discours) semblent être homogènes et ce au-delà des variables étudiées : l'apprentissage du solfège et la pratique en contexte naturel.

Cette partie étudiera plus particulièrement les ressemblances ou différences qui peuvent exister dans les représentations sur les formes sonores et écrites des langages étudiés (la musique et le discours) afin de voir si ces relations dépassent les caractéristiques de la frontière psychosociologique et disciplinaire entre ces domaines. En outre, si le fait que les musiciens soient également des locuteurs(-plurilingues) était une variable remarquable, nous devrions retrouver des tendances similaires entre les panels LPR/LPNR et MAS/MSS. Nous tenons à préciser, à nouveau, que les musiciens ont été interrogés sur leur rapport au langage musical et les locuteurs-plurilingues au langage verbal uniquement. Nous avons ainsi limité le biais du fait que certains musiciens puissent être plurilingues et certains locuteurs-plurilingues, des musiciens. Cependant, nous pouvons comparer ces données étant donné l'homogénéité sur le plan socio-culturel de ces panels en termes d'âge (15-25 ans), de culture (occidentale) et cognitif (haut niveau d'étude). Nous noterons, néanmoins, que les effectifs étant différents, la représentativité sera différente en fonction des groupes. L'analyse portera sur cinq dimensions : le rôle du locuteur/musicien, les aspects intra- et interpsychiques, émotionnels et l'influence de l'écrit sur les représentations des formes orales.

Nous comparerons tout d'abord, les représentations liées au **rôle du locuteur/musicien**. Nous pouvons constater que le rapport à la « *norme* » est totalement différent dans un groupe et dans l'autre (Tableau 37).

		Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Parler comme un natif est le plus important pour bien parler une langue étrangère	LPR	<b>22%</b>	<b>36%</b>	<b>26%</b>	11%	5%
	LPNR	<b>20%</b>	<b>44%</b>	<b>27%</b>	8%	2%
Transmettre des émotions est le plus important pour bien parler une langue étrangère	LPR	18%	<b>50%</b>	<b>25%</b>	6%	2%
	LPNR	15%	<b>37%</b>	<b>34%</b>	12%	3%

		Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
Un bon musicien doit seulement reproduire les intentions du compositeur.	MAS	3%	11%	<b>34%</b>	<b>27%</b>	<b>25%</b>
	MSS	1%	11%	<b>34%</b>	<b>30%</b>	<b>25%</b>
Un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète.	MAS	<b>57%</b>	<b>37%</b>	5%	1%	0%
	MSS	<b>62%</b>	<b>35%</b>	3%	0%	0%

**Tableau 37:** Résultats comparatifs en pourcentages de réponses au questionnaire à échelle d'attitude pour deux affirmations sur le rôle du locuteur / musicien. En gras, les pourcentages supérieurs ou égaux à 20% et en gras surligné gris, les pourcentages supérieurs ou égaux à 50%.

La partition est posée en tant que norme, étant donné qu'elle reflète les intentions de l'auteur. L'œuvre initiale n'existe que dans l'esprit du compositeur. Dans le cas du discours, le locuteur natif est devenu dans la doxa, une des normes prédominantes. Il est intéressant de constater que même si les deux groupes (locuteur-plurilingues et musiciens) affirment une tendance forte en faveur d'intégrer une charge émotionnelle et personnelle dans l'acte sonore, les locuteurs-plurilingues chercheraient à atteindre la norme du « natif » tandis que les musiciens devraient s'écarter de la norme écrite (la partition), se positionnant ainsi comme des interprètes, et non comme des exécutants (Corbel, 2005 : 78). Le musicien ne pourrait intégrer sa personnalité dans l'œuvre que s'il se détache des intentions premières du compositeur afin d'y intégrer son intentionnalité émotionnelle. Le musicien ne pourrait intégrer sa personnalité dans l'œuvre que s'il se détache des intentions premières du compositeur. Le locuteur-plurilingue ne voit pas d'opposition entre d'une part, son intentionnalité émotionnelle et d'autre part, le fait d'atteindre une norme externe : le « natif ».

Afin d'approfondir cette apparente opposition entre émotion et « norme », il nous semble intéressant d'analyser les rapports au **niveau intrapsychique** afin de voir si cette opposition perdure à ce niveau. Les musiciens montrent une forte tendance à penser aux différentes caractéristiques du type de musique qu'ils jouent que ce soit à l'aide d'une partition ou de manière improvisée. Ce sentiment est renforcé chez les musiciens ayant fait du solfège. Il est intéressant de voir que cette tendance ne se retrouve pas chez les locuteurs-plurilingues. En outre, il n'y a pas de tendance remarquable sur le fait que les locuteurs pensent aux différentes

caractéristiques de la langue quand ils parlent dans leur langue maternelle ni avant de lire un texte dans une langue étrangère. Néanmoins, nous pouvons remarquer que les deux panels LPR et LPNR montrent une tendance positive, pour ce genre de pratique, avant d'utiliser une langue étrangère à l'oral. Il apparaîtrait, par conséquent, que, sur le plan cognitif, la pratique de la musique, et ce même à un niveau de pratique important (MAS), reste une pratique qui demande une réflexion métacognitive liée à l'ensemble des paramètres qu'il faut maîtriser dans l'acte de jouer. L'oralisation de la L1 est un acte automatique qui plus est chez les locuteurs-plurilingues n'ayant pas une pratique régulière d'une langue étrangère. Seule la pratique orale d'une langue étrangère semble imposer chez les LPR et LPNR une réflexion métacognitive. Cette opposition se retrouve lorsque les musiciens montrent de l'importance dans le « *bien* » jouer pour la qualité du son contrairement aux locuteurs-plurilingues qui cherchent à transmettre du sens.

Cependant, les quatre panels (LPR/LPNR/MAS/MSS) montrent plus d'intérêt pour le fait d'être à l'aise même s'il y a des erreurs. La norme du « *bien jouer* » se trouverait donc chez le locuteur-plurilingue dans une interaction réussie tandis que le musicien chercherait à ce que sa perception interne et externe coïncide le plus possible tout en n'incluant pas le spectateur dans cette relation. Ce parallèle se retrouve lorsque les quatre panels ont l'impression lorsqu'ils changent de langue/style de musique d'être la même personne tout en agissant différemment avec les autres (interlocuteurs, spectateurs, etc.). Une minorité dans les quatre cas pense changer de personnalité. Les individus auraient donc une conscience stable de qui ils sont en tant que personne quel que soit le type de production qu'ils émettent et qu'ils souhaitent garder lorsqu'ils affirment vouloir garder leur identité personnelle lorsqu'ils changent de musique/langue. Cependant, les quatre panels montrent peu d'intérêt dans la conservation de leur identité culturelle et montrent une importance remarquable dans l'intégration des éléments culturels de la langue ou des spécificités d'un style de musique. Néanmoins, il est intéressant de constater qu'un nombre très important (80%) de musiciens (MAS et MSS) pensent changer d'état d'esprit lorsqu'ils changent de style de musique, contrairement aux locuteurs-plurilingues (LPR et LPNR) qui ne montrent pas de tendance probante. Il semblerait que la doxa proposant une co-construction de la pensée-« langue » favoriserait un continuum de distanciation dans la frontière entre les discours (les « langues ») (voir 2.3.1).

A l'opposé, la pratique musicale, quel que soit le type d'apprentissage, s'organiserait en espaces cognitifs cohérents pour chaque type de musique que le musicien peut produire. Les musiciens s'apparenteraient à des acteurs capables de « *jouer* » différents rôles éphémères sans que leur identité profonde soit modifiée. Nous faisons l'hypothèse que nous pourrions retrouver ce genre de tendance chez des locuteurs lors d'un changement de registre de langue dans des contextes

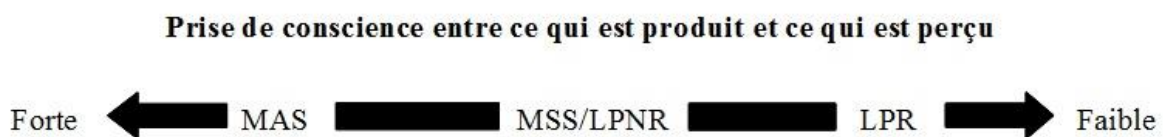
particuliers (situation académique *versus* relâchée, etc.).

Cette notion d'espace cognitif apparaît également lorsque les quatre panels présentent une tendance forte concernant des sensations différentes lorsqu'ils changent de langue ou de style de musique. Il semblerait que cette notion se développe au-delà de la sphère purement cognitive, mais se projette également sur la sphère physique, de manière rationalisée ou non, et permettant ainsi de catégoriser un réseau de caractéristiques propres à une performance donnée. Nous terminerons cette analyse intrapsychique en montrant qu'aucun des panels n'a montré de tendance remarquable concernant le fait de penser en même temps qu'ils parlent ou jouent de la musique. Il serait intéressant de décomposer cet aspect afin de voir dans une future étude si aucune tendance ne peut être mise en évidence ou si les facteurs individuels ne permettent pas de faire émerger de tendance importante.

Après avoir analysé les similitudes et différences entre les panels de musiciens et locuteurs-plurilingues, nous nous intéresserons aux plans **interpsychique** et **émotionnel** afin d'apporter un regard complémentaire. Nous pouvons tout d'abord constater que les quatre panels montrent une tendance dans une prise de conscience dans la différence entre ce qui est produit (par soi) et ce qui est perçu (par autrui). Cette prise de conscience est la plus forte chez les musiciens ayant fait du solfège puis comparable chez les MSS et LPNR et enfin plus faible chez les LPR (Tableau 38 et Figure 35).

	MAS		MSS	
	Tout à fait	Plutôt	Tout à fait	Plutôt
J'ai l'impression que ce que j'entends quand je joue n'est pas la même chose que ce qu'entendent les spectateurs.	30%	35%	14%	40%
	65%		54%	
	LPR		LPNR	
J'ai l'impression que ce que j'entends quand je parle n'est pas la même chose que ce qu'entendent mes interlocuteurs.	9%	27%	12%	32%
	36%		44%	

**Tableau 38:** Pourcentage de réponse des différents panels concernant la prise de conscience entre ce qui produit et ce qui est perçu.



**Figure 35 :** Continuum de prise de conscience entre ce qui est produit et ce qui est perçu chez les 4 panels étudiés.

Les MAS ont une pratique (orchestre, groupes semi-professionnel et amateur) plus variée que les MSS ce qui peut expliquer cette plus grande sensibilité aux distorsions du son dans l'espace. Les LPNR présentent une forte différence entre leur productions en langue(s) première(s) et les LE, de par leur niveau, ce qui implique une plus grande prise de conscience entre ce qu'ils aimeraient produire et ce qu'ils produisent en réalité. Cette considération peut également expliquer que les LPR aient la perception la moins sensible de cette différence puisque qu'ils peuvent valider plus facilement une interaction en LE et par conséquent les différences de forme sont minimales étant donné le caractère co-construit du sens dans l'interaction. Malgré une prise de conscience des différences entre soi et autrui, les musiciens montrent également peu d'intérêt à obtenir un retour positif des spectateurs dans le « *bien jouer* » tandis que les locuteurs-plurilingues veulent une fois encore valider l'interaction. Toutefois, les quatre panels s'accordent à montrer une forte tendance envers l'influence de la timidité sur la qualité de l'oral/de jeu (Tableau 39).

	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout	Tout à fait	Plutôt	Moyen	Plutôt pas	Pas du tout
	LPR					LPNR				
La timidité influence la qualité de l'oral.	<b>54%</b>	<b>30%</b>	8%	5%	3%	<b>64%</b>	<b>21%</b>	7%	6%	2%
	MAS					MSS				
La timidité influence la qualité du jeu.	<b>39%</b>	<b>37%</b>	12%	7%	4%	<b>42%</b>	<b>35%</b>	12%	6%	5%

**Tableau 39:** Résultats comparatifs en pourcentages de réponses au questionnaire à échelle d'attitude pour deux affirmations sur les effets de la timidité sur la qualité de l'oral/de jeu. En gras surligné gris : pourcentages  $X =$  ou  $> 50\%$  ; en gras : pourcentages  $30\% =$  ou  $< X$ .

Afin de terminer notre réflexion et après avoir porté un regard général, nous nous centrerons sur **les relations entre objets sonore et écrit**. Même si, de prime abord, la lecture aiderait plus les locuteurs-plurilingues à mieux prononcer que les musiciens à mieux jouer, dans leurs représentations, nous constatons que pour une très faible minorité les formes sonores et écrites fonctionnent de la même manière. En réalité, les quatre panels montrent une forte sensibilité sur les différences entre ces deux modalités, même si les locuteurs-plurilingues mettent l'accent sur le fait qu'ils fonctionnent différemment et les musiciens sur le fait qu'ils soient naturellement liés.

Néanmoins, les quatre panels semblent s'accorder sur le fait qu'il n'est pas plus difficile de lire un texte/une partition que de parler spontanément à l'oral en LE/improviser. Toutefois, pour une majorité d'entre eux, leur prononciation/jeu serait meilleur(e) quand ils lisent un texte/une partition. Les tendances sont plus mitigées lorsqu'il s'agit de savoir si la prononciation serait meilleure lorsqu'ils parlent spontanément à l'oral/improvisent (Tableau 40).

	LPR	LPNR	MAS	MSS
Il est plus difficile de lire un texte/une partition que parler spontanément à l'orale en LE/improviser	--	--	-	+/--
Ma prononciation/mon jeu est meilleur(e) quand je parle spontanément à l'oral/improviser	++/-	+	+/-	++/-
Ma prononciation/mon jeu est meilleur(e) quand je lis un texte/une partition	+	+	+	+/-
Lecture/partition aide mieux prononcer/jouer	+	++	+	+/-

**Tableau 40:** Synthèse des positionnements des quatre panels concernant les liens écrit-son dans un questionnaire à échelle d'attitude. Les signes correspondent à : (-) ou (+) tendance légèrement négative ou positive ; (- -) ou (++) tendance fortement négative ou positive ; (++/-) aucune tendance remarquable mais plutôt positive et (+/-) plutôt négative.

Ainsi, de manière générale, il semblerait que dans les représentations des sondés (musiciens et locuteurs-plurilingues) la forme écrite aiderait à la mise en son et n'imposerait pas de difficultés supplémentaires.

Selon tous ces éléments, nous pouvons dire que les rapports individu-son-écrit entre la parole et les locuteurs-plurilingues sont largement influencés par le caractère pragmatique des « langues ». Les musiciens ont un rapport plus introspectif en lien avec l'intentionnalité (explicite) émotionnelle que le jeu musical implique. Dans le contexte de notre étude, les variables externes étudiées ne semblent pas apporter de modification importante des représentations. Il nous reste donc à réfléchir sur l'influence d'un changement de médium de langage comme c'est le cas pour la langue des signes.

#### 4.3.4 Une modalité du signe ?

Après avoir mis en évidence, les similitudes et différences entre les langages verbal et musical qui comportent tous les deux, deux modalités d'expression : sonore et écrite, nous nous intéresserons au panel de locuteurs-signeurs afin de voir si le médium de communication (physique *versus* sonore) peut modifier les rapports au langage.

Nous devons avant tout noter, d'après le profil socio-culturel du panel de locuteurs-signeurs, que les sondés ont une approche de la langue des signes (LS) en tant que locuteurs tardifs puisqu'ils ont majoritairement découvert le langage par une langue orale puis une langue

des signes à l'âge de 15 et plus soit après la période critique (pour la langue orale : Flege *et al.*, 1999 ; Kuhl, 2010 ; Simmonds, 2011 ; pour la langue des signes : Newman, et al., 2002). Ils ont pour la grande majorité pris des cours formels de langue des signes et par conséquent ont pu développer un regard métacognitif sur cette « *langue* ». L'analyse permettra de mettre en évidence si des différences psychocognitives apparaissent dans la pratique d'une « *langue* » non-sonore même si les sondés n'ont pas pu développer des réseaux neurophysiologiques spécifiques de par leur apprentissage majoritairement tardif de la LS.

Nous focaliserons l'analyse sur les points traités précédemment, soit : le rôle du locuteur, les aspects intra- et interpsychiques, émotionnels et l'influence de l'écrit sur les représentations des formes signées.

Concernant le **rôle du locuteur**, nous pouvons constater que les locuteurs-signeurs répondent de la même manière que les locuteurs-plurilingues, soit en faveur d'une tendance vers la « *norme* » du locuteur natif (LS et LO) mais également en faveur de la transmission des émotions. Il semblerait donc qu'à ce niveau, la modalité ne change pas la relation à la « *norme* ». Si nous comparons ces résultats à ceux des musiciens, il semblerait que le caractère pragmatique spécifique des langues soit à l'origine de cette nécessité d'atteindre une norme tout en favorisant l'intentionnalité émotionnelle. Nous retrouvons cet élément lorsque le panel montre que le plus important pour bien communiquer, c'est d'être compris et à l'aise et ce malgré de possibles erreurs. Cette tendance est strictement similaire à celles constatées chez les LPR et LPNR.

Sur le plan **intrapsychique**, les locuteurs-signeurs montrent une forte tendance sur le fait de penser aux différentes caractéristiques de la langue des signes avant de signer des phrases d'un texte. Par contre, le fait d'y penser pendant ou avant l'acte de signer (de manière générale) ne représente qu'une légère tendance. Le panel montre donc une approche plutôt automatisée de l'acte de signer comme nous avons pu le voir chez les locuteurs-plurilingues ayant une pratique régulière en contexte naturel démontrant une approche peu métacognitive dans l'acte lui-même. Ce résultat paraît étrange étant donné le fait que la majorité du panel a appris sa LS de manière tardive et par des cours formels. Toutefois, en ce qui concerne l'acte de signer un texte, il semblerait que ce dernier demande une réflexion plus importante et moins automatisée probablement explicable par un coût cognitif important lié à un premier traitement graphème-phonème (Dehaene, 2007 : 69, entre autres) puis une transcription motrice en signe. Ce coût supplémentaire est en corrélation avec l'apprentissage tardif de la LS et donc un traitement auditif privilégié dans le traitement de l'écrit.

Notre panel de locuteurs-signeurs montre les mêmes tendances que les quatre précédents panels (LPR/LPNR/MAS/MSS) dans le fait de penser être les mêmes personnes tout en agissant

différemment avec les autres quand ils changent de langue (LO/LS). Ils ne montrent pas non plus de tendance particulière sur le fait d'être dans le même état d'esprit ou non lorsqu'ils changent de langue (LO/LS) comme nous avons pu le constater chez les locuteurs-plurilingues. Il semblerait donc que les individus aient une perception stable de la personne qu'ils sont et ce malgré le changement de langue (LO/LS) ou de style de musique. Tous perçoivent néanmoins le fait que ces changements modifient leur façon d'agir avec les autres. Mais seuls les musiciens ont une conscience importante d'un changement d'esprit lorsqu'ils changent de style de musique. Nous pouvons donc affirmer que la modalité physique de la langue des signes n'apporte pas de spécificité à celle sonore de la « *langue* ».

Sur le plan **interpsychique**, les locuteurs-signeurs montrent une tendance négative au fait qu'ils auraient l'impression que ce qu'ils signent n'est pas la même chose que ce que perçoivent leurs interlocuteurs. Cette tendance peut être expliquée par le caractère plus stable d'un stimulus visuel que celui d'un stimulus sonore ; ainsi les locuteurs-plurilingues et les musiciens perçoivent une « *vaporéité* » dans l'objet sonore ce qui implique une prise de conscience sur les distorsions probables entre la production et la perception tandis que le signe physique serait plus stable et donc enclin à moins de variation entre celui qui produit et celui qui perçoit. Néanmoins, l'influence de facteurs émotionnels tels que la timidité est reconnue par tous les panels pour les langages qu'ils soient oral, signé ou musical. Sur le plan culturel, nous ne remarquons pas de spécificité particulière car les quatre panels montrent une volonté, lorsqu'ils changent de « *langue* »/style de musique, d'intégrer les éléments spécifiques de la « *langue* »/musique cible, de garder leur identité personnelle et ce au détriment de leur identité culturelle.

Nous terminerons notre analyse en nous intéressant à l'influence de la modalité de communication sur **les rapports à l'écrit**. Nous pouvons remarquer tout d'abord que comme les quatre autres panels, les locuteurs-signeurs pensent que l'écriture et l'oral fonctionnent différemment même s'ils sont naturellement liés. Nous retrouvons les mêmes proportions qu'avec les locuteurs-plurilingues ce qui nous permet d'affirmer que l'apprentissage tardif d'une langue des signes ne modifie pas les relations entre l'écriture et l'oral chez ces personnes. Il est cependant intéressant de remarquer que contrairement aux quatre autres panels, une grande majorité des locuteurs-signeurs montrent une forte tendance négative en ce qui concerne le fait que la lecture les aiderait à mieux signer. Ce résultat est d'autant plus intéressant que les LPRN n'affichent pas cette tendance négative. Or les LPRN et les LS ont appris tardivement (10 ans et plus) leur(s) première(s) langue(s) étrangère(s) ou leur(s) première(s) langue(s) des signes. Nous pouvions, par conséquent, nous attendre à des résultats similaires. En réalité, le groupe



des musiciens n'ayant pas fait de solfège propose les résultats les plus semblables en n'affichant pas de tendance importante à savoir si la partition aide à mieux jouer. Nous pensons que cette corrélation est un épiphénomène dont l'origine serait distincte pour les panels LS et MSS.

Le panel de locuteurs-signeurs montre qu'il est plus difficile de signer des phrases d'un texte que de signer spontanément, qu'il n'est pas plus facile de lire des phrases d'un texte en signant que de les lire à l'oral et qu'ils ne communiquent pas mieux quand ils signent spontanément que quand ils parlent à l'oral spontanément. L'ensemble de ces données renforce l'analyse faite précédemment visant à dire que chez des locuteurs tardifs d'une langue des signes, le rapport entre graphème-signes est moins naturel que celui pour des langues orales (L1/LM ou LE) et par conséquent requiert un poids cognitif supérieur. Nous pouvons ajouter à cela que nous ne pouvons voir de tendance remarquable chez les locuteurs-signeurs sur le fait que leur prononciation serait meilleure quand ils lisent un texte à l'oral que quand ils parlent spontanément à l'oral.

#### **4.4 Conclusion générale de l'étude 1**

Le but principal de cette étude quantitative était d'apporter un éclairage multimodal en comparant les représentations sur plusieurs types de langages : langage verbal oral, des signes) et musical. Nous souhaitons mettre en évidence des liens ou ruptures afin de porter un regard critique des liens musique-parole dans une perspective psychosociale. Cette étude faisait l'objet de trois questionnements principaux auxquels nous répondrons point par point :

1. le type de langage verbal ou musical favorise-t-il l'émergence de représentations différentes sur le traitement de l'objet sonore et sa forme écrite ?

Les représentations des musiciens et des locuteurs-plurilingues respectivement pour les langages musical et verbal font apparaître de nombreuses différences que ce soit sur le rôle de celui qui « *met en son* », sur les rapports intra- et inter-psychique ou sur les liens forme orale/écrite. Ce que nous retenons principalement, c'est l'impact du prisme pragmatique pour le langage verbal et de l'intentionnalité émotive sur les représentations des sondés correspondants. En effet, l'objectif principal du locuteur-plurilingue est de valider des interactions tandis que le musicien a un rapport plus narcissique puisqu'il considère que le plus important pour bien jouer est d'être à l'aise malgré des erreurs ainsi qu'avoir un « *bon* » son. Le retour positif des spectateurs est secondaire, d'après les résultats obtenus dans notre questionnaire. Il semblerait qu'un « *bon* » locuteur est celui qui sait co-construire un échange dans une perspective

interpsychique tandis que le « *bon* » musicien serait celui qui, à partir d'une approche intrapsychique, pourrait solliciter autrui à travers une intentionnalité émotionnelle efficace.

2. des pratiques différentes (solfège, pratique régulière) favorisent-elles le renforcement de certaines représentations sur les formes écrite et sonore ?

De manière générale, nous n'avons pas observé de tendances importantes entre les panels de musiciens (MAS et MSS) et les locuteurs-plurilingues (LPR et LPNR). Il semblerait que le suivi des cours de solfège et la pratique régulière d'au moins une langue étrangère en contexte naturel, respectivement chez les musiciens et les locuteurs-plurilingues, ne modifient pas de manière importante les représentations des sondés. Néanmoins, nous avons pu voir que ces variables externes peuvent être à l'origine de pratiques différentes qui vont modifier certaines représentations. Les musiciens ayant fait du solfège ont en général commencé plus tôt et ont des pratiques plus variées ayant un impact notamment métacognitif en lien à la partition. Nous avons également remarqué que les individus pratiquant régulièrement au moins une langue étrangère dans un contexte naturel ont en moyenne commencé à apprendre les langues étrangères de manière plus précoce. Contrairement aux musiciens, cette donnée inhibe l'approche métacognitive de l'objet oral en LE en faveur d'un traitement plus implicite et automatique qui pourrait favoriser un traitement plus rapide de l'information verbale.

3. la modalité de langage (langue des signes) influe-t-elle sur les représentations des locuteurs ?

Nous avons constaté que la pratique d'au moins une langue des signes ne modifie pas les représentations des individus sur leur rapport au langage oral. Ce point peut être expliqué par le fait que la majorité des sondés sont des locuteurs-signeurs tardifs. Néanmoins, nous avons pu mettre en évidence que la modalité physique du signe rend la forme écrite plus éloignée de la première. Contrairement aux individus ne pratiquant pas de LE en contexte naturel qui pensent qu'il est plus facile de lire un texte en LE que de parler spontanément, les locuteurs-signeurs pensent qu'il est plus difficile de signer un texte que de signer spontanément. La modalité signée imposerait donc chez des apprenants tardifs des contraintes spécifiques à la modalité physique du signe.

### **Synthèse de positionnement**

Cette étude a été la première à être lancée dans ce travail de thèse. Elle marque le prolongement d'une étude sur un nombre de répondants très limité menée lors de mon mémoire de master 2. Elle a été le point de départ de cette réflexion sur les liens entre musique et parole afin de mettre en évidence des tendances fortes dans le lien qui unit locuteurs-plurilingues et langage verbal ainsi que musiciens et langage musical. Le grand nombre de répondants contrebalance le fait que ce type de questionnaire ne mesure que des représentations et non des pratiques réelles puisque cela permet d'avoir une représentativité suffisante pour émettre des conclusions. Néanmoins, elle permet de montrer comment les caractéristiques de l'objet (verbal ou musical) peut impacter ou non la relation que l'individu lui porte indépendamment du fait qu'elles utilisent le même médium de transmission : le son. Cette étude a permis, notamment, d'initier une réflexion sur la focalisation des individus sur les aspects sonores et non les éléments sémantiques de la parole. Dans cette thèse, il est important de voir qu'étant donné le caractère interactionnel majeur du langage verbal, l'attention des individus se porte majoritairement sur la validation des interactions plutôt que les modalités de production/perception. A ce titre, il est intéressant de constater que les locuteurs-plurilingues qui pratiquent au moins une langue en contexte naturel sont ceux qui montrent la tendance la plus faible dans la prise de conscience entre ce qui est produit et ce qui est perçu par autrui. Plus le système langagier devient automatique, moins d'attention est portée sur la forme. C'est à ce niveau que cette étude nous amène vers une prise de conscience de l'intérêt de la transdisciplinarité qui permet de considérer le domaine dans lequel on se place à travers le prisme d'un autre domaine connexe. En outre, ici, la didactique de la musique permet d'être une source d'inspiration afin de penser à de nouvelles possibilités en didactique de la prononciation. Ces aspects permettraient aux apprenants de focaliser sur le son et leur performance en développant leur perception proprioceptive dans le but de déconstruire le système de la/les L1(s) vers un nouveau système articulaire (cf. les paramètres articulatoires ; Honikman (1964)). Cette étude m'a permis de faire deux publications l'une sur la didactique de la prononciation (Miras, 2013) et l'autre sur les représentations de la lecture-écriture (Miras, à paraître).

## 5. Etude 2 : expérimentation psychoacoustique

### Résumé

Cette étude psychoacoustique a pour but de contribuer aux connaissances sur les processus du traitement de la parole afin de confirmer ou infirmer les travaux en neurocognition actuels qui tendent vers l'idée que la pratique musicale aurait un impact sur les capacités auditives (Peretz & Coltheart, 2003 ; Chobert & Besson, 2012 ; Kraus & Chandrasekaran, 2010 ; entre autres). Nous chercherons à déterminer si une tâche d'écoute musicale instrumentale peut favoriser un traitement acoustique des unités sonores de la parole et donc permettre un traitement perceptif plus rapide des informations acoustiques de stimuli vocaliques au niveau segmental chez des apprenants du français.

Pour ce faire, nous avons mis en place un test de discrimination AXB avec un intervalle interstimulus (ISI) égal à 200 ms afin de tester le traitement acoustique des unités de la parole (Pisoni & Tash, 1974). Nous nous focalisons sur sept paires de voyelles articulatoirement proches : /i/-/e/, /e/-/ɛ/, /y/-/u/, /ø/-/œ/, /o/-/ɔ/, /ɔ/-/a/. Nous montrerons la pertinence des stimuli grâce à leur analyse acoustique et à un test de perception avec 8 auditeurs francophones natifs sur ces stimuli. Au total, nous avons obtenu 112 triplets pour chaque test sur la base de 7 paires vocaliques x 4 répartitions x 4 répétitions.

Chaque apprenant du français passe deux tests de discrimination AXB ordonnés différemment de manière semi-aléatoire. Entre les deux tests, chaque sujet a une tâche parmi trois possibles qui consistent en une exposition, soit à un document visuel, soit un document audio de parole en français, soit un document audio de musique instrumentale suivi d'un court questionnaire. Les trois documents ont été sélectionnés à l'aide de deux pré-études sous la forme d'un questionnaire en ligne rempli par des musiciens, non-musiciens, pour la première et des experts en art, pour la deuxième. Ces deux pré-études permettent de sélectionner trois documents perçus comme proches en termes de tempo et d'impact émotionnel afin de limiter tout biais lié à des changements physiologiques pouvant conduire à des changements cognitifs (Bernardi *et al.* 2006). Pour terminer, nous avons pu retenir trois documents (visuel, de parole en français et de musique instrumentale) qui suggèrent un tempo modéré et un impact émotionnel lié à des termes comme « *calme, paisible, serein et monotone* ».

L'étude concerne 40 locuteurs francophones dont 30 non-natifs du français. Ces derniers ont été regroupés en trois groupes de dix sujets en fonction des trois tâches possibles. Pour chaque groupe, nous avons retenu cinq musiciens et cinq non-musiciens. Afin d'avoir un point de

référence, dix sujets locuteurs natifs du français n'ont été confrontés qu'au document de musique instrumentale. Ce groupe comprend également cinq musiciens et cinq non-musiciens. L'ensemble des 30 locuteurs non-natifs du français était des apprenants du français lors de la passation du test dans des groupes de niveau A2/B1.

Les données ont été analysées de manière statistique en prenant en compte le taux de discrimination d'une part et le temps de réaction d'autre part. En ce qui concerne l'analyse des données parmi les apprenants du français, nous avons pu montrer un faible taux d'erreur moyen, soit moins de 2%. Aucun effet significatif n'a pu être démontré à l'exception d'un effet marginalement significatif ( $F(1,24)=4,015$ ,  $p=0,057$  ;  $F(1,6)=64,751$ ,  $p=0,036$ ) du statut musical avec un meilleur score moyen des musiciens de 1,76%. Concernant le temps de réaction, nous avons pu montrer deux effets significatifs pour la mesure répétée du Test ( $F(1,24)=12,284$ ,  $p=0,002$  ;  $F(1,6)=233,315$ ,  $p<0,0001$ ) avec une diminution du temps de réaction de 85,11 ms lors du deuxième test (Test 1= 355,80 ms et Test 2= 270,69 ms) et de l'effet principal du Statut ( $F(1,24)=4,381$ ,  $p=0,047$  ;  $F(1,6)=324,149$ ,  $p<0,0001$ ) avec un meilleur score par les musiciens de 148,12 ms (Musiciens= 239,19 ms et Non-Musiciens= 387,31 ms). Nous constatons également une interaction marginalement significative Test X Statut ( $F(1,24)=3,297$ ,  $p=0,082$  ;  $F(1,6)=16,637$ ,  $p=0,007$ ) d'après laquelle, les musiciens répondent d'autant plus rapidement que les non-musiciens lors du test 2 (164,44ms) que lors du test 1 (107,14ms).

Lorsque nous comparons les résultats des apprenants et des locuteurs natifs du français, nous constatons qu'il n'y a pas d'effet significatif en ce qui concerne le taux de discrimination selon les différents facteurs étudiés. Nous constatons notamment que le taux moyen d'erreur est de moins de 2% et qu'il n'y a pas de différence significative entre les apprenants et les locuteurs natifs du français. Au regard du temps de réaction, nous avons retrouvé les effets significatifs démontrés lors de l'étude avec les apprenants seuls, soit un effet significatif pour la mesure répétée du Test ( $F(1,16)=22,732$ ,  $p<0,0001$  ;  $F(1,6)=204,656$ ,  $p<0,0001$ ) avec une diminution du temps de réaction de 98,61 ms lors du deuxième test (Test 1 = 325,43 ms et Test 2 = 226,81 ms) et de l'effet principal du statut musical ( $F(1,16)=7,242$ ,  $p=0,016$  ;  $F(1,6)=324,990$ ,  $p<0,0001$ ) avec un meilleur score par les musiciens de 161,39 ms (Musiciens = 195,43 ms et Non-Musiciens = 356,81ms). Nous avons également pu remarquer une interaction significative Test X Statut musical ( $F(1,16)=9,691$ ,  $p=0,007$  ;  $F(1,6)=38,221$ ,  $p=0,001$ ) ce qui montre que les musiciens sont meilleurs lors du test 2 qu'ils

ne l'étaient lors du test 1 par rapport aux non-musiciens.

En conclusion, nous avons pu retrouver au niveau comportemental les différences entre musiciens et non-musiciens décrites dans les travaux actuels en neurocognition susmentionnés. Nous avons pu montrer, en outre, que les musiciens répondaient significativement plus rapidement que les non-musiciens à un test de discrimination AXB d'oppositions fines de phonèmes vocaliques du français et ce que ce soit chez des apprenants ou des locuteurs natifs du français. Nous avons également constaté, chez l'ensemble des groupes, une diminution significative du temps de réaction entre les tests 1 et 2 ce qui va dans le sens des travaux de Bygate (2001) et Bei (2013) sur une progression qualitative et quantitative des résultats par simple répétition de la tâche. Au regard du taux de discrimination, nous n'avons pu montrer aucun effet significatif à l'exception d'un effet marginalement significatif en faveur des musiciens parmi les apprenants de français. Nous avons également constaté un faible taux d'erreur moyen de moins de 2%. Ce point va dans le sens des travaux selon lesquels la nativisation/assimilation phonologique (le « crible phonologique ») ne prendrait place que de manière descendante mais que le traitement acoustique ne subirait pas directement l'influence de la/les L1(s). Pour terminer, nous n'avons pu montrer aucun effet significatif parmi les trois groupes d'apprenants du français. Cela montre que l'influence de la musique instrumentale ne semble avoir d'impact qu'à travers une période de pratique importante permettant d'amorcer des changements sur le plan neurocognitif. Cet aspect fait écho à la notion d'expertise (Gladwell, 2008). Il semblerait donc de manière générale que l'on puisse mesurer un impact de la musique instrumentale sur les capacités auditives d'unités langagières ; néanmoins, cet impact n'apparaît qu'à travers une pratique à long terme et non un temps d'exposition court.

## 5.1 Méthodologie et questions de recherche

L'ensemble des travaux qui portent sur les liens entre musique et parole ne permet pas actuellement de prendre une position claire étant donné la complexité de la nature de ces derniers. Comme nous avons pu le voir dans la section 1.2, les travaux en neurocognition tendent plutôt vers un traitement partiellement modulaire (Peretz & Coltheart, 2003) mais non encapsulé (Chobert & Besson, 2012), de la musique et la parole. Néanmoins, Chobert & Besson (2012 : 71) ont notamment montré que « *l'apprentissage musical ne facilite pas seulement le traitement des paramètres acoustiques tel que la durée ou la fréquence, mais améliore également la perception de paramètres linguistiques plus abstraits tel que le VOT* ». Kraus & Chandrasekaran (2010 : 599) ont décrit « *des résultats qui vont en faveur du fait que les capacités fines auditives des musiciens, qui sont acquises au fil des années d'entraînement, ont des répercussions dans d'autres domaines comme la parole, le langage, les émotions et les processus auditifs* »<sup>69</sup>. Cependant, l'ensemble de ces travaux porte majoritairement soit sur un entraînement de plusieurs mois chez des enfants soit chez des adultes ayant pratiqué la musique pendant des années. Peu d'études portent sur l'influence de la musique sur les capacités auditives chez des adultes musiciens ou non avec une durée d'exposition limitée dans le temps. Nous savons toutefois (voir section 1.2.2) que, selon Bigand (2004 : 215), l'apprentissage implicite de la musique chez des auditeurs sans formation musicale provoquerait des réorganisations cérébrales plus massives que l'apprentissage explicite. A l'opposé de ce paradigme, nous constatons l'existence de pratiques pédagogiques comme la méthode Speedlingua© (section 3.3.1) qui promet qu'une écoute musicale soutiendra la médiation de la prononciation, hypothèse qui n'est soutenue par aucune donnée empirique.

Notre étude aura donc pour but de mettre en corrélation des résultats en neurocognition sur l'impact de l'entraînement musical sur les capacités auditives avec le paradigme de la didactique des langues dans lequel, les groupes sont souvent hétérogènes (musiciens, non-musiciens, auditeurs expérimentés, auditeurs amateurs), les contraintes de temps importantes (formations de quelques heures, semaines, mois) et les pratiques pédagogiques peu confrontées à des protocoles expérimentaux. Nous tenterons principalement de savoir si une tâche d'écoute musicale instrumentale peut favoriser un traitement acoustique des unités sonores de la parole

---

69 « *we describe data that support the view that the fine-grained auditory skills of musicians, which are acquired through years of training, percolate to other domains, such as speech, language, emotion and auditory processing* »

et donc permettre un traitement perceptif plus rapide des informations acoustiques de stimuli vocaliques au niveau segmental. Notre hypothèse est qu'une tâche musicale pourrait avoir, à court terme, un effet d'amorçage cognitif vers une focalisation pour un traitement acoustique plus fin et donc limiter l'impact du traitement phonologique et par conséquent le processus de nativisation/assimilation. De ce fait, les auditeurs auraient une meilleure perception des caractéristiques acoustiques des unités phonétiques proposées. Nos questionnements expérimentaux seront les suivants :

1. Existe-t-il des différences significatives entre musiciens et non-musiciens dans une tâche de discrimination AXB des phonèmes vocaliques du français en ce qui concerne le temps de réaction et le taux de discrimination ?
2. Une exposition à une musique instrumentale peut-elle favoriser le traitement des phonèmes vocaliques du français au niveau acoustique?
3. Existe-t-il une différence de traitement entre apprenants et locuteurs natifs du français?

Pour ce faire, nous définirons tout d'abord notre protocole expérimental issu de la méthodologie en phonétique-phonologie à travers le matériel et les méthodes employés. Nous définirons par la suite la modalité de passation du test : le cadre, les participants et les variables, entre autres. Nous terminerons par l'interprétation et l'analyse des résultats.

Nous souhaitons préciser que le caractère pluri-inter-transdisciplinaire de notre thèse nous a poussé à faire des compromis expérimentaux en faveur d'un domaine ou d'un autre. En effet, certains outils phonétiques ne permettent pas de prendre directement en compte le paradigme expérimental qui peut s'appliquer actuellement en didactique. Nous assumons donc que cette étude se placera du côté de la phonétique ou de la linguistique appliquée, à travers une analyse réductionniste, mais qu'elle nous permettra d'envisager des applications et d'autres protocoles de recherche plus en adéquation avec la didactique des langues.

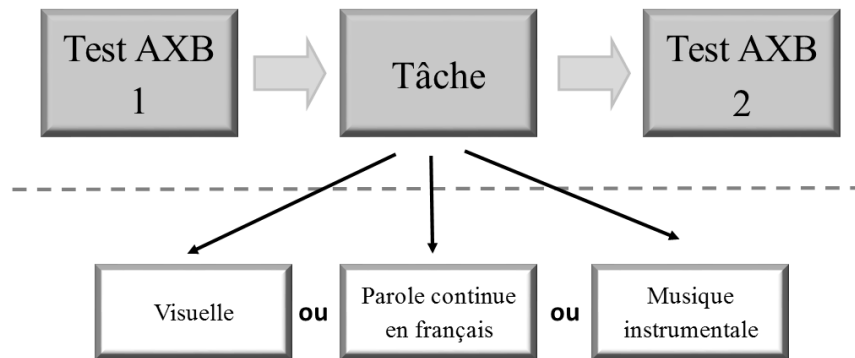
### **5.1.1 Matériel et Méthodes**

Nous présenterons dans cette section le déroulement général du test à travers les différentes étapes qui le composent : le questionnaire, les tests AXB et les tâches. Nous verrons par la suite les résultats aux différentes pré-études mises en place afin de choisir les documents utilisés comme tâches. Nous mentionnons dès à présent que l'ensemble des documents sonores sont disponibles dans un dossier partagé *Googledrive* en ligne (voir partie « documents sonores » de l'étude 2 en annexes)



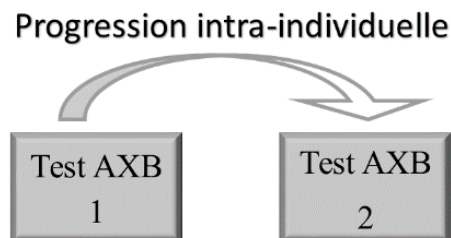
### 5.1.1.1 Déroulement général du test

Le test général se compose de trois grandes étapes qui sont la passation d'un premier test AXB (test 1) puis d'une tâche (parmi trois tâches possibles) et enfin d'un deuxième test AXB (test 2) (Figure 36). Un questionnaire, passé au préalable, permet de cibler chaque individu dans un groupe précis dont nous définirons les caractéristiques ultérieurement.



**Figure 36:** Déroulement des trois phrases principales du test général.

Chaque répondant doit, tout d'abord, passer un premier test de discrimination AXB que nous nommerons test 1. Il passera ensuite par une des trois tâches : soit visuelle, soit de parole continue en français ou soit de musique instrumentale. Pour finir le sujet repassera un test de discrimination AXB que nous nommerons test 2.



**Figure 37:** Mesure de la progression entre les tests 1 et 2 afin de mettre en évidence la progression intra-individuelle.

L'analyse finale des données de la recherche portera sur la mesure pour chaque répondant de la progression de son score entre les tests 1 et 2 (Figure 37).

### 5.1.1.2 Les auditeurs

Les auditeurs, de cette étude, sont répartis en deux catégories que nous présenterons successivement (Figure 38) :

- des apprenants du français comme langue étrangère/seconde,
- des locuteurs natifs du français comme langue première et/ou langue de scolarisation et/ou de première socialisation et/ou familiale.



**Figure 38:** Les différents groupes d'auditeurs retenus pour l'étude. Les tâches sont une exposition à un document visuel (Visuelle) ou de parole continue en français (Parole FR) ou de musique instrumentale (Musique).

Chaque sous-groupe est composé de 5 sujets soit au total 10 sujets par groupe (tâche). Au final, nous avons sollicité la participation de 30 apprenants du français (15 musiciens et 15 non-musiciens) et 10 locuteurs « natifs » (5 musiciens et 5 non-musiciens).

Les apprenants du français ont été recrutés dans différentes institutions: collaboration avec l'Alliance française, diplôme universitaire d'étude française ou cours d'hiver de l'Université Sorbonne Nouvelle – Paris 3. Une affichette (annexe 26) a été diffusée afin d'obtenir le retour de personnes volontaires. Le profil des auditeurs a été dressé à l'aide d'un questionnaire Googledocument© (annexe 27) qui a systématiquement été rempli au moins une semaine avant la passation du test. Le questionnaire était disponible en français ou en anglais afin d'être sûr que chaque apprenant puisse comprendre les consignes s'il ne se sentait pas totalement à l'aise avec le français. Nous commencerons par spécifier les profils socio-culturels et langagiers généraux des apprenants du français (l'ensemble des réponses est donné en annexe 28) :

- **Nationalités** : les nationalités sont variées et représentent la diversité des apprenants du français en France. Les nationalités les plus représentées sont : russe (5), américaine et brésilienne (4 sujets chacune), puis colombienne (3), ukrainienne, australienne (2) et enfin avec une occurrence par nationalité : chinoise, iranienne, syrienne, argentine, suédoise, péruvienne, espagnole, irlandaise, taiwanaise, géorgienne et chilienne.
- **Genre** : 24 sujets se déclarent être une femme et 6 un homme.
- **Lieu de vie** : la majorité déclare habiter actuellement en France (26/30) même si l'intégralité a séjourné en France lors de la passation du test.
- **Age** : la plupart des sujets avaient entre 20 et 50 ans (23/30), une minorité avait moins de 20 ans (2/30) ou plus de 50 ans (5/30).
- **Lieu et langue principale de scolarisation** : nous ne pouvons traiter les données liées à ces deux questions de notre formulaire à cause d'une ambiguïté dans la formulation ayant provoqué des discordances dans les réponses. Un certain nombre de répondants a notamment mentionné le français comme langue principale de scolarisation car ils ont principalement étudié le français à l'université en tant que contenu mais pas forcément comme langue d'enseignement.

- **Profession** : les sujets ont des professions variées même si nous constatons que la majorité provient de classes moyennes supérieures à élevées.
- **Langues premières** : les principales langues premières mentionnées sont l'espagnol (7/30), l'anglais (5/30), le portugais (4/30), le russe (4/30) puis le chinois (2/30), l'ukrainien (2/30) et enfin avec une occurrence par langue : le japonais, l'arabe, le suédois, le catalan, le géorgien et le persan. Au total, nous avons 13 langues mentionnées en tant que langues premières.
- **Plurilinguisme** : les sujets de cette étude sont en majorité des locuteurs plurilingues dans le sens où l'intégralité a appris au moins une langue étrangère (LE1 : 30/30), où une grande majorité a appris une deuxième langue étrangère (LE2) (28/30), la moitié une troisième langue étrangère (LE3) (15/30) et une minorité une quatrième langue étrangère (LE4) (4/30). La majorité a appris sa première langue étrangère entre 5 et 15 ans (21/30).
- **Langues étrangères apprises** : les deux langues étrangères les plus mentionnées sont le français (30/30) et l'anglais (23/30) qui est majoritairement appris en LE1 (19/30). Au total, nous avons 12 langues mentionnées en tant que langues étrangères apprises.
- **Le français** : de manière générale, le français est la deuxième ou troisième langue étrangère apprise (24/30). Ils ont principalement appris le français à l'âge de 10 ans et plus (28/30). Ils ont appris le français de manière formelle en prenant des cours (22/30). Ils ont tous été en France pour des courtes (24/30) ou longues (17/30) durées.

L'ensemble des sujets était encore lors de l'étude des apprenants du français dans une structure institutionnelle. Nous constatons, de manière générale, que l'intégralité des apprenants suit actuellement une formation de/en français dans des groupes de niveau CECRL de A2 à B2. Ces niveaux sont informatifs car si l'Alliance français détermine le niveau de ses apprenants d'après un test de placement, le DUEF et les cours d'hiver à l'Université Sorbonne Nouvelle – Paris 3 ne garantissent qu'un niveau minimum pour avoir accès à l'inscription. A travers le questionnaire mené, nous avons pu compléter ce point au regard de 16 personnes ayant passé une certification du français au cours de deux dernières années. Nous constatons que la majorité de ces derniers a un niveau B1/B2 (11/16) et une minorité un niveau A2 (1/16) ou C1-C2 (3/16). Pour aller plus loin, nous avons demandé aux sujets de s'autoévaluer sur une échelle de 1 (très faible) à 7 (excellent) sur trois points en français : l'écrit, l'oral et la prononciation (Tableau 41).

		Groupe 1		Groupe 2		Groupe 3	
		Tache Visuelle		Tache Parole FR		Tache Musicque	
		M	NM	M	NM	M	NM
Ecrit	Moyenne (écart type)	3,6 (1,1)	4,2 (1,3)	4 (0,7)	3,8 (1,1)	3,6 (1,5)	3,6 (1,1)
		3,9 (1,2)		3,9 (0,9)		3,6 (1,0)	
Oral		4,4 (1,1)	4,2 (1,5)	4,2 (1,1)	3,8 (1,5)	4,4 (1,0)	4,4 (1,8)
		4,3 (1,2)		4 (1,2)		4,4 (1,5)	
Prononciation		4,2 (1,1)	4 (1)	4,6 (1,1)	3,8 (1,1)	4,2 (0,8)	4,4 (2,1)
		4,1 (1)		4,2 (1,1)		4,3 (1,7)	

**Tableau 41:** Moyennes et écart type obtenues dans chaque groupe sur l'autoévaluation des sujets de leur niveau de français écrit, oral et de prononciation sur une échelle de 1 (très faible) à 7 (excellent). M : musiciens ; NM : non musiciens.

Nous remarquons qu'en moyenne, les différents groupes montrent des profils d'auto-évaluation similaires allant de 3,6 à 4,6 sur 7. En ce qui concerne la prononciation, les moyennes sont proches pour les trois groupes et comprises entre 4,1 et 4,3 sachant que l'écart-type moyen est de 1,3. Les musiciens des groupes 1 et 2 se sont autoévalués avec un niveau supérieur de 0,5 en moyenne par rapport aux non-musiciens. Cette tendance est inverse pour le groupe 3 avec une différence de 0,2 au profit des non-musiciens.

**Profil général :** au regard de ces premières informations, nous pouvons dire que les trente sujets sont de nationalités, langues premières, d'âges et de genre différents même si nous constatons une majorité de femmes. La majorité provient de la catégorie socio-professionnelle des classes moyennes supérieures. Ils ont tous vécu en France pour au moins un court séjour. Ce sont tous des locuteurs plurilingues d'au moins 3 langues (L1 + 2 LE). Le français est majoritairement la 2<sup>ème</sup> ou 3<sup>ème</sup> langue étrangère apprise et ce de manière tardive (10 ans et plus). La majorité a appris le français par des cours formels de langue. Cependant, leur niveau paraît homogène.

Nous poursuivons notre description par celle du profil musical des apprenants du français :

- **Place de la musique :** plus de la moitié des sujets déclarent que la musique a une place importante dans leur vie (19/30).
- **Ecouter de la musique :** la plupart des sujets écoutent de la musique régulièrement soit tous les jours (18/30) soit deux à trois fois par semaine (10/30).
- **Type de musique écouté :** le style le plus écouté est la musique pop (21/79 occurrences) puis le classique (14/79), le rock (13). En tout, 18 styles ont été mentionnés.

Au début du test, nous souhaitons dissocier les musiciens et non-musiciens à partir d'une réponse déclarative à l'affirmation « je me considère comme ». Cependant, au regard, des

premiers retours, nous nous sommes rendu compte que des personnes ayant joué / qui jouent de la musique pendant / depuis des années ne se déclaraient pas forcément comme « musicienne ». Nous avons donc choisi de dissocier ces deux groupes à travers l'affirmation : « je fais / j'ai fait de la musique (instrument, chant, etc.) » à laquelle nous avons obtenu 15 « oui » et 15 « non ». Nous allons présenter maintenant le profil de notre cohorte de musiciens :

- **Nombre d'années de pratique** : une légère majorité a moins de 10 ans de pratique (9/15) et l'autre moitié a plus de 10 ans de pratique (6/15).
- **Age de début de pratique** : la moitié des musiciens sont des musiciens précoces et ont commencé la pratique de la musique à l'âge de moins de 10 ans (7/15) et l'autre moitié des musiciens tardifs en ayant commencé la musique après 10 ans (8/15).
- **Rythme de pratique** : la majorité des musiciens sont des musiciens occasionnels et jouent une fois par mois ou moins (11/15). Une minorité est des musiciens réguliers (tous les jours ou deux fois par semaine) (4/15).
- **Instruments joués** : les instruments joués les plus mentionnés sont le piano (6 occurrences), le chant (5), la guitare (2) et enfin le violoncelle (2). En tout, neuf instruments ont été mentionnés.
- **Type d'apprentissage** : la moitié (7/15) a appris la musique avec des cours formels et l'autre non (8/15).

**Profils des musiciens** : de manière générale, les musiciens de notre cohorte ont des profils variés avec des musiciens précoces ou tardifs et un apprentissage formel ou non. Cependant, une majorité est constituée de musiciens occasionnels.

Nous souhaitons mentionner, dès à présent, que compte tenu d'une volonté épistémologique de vérifier la possibilité de valider des résultats obtenus dans des paradigmes expérimentaux très contrôlés en neurocognition, dans une étude comportementale avec des sujets plus hétérogènes correspondant à la réalité de la « classe », nous n'avons pas tenté de contrôler plus nettement le profil de la cohorte de musiciens. Cependant, afin d'être en mesure d'analyser tout biais quasi-expérimental dû au profil des sujets, nous allons porter un regard intergroupe des apprenants du français sur cinq points pour l'ensemble des groupes : le niveau de français, l'âge d'apprentissage de la première LE, l'âge de premier apprentissage du français, la fréquence d'écoute de musique ; et pour les musiciens : le nombre d'années de pratique, l'âge de début de pratique de la musique et la fréquence de cette pratique. Ces différents points sont susceptibles d'avoir un impact les capacités de discrimination des sons de langues étrangères.

Institution	Formation	Niveau CECRL	Groupe 1		Groupe 2		Groupe 3		TT
			M	NM	M	NM	M	NM	
Alliance française		A2	1	1	1	1	1	1	6
		B2	1	1	1	1	1		5
Université Paris 3	DUEF	B1/B2	2	2	2	2	2	3	14
	Cours d'hiver	A2					1	1	2
		B1/B2	1	1	1	1			4
<b>Total</b>			5	5	5	5	5	5	30

**Tableau 42:** Effectifs des apprenants en fonction des institutions dans lesquelles ils suivent une formation de français. M : musiciens ; NM : non musiciens.

Comme nous le constatons à travers le Tableau 42, les groupes sont équilibrés en ce qui concerne les niveaux en français d'après le niveau théorique des formations suivies. Il ne semblerait pas que le niveau de français des auditeurs puisse provoquer un biais dans les analyses statistiques.

Age	Apprentissage première langue étrangère							Premier apprentissage du français							
	1M	1NM	2M	2NM	3M	3NM	TT	1M	1NM	2M	2NM	3M	3NM	TT	
0-5					3	1	4								0
5-10	2	2	3	1	1	1	10				1		1		2
10-15	1	2	1	4	1	2	11	1	1	1	1	1			5
15 et +	2	1	1			1	5	4	4	4	3	4	4		23
Total	5	5	5	5	5	5	30	5	5	5	5	5	5	5	30

**Tableau 43 :** Effectifs des apprenants en fonction de l'âge d'apprentissage de la première langue étrangère (à gauche) et de premier apprentissage du français (à droite). M : musiciens ; NM : non musiciens.

Au regard du Tableau 43, nous pouvons constater que de manière générale, les différents groupes sont semblables en termes d'âge d'apprentissage de la première LE et celui de premier apprentissage du français. Nous pouvons cependant noter que le groupe 3 de musiciens (3M) se démarque par le fait qu'une majorité d'auditeurs a appris leur première langue étrangère de manière précoce contrairement aux autres groupes

**Biais langagier intergroupe possible :** d'après cette première analyse des profils des auditeurs, il semblerait qu'un biais en rapport au niveau hypothétique en français soit peu probable étant donné que les groupes sont également répartis. Cependant, concernant le profil linguistique des trois groupes, il est envisageable que le groupe 3 ait de meilleurs résultats étant donné que les musiciens de ce groupe (3M) sont des locuteurs plurilingues plus précoces que les autres.

Concernant le profil musical des auditeurs, il ne semble pas y avoir de différences majeures étant donné que dans les 3 groupes, les sujets sont des auditeurs réguliers de musique (Tableau 44).

Fréquence d'écoute de la musique	1M	1NM	2M	2NM	3M	3NM	TT
Tous les jours	2	4	3	1	4	4	18
Deux à trois fois par semaine	2		2	4	1	1	10
Une à deux fois par mois							
Moins d'une fois par mois	1	1					2
Total	5	5	5	5	5	5	30

**Tableau 44** : Effectifs des apprenants en fonction de la fréquence d'écoute de musique

Ans	Nb d'années de pratique				Age de début de pratique				Fréquence	Rythme de pratique			
	1M	2M	3M	TT	1M	2M	3M	TT		1M	2M	3M	TT
0 à 5	1	2	2	5			1	1	Tous les jours	1		1	2
5 à 10	2	2		4	2	1	3	6	Deux fois par semaine	1		1	2
10 à 15	1	1	1	3	1	2		3	une fois par mois	1	1		2
15 et +	1		2	3	2	2	1	5	moins d'une fois par mois	2	4	3	9

**Tableau 45** : Effectifs des apprenants en fonction du nombre d'années de pratique d'un instrument (à gauche), de l'âge de début de pratique (au centre) et la fréquence de pratique (à droite).

En ce qui concerne le profil spécifique des musiciens (Tableau 45), il semblerait qu'il y ait une homogénéité entre les trois groupes pour le nombre d'années de pratique. Toutefois, le groupe 3 aurait commencé la musique à un âge plus précoce que les deux autres groupes et le groupe 2 serait composé majoritairement de musiciens très occasionnels contrairement aux groupes 1 et 3.

**Biais musical intergroupe possible** : il semble peu probable qu'un biais sur l'expertise d'écoute musicale puisse être en jeu puisque les 3 groupes montrent une homogénéité. Cependant, le groupe 3 pourrait montrer de meilleurs résultats car il est composé de plus de musiciens précoces que les autres groupes contrairement au groupe 2 composé d'une majorité de musiciens très occasionnels.

Nous terminerons par la description du profil des auditeurs natifs du français :

- **Nationalités** : l'ensemble des dix auditeurs natifs du français sont de nationalité française (dont une bi-nationalité), vivent en France et ont suivi principalement leurs études en France et en français.

- **Niveau académique** : ils ont tous un niveau académique équivalent au Baccalauréat ou supérieur (dont 2 niveau licence et 2 niveau master)
- **Profession** : 5 sujets se déclarent étudiant-e, 2 enseignant-es, 1 manager, 1 sans emploi et 1 en congé de formation.
- **Langue première** : l'ensemble des dix auditeurs est de langue première française (dont un bilingue libanais).
- **Langues étrangères** : 9/10 a l'anglais pour première langue étrangère et un a déclaré le basque. L'espagnol et l'allemand sont les deux deuxièmes langues étrangères apprises (respectivement 4 occurrences) puis l'anglais et le latin (1 occurrence). 6 sujets déclarent avoir trois langues étrangères dont le japonais, le chinois, l'espagnol, le latin, l'italien et le portugais. 4 déclarent avoir quatre langues étrangères dont le chinois, le vietnamien, l'espagnol et le créole réunionnais.
- **Age apprentissage première langue étrangère** : l'intégralité a appris sa première langue étrangère entre l'âge de 5 et 15 ans (dont 4 entre 5-10 ans et 6 entre 10-15 ans).
- **Expertise linguistique** : la majorité déclare ne pas avoir d'accent régional ou léger (2/10). 9 sujets sur 10 ont déjà suivi des cours de linguistique dont 6 ont déjà suivi des cours explicites de phonétique-phonologie.
- **Trouble de l'audition** : aucun des sujets n'a de troubles de l'audition sauf un qui déclare qu'ils sont mineurs

**Profil général** : au regard de ces informations, nous pouvons dire que les dix sujets locuteurs natifs du français correspondent à ce que l'on peut s'attendre de locuteurs du français en France métropolitaine. Toutefois, nous prenons en compte que ce sont majoritairement des individus ayant un haut niveau métacognitif et linguistique de par le fait qu'ils suivent tous actuellement des cours à l'université dont un nombre important a déjà suivi des cours de linguistique.

Nous allons maintenant dresser le profil musical des sujets francophones natifs. Les musiciens et non-musiciens ont été départagés de la même manière que les apprenants du français.

- **Place de la musique** : la majorité des sujets déclarent que la musique a une place importante dans leur vie (9/10) et l'intégralité a déjà suivi des cours de musique l'école primaire, collège, lycée.
- **Ecouter de la musique** : la plupart des sujets écoutent de la musique régulièrement soit tous les jours (7/10) soit deux à trois fois par semaine (3/10).
- **Type de musique écouté** : le style le plus écouté est la musique pop (8 occurrences) puis le rock (7). En tout, 9 styles ont été mentionnés.

Afin d'aller plus loin, nous allons préciser le profil des musiciens.

- **Nombre d'années de pratique** : deux sujets indiquent qu'ils jouent de la musique depuis moins de 5 ans, 2 entre 10-15 ans et un plus de 15 ans.



- **Age de début de pratique** : un sujet a commencé la musique avant l'âge de 5 ans, 2 entre 5-10 ans et 2 entre 10-15 ans.
- **Rythme de pratique** : la majorité des musiciens sont des musiciens occasionnels et jouent une fois par mois ou moins (3/5). Une minorité est des musiciens réguliers (tous les jours ou deux fois par semaine) (2/5).
- **Instruments joués** : les instruments joués sont variés : 9 instruments au total sont mentionnés.
- **Type d'apprentissage** : 3 sujets sur 5 ont appris la musique avec des cours.

**Profils des musiciens** : d'après ces informations, il semblerait que les musiciens du panel de locuteurs natifs du français soient de profil hétérogène.

Nous noterons, dès à présent, que nous n'avons pas pris en compte pour notre étude six sujets parmi les apprenants du français qui avaient passé notre test mais dont les conditions expérimentales n'étaient pas optimales pour des temps de réaction fiables. Les différentes raisons qui ont motivé le fait de ne pas retenir ces six sujets ont été : des problèmes techniques liés à E-prime<sup>70</sup>, des difficultés à comprendre la consigne même en anglais, des perturbations dans le test à cause d'un téléphone et des non-réponses dans un des deux tests. Nous avons également écarté un locuteur natif du français car il était étudiant en Master en neurocognition et connaissait par conséquent les biais d'un test AXB et il savait dès le début du test qu'une stratégie AX était possible.

### 5.1.2 Les test AXB

Dans cette partie, nous définirions les choix sous-jacents le test AXB depuis sa conception jusqu'à sa passation.

#### 5.1.2.1 Choix du test

Werker et Tee (1984b : 1876) ont pu démontrer que la parole supporte trois types de traitement : acoustique, phonétique et phonémique (voir section 1.2.1). Ces traitements sont concomitants pendant une courte période car les traitements acoustiques et phonétiques font intervenir l'analyse rapide de caractéristiques acoustiques des sons de la parole tandis que le traitement phonémique fait appel à un stockage à long terme de catégorie sonore. Kamiyama (2009 : 109) rappelle qu'il existe trois types d'expérience en perception phonétophonologique : catégorisation, identification et discrimination (Figure 39).

---

<sup>70</sup> E-prime est un logiciel de conception d'expériences. Il permet de présenter diverses informations (visuelle ou auditive) à un sujet et recueillir ses réponses ainsi que le temps de réaction.

**A. Perception**

	(cible)	(contexte)	(mesure)
-- catégorisation (réponse / transcription libre, choix forcé) -- détection de phonème, de mot -- discrimination de type AX, AXB, ABX,	-- consonnes et voyelles -- syllabes -- accent lexical -- mots	-- contexte vide -- syllabes -- mots / logatomes -- syntagmes -- phrases -- discours	-- réponse donnée par l'auditeur -- temps de réponse

**Figure 39:** Les trois types d'expérience en perception phonético-phonétique avec mise en avant des différentes cibles, contextes et mesures possibles (Kamiyama, 2009 : 109).

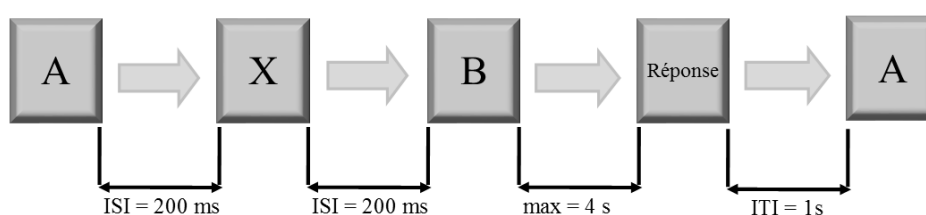
Nous focaliserons notre attention sur le test de discrimination AXB. Dans ce dernier, des triplets de sons sont présentés à un individu qui doit déterminer, pour chaque triplet, si le deuxième son est le même que le premier ou le troisième. Dans le test AXB, le stimulus testé se trouve à égale distance des stimuli repères contrairement à un test ABX dans lequel l'occurrence A est plus phonologisée que B étant donné qu'elle se situe temporellement plus loin de l'occurrence X et de la prise de décision (Pisoni, 1973). De plus, les sujets ont tendance à montrer un biais important envers « B = X » (Schouen, Gerrits & Van Hessen, 2003 dans Gerrits & Schouten : 364). AXB est une variante de ABX ; toutefois Van Hessen & Schouten (1999 dans Gerrits & Schouten : 364) ont démontré que souvent les répondants ignorent le troisième stimulus (B) mais Gerrits (2001 : 44) a trouvé de profondes différences entre les tests de discrimination AX et AXB : « cela implique que les auditeurs ont traité les trois stimuli pour la discrimination AXB, plutôt que de le réduire à un traitement AX ou XB comme indiqué dans les travaux de Pastore et al. (1976) et de Van Hessen and Schouten (1999) »<sup>71</sup>. Une étude de Thantibundhit & al. (2012 : 4) a montré que, par rapport à un test ABX, les tests AXB sont préférés par les sujets (73%) et montrent des taux d'erreurs inférieurs.

Nous avons choisi pour notre étude un test de discrimination car comme le mentionne Kamiyama (2009 : 110) : « que ce soit de AX, ABX, ou AXB, les tests de discrimination peuvent être réalisés auprès des non apprenants de la langue testée, car il s'agit de comparer les stimuli présentés, et non de reconnaître les stimuli comme membre d'une catégorie non native ». Il ajoute, en accord avec les travaux de Pisoni & Tash (1974), que :

<sup>71</sup> "This implies that listeners have processed all three stimuli in AXB discrimination, instead of reducing AXB to AX or XB as was indicated by results of Pastore et al. (1976) and Van Hessen and Schouten (1999)"

« les résultats des expériences de perception peuvent varier selon les types de tâche, par exemple, entre les tâches d'identification et de discrimination (Flege et Hillenbrand, 1984). La tâche d'identification demande une classification fondée davantage sur les catégories phonémiques (donc linguistiques) que la tâche de discrimination, qui relève davantage des différences acoustiques (donc physiques) qui se trouvent entre les stimuli, même si les stimuli d'une même paire ou d'un même triplet ne se succèdent pas immédiatement » (Kamiyama, 2010 : 110).

Dans le cadre de notre réflexion théorique (section 1.2.1), nous avons vu que la pratique musicale pouvait avoir une influence sur la capacité auditive d'autres objets sonores (Kraus & Chandrasekaran, 2010 : 600) mais également des paramètres linguistiques plus abstraits (Chobert & Besson, 2012 : 71). Il nous semble donc pertinent de mettre en place un test de discrimination avec un intervalle interstimulus (ISI) court : 200ms dans notre cas afin de tester un traitement plutôt acoustique et phonétique que phonologique. Pisoni (1973) a mis en évidence le fait que plus petit est l'ISI plus le traitement est auditif tandis que plus il est long plus le traitement est catégoriel, même si Gerrits & Schouten (2004:364) affirment que « au regard du relatif court empan de mémoire auditive (200-300 msec.), le plutôt haut degré de perception catégorielle que l'on retrouve souvent avec un test AXB peut refléter d'après Massaro and Cohen (1983), au recours unique à la mémoire phonétique »<sup>72</sup>. De plus, Gerrits & Schouten (2004:366) affirment que « le traitement du signal acoustique n'est pas terminé après 100-200 msec »<sup>73</sup>. L'intervalle inter-trial (ITI), soit le temps entre deux essais, était de 1 seconde (Hallé, Best & Bachrach, 2003 : 2894) (Figure 40).



**Figure 40:** Déroulement du test AXB avec mise en avant des intervalles interstimulus et inter-trial.

Nous avons choisi de développer un test AXB portant sur des oppositions phonologiques proches afin de cibler des différences acoustiques et phonétiques fines. Cette caractéristique associée au caractère naturel des stimuli permet de respecter d'une part un paradigme de

<sup>72</sup> "In view of the relatively short time span of auditory memory (200–300 msec), however, the rather high degree of categorical perception often found with the ABX task may, according to Massaro and Cohen (1983), reflect the exclusive use of phonetic memory"

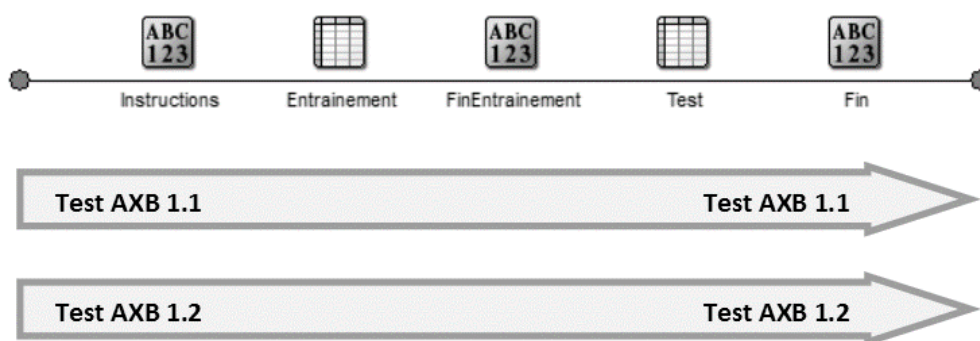
<sup>73</sup> "Processing of the auditory signal is not terminated after 100–200 msec"

recherche en phonétique-phonologie qui impose des paramètres expérimentaux spécifiques mais, d'autre part, des besoins particuliers en didactique des langues qui tend vers un travail d'unités du vivant.

### 5.1.2.2 Déroulement général du test AXB

Le test a été développé sur E-Prime (Psychology Software Tools, 2013) afin d'avoir une mesure fiable du temps de réaction dont l'impact a été démontré dans le « *the reaction time (RT) matching paradigm* » développé par Posner *et al.* (Posner & Mitchell, 1967; Posner, 1969; Posner, Boies, Eichelman, & Taylor, 1969). Comme le rappellent Pisoni & Tash (1974 : 286) : « *cette procédure offre la possibilité d'examiner les niveaux de traitement sur lesquels les comparaisons se font en mesurant le temps de traitement requis pour chaque type de comparaison* »<sup>74</sup>.

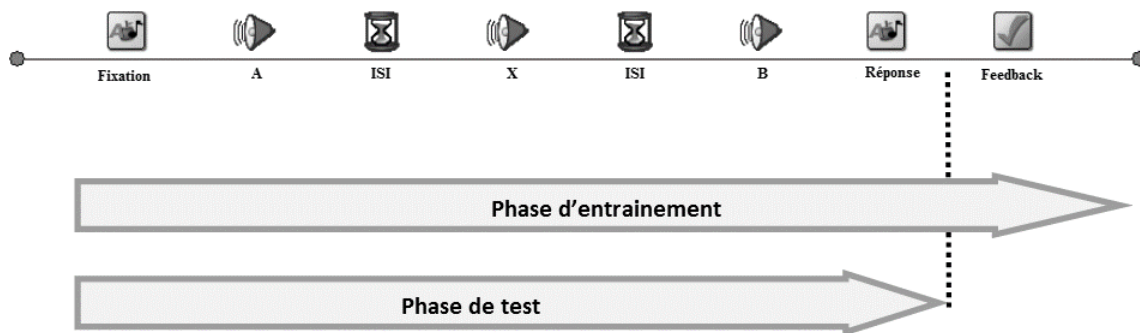
Les deux tests AXB 1 et 2 ont été développés de la même manière en cinq étapes (Figure 41).



**Figure 41:** Déroulement des tests AXB 1 et 2 en cinq étapes selon l'affichage E-Prime.

La phase d'instructions est donnée en annexe 29. La consigne donnée aux sujets était de répondre pour tous les stimuli même s'ils devaient « deviner » la réponse et ils devaient répondre aussi rapidement que possible une fois qu'ils étaient sûrs de leur réponse (Hallé, Best & Bachrach, 2003 : 2894). Cela nous a permis de prendre en compte que ce que Hallé & Best (2007 : 2903) appellent une « *variante de la procédure d'un test de discrimination AXB, le « free RT speed AXB paradigm »* » et qui a démontré son efficacité pour renforcer les différences de temps de réaction entre les groupes ou les contrastes entre stimuli. De manière générale, les phrases d'entraînement et de test sont similaires à l'exception d'un feedback lors de la phase d'entraînement (Figure 42).

<sup>74</sup> « *This procedure provides an opportunity to examine the level of analysis at which comparisons are made by measuring the processing time required for different types of comparisons* »



**Figure 42:** Déroulement des phases d'entraînement et de test selon l'affichage E-Prime. La phase test comprend sept étapes et celle d'entraînement, une de plus soit le « feedback ».

Tout d'abord, l'auditeur voit une croix qui lui permet de fixer son attention pendant que les sons sont joués: « A » puis ISI (200 ms) puis « X » puis ISI (200 ms) puis « B » ; puis il doit choisir sur le clavier de l'ordinateur si le deuxième son est le même que le premier ou le troisième en appuyant respectivement sur la touche « bleu » ou « vert » représentée par des gommettes de couleur sur les touches « CTRL » et « flèche droite » (Figure 43).



**Figure 43:** Placement des touches de réponse sur le clavier d'ordinateur pour les tests AXB. La pastille bleue est sur la touche « CTRL » et la verte sur « flèche droite ».

La phase d'entraînement permet notamment aux auditeurs de se familiariser avec la tâche qui va leur être demandée. La phase d'entraînement consiste en un test AXB de huit triplets soit deux paires vocaliques selon quatre répartitions possibles (2 paires\*4 répartition = 8 triplets) (Tableau 46).

1	/y/	/o/	/o/	+	/o/	/a/	/a/	=	1	/y/	/o/	/o/	Total
2	/y/	/y/	/o/		/o/	/o/	/a/		2	/a/	/a/	/o/	
3	/o/	/y/	/y/		/a/	/o/	/o/		3	/y/	/y/	/o/	
4	/o/	/o/	/y/		/a/	/o/	/a/		4	/o/	/o/	/a/	
Opposition /o/-/y/			Opposition /o/-/a/						5	/o/	/y/	/y/	
									6	/a/	/o/	/o/	
									7	/o/	/o/	/y/	
									8	/o/	/a/	/a/	

**Tableau 46:** Répartition des stimuli du test AXB pour la phase d'entraînement. Au total, huit triplets sont organisés de manière semi-aléatoire selon deux oppositions vocaliques et quatre répartitions possibles.

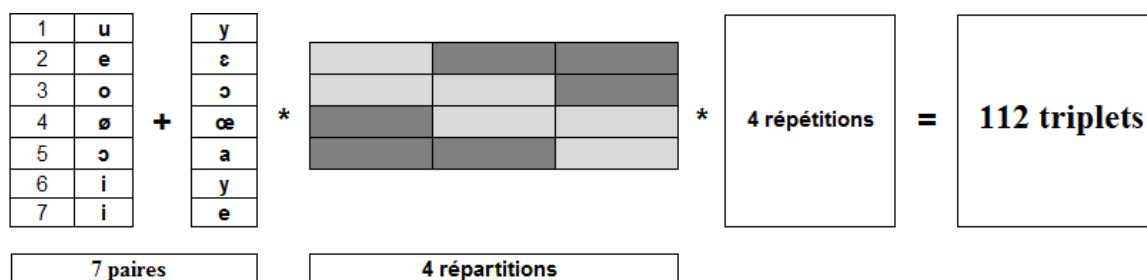
Les oppositions ont été sélectionnées afin de ne pas être présentes dans le test AXB de la phase de test et leur répartition semi-aléatoire a permis de ne pas mettre deux triplets successifs portant sur les mêmes oppositions vocaliques.



**Figure 44:** Affichage des feedbacks positif (à gauche) et négatif (à droite) au test AXB.

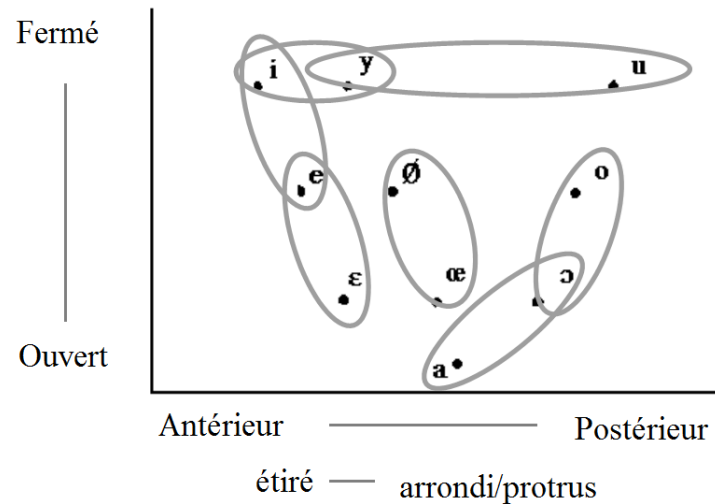
La phase d'entraînement comporte des « feedbacks » (retours) pour chaque triplet afin que l'auditeur puisse savoir s'il a compris la tâche ou non (Figure 44).

La phase de test, pour les tests AXB 1 et 2, se déroule de la même manière que la phase d'entraînement à l'exception du fait qu'il n'y a pas de « feedback » (retour). En général, ce feedback permet de consolider la réponse correcte, de rassurer le sujet sur sa compréhension de la consigne, et d'améliorer les capacités des participants avant le début du test (Collet *et al.*, 2014 : 398). Elle consiste en un test AXB de 112 triplets soit sept paires vocaliques selon quatre répartitions possibles répétées quatre fois (7 paires\*4 répartitions\*4 répétitions = 112 triplets) (Tableau 47).



**Tableau 47:** Répartition des stimuli du test AXB pour la phase de test. Au total, 112 triplets sont organisés selon sept paires vocaliques, quatre répartitions possibles et quatre répétitions.

Les paires ont été choisies afin d'obtenir des duos phonologiquement proches en français standard et par conséquent les différences acoustiques et phonétiques seront fines (Tableau 47 et Figure 45).

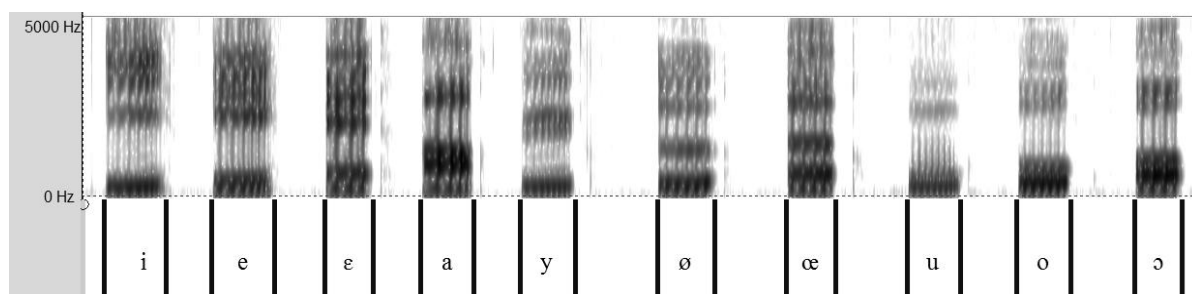


**Figure 45:** Triangle vocalique du français standard. Les sept paires de la phase test du test AXB sont entourées. En abscisse, l'antériorité/postériorité et en ordonnée, l'aperture.

Les 112 triplets ont été organisés de manière semi-aléatoire afin de s'assurer qu'il n'y ait pas de triplets successifs portant sur les mêmes oppositions vocaliques. L'organisation semi-aléatoire est différente pour les tests 1 et 2 afin de limiter les biais provoqués par une progression liée à l'habitude des triplets de la tâche. L'organisation chronologique des triplets pour les tests 1 et 2 est disponible en annexe 30. Les triplets ont été organisés de manière semi-aléatoire à l'aide de la fonction Excel « ALEA » puis d'un tri croissant des données ainsi créées. L'attribution d'un chiffre et d'une couleur pour chaque paire vocalique a permis de contrôler la non-présence de deux mêmes paires successives dans le test.

### 5.1.2.3 Les stimuli

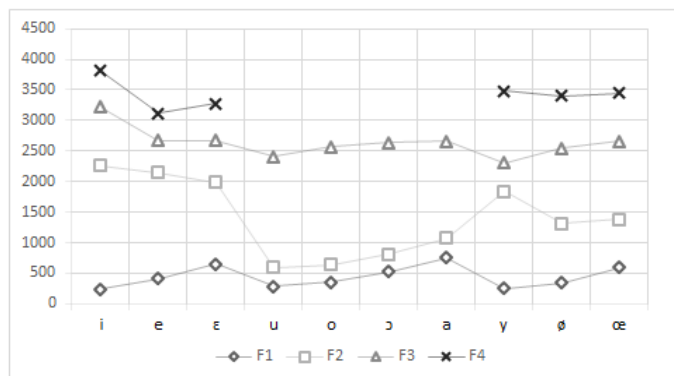
Les stimuli ont été enregistrés par un locuteur homme de 24 ans dont le français est sa langue familiale, de scolarisation et de socialisation. Il a produit les dix voyelles orales du français standard (en ne comptant pas l'opposition /a/-/ɑ/) avec un micro-casque AKG C520L dans la chambre sourde de l'Institut de Linguistique et Phonétique Générales et Appliquées (ILPGA) à Paris sur un ordinateur MAC OS X et le logiciel *Soundforge*. Les stimuli sont disponibles dans un dossier partagé *Googledrive* en ligne (voir partie « documents sonores » de l'étude 2 en annexes).



**Figure 46:** Représentation spectrale sur Praat des dix voyelles orales du français standard enregistrées à l'isolée par LOC1 et utilisées comme stimuli pour les tests AXB. En abscisse, le temps et en ordonnée, la fréquence (en Hz).

Les stimuli ont été analysés et traités sur Praat afin de mesurer les valeurs formantiques de chacune des voyelles (Figure 46 et Figure 47). Nous avons utilisé le script Praat de Cédric Gendrot « log\_file4.praat »<sup>75</sup> qui permet d'analyser de manière semi-automatique les valeurs de durée, fréquence fondamentale (F0) et formants.

	F1	F2	F3	F4
i	245	2259	3221	3810
e	413	2142	2667	3113
ε	649	1991	2669	3274
u	284	593	2412	
o	350	644	2565	
ɔ	525	811	2640	
a	751	1073	2663	
y	256	1839	2305	3487
ø	342	1316	2550	3405
œ	591	1378	2656	3443

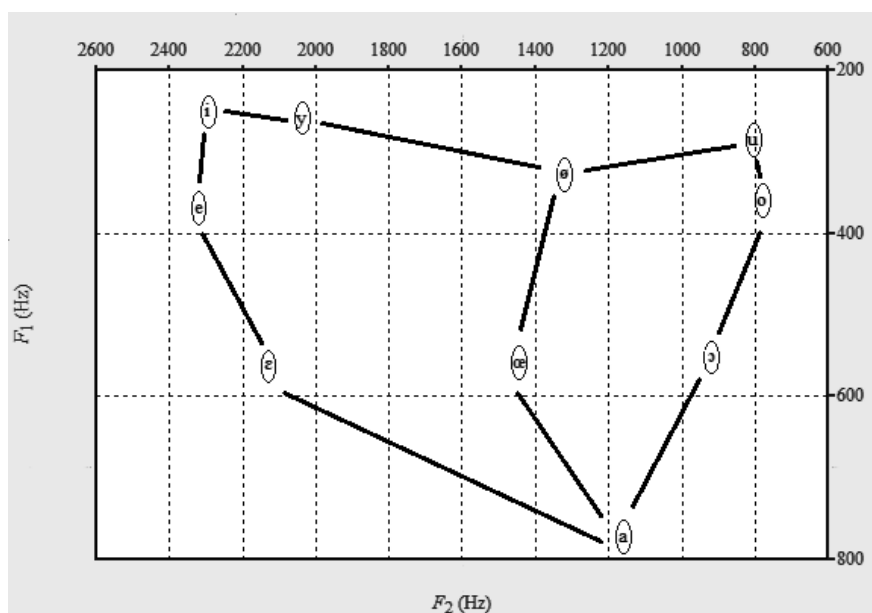


**Figure 47:** Analyse formantique (formants F1 à F2) des stimuli du test AXB. Les valeurs sont présentées sous forme de tableau (à gauche) et dans un graphique en nuage de points (à droite). Le F4 des quatre voyelles postérieures et le /a/ n'a pas pu être déterminé. Valeurs en Hz.

A partir de ces données, nous avons pu afficher le triangle vocalique de LOC1 grâce au script Praat de Cédric Gendrot : « triangle\_vocalique.praat »<sup>75</sup> qui permet de représenter le triangle vocalique (F1/F2) à partir des données récupérées par « log\_file4.praat » (Figure 48). Nous rappelons que l'affichage du rapport F1-F2 n'est pas suffisant pour caractériser les phonèmes vocaliques (Vaissière, 2009 : 31-2 ; section 3.1.3) ; cependant la visualisation en triangle vocalique reste un indicateur pertinent dont il faut garder à l'esprit les limites. Il nous permet notamment de nous rendre compte que le [a] à l'isolée de LOC1 est phonétiquement proche d'un [a] ; toutefois ce n'est pas un problème dans le cas présent car notre étude ne focalise pas sur les représentations phonologiques mais sur des oppositions phonétiques.

<sup>75</sup> [http://ed268.univ-paris3.fr/lpp/pages/EQUIPE/gendrot/page\\_web/scripts.htm](http://ed268.univ-paris3.fr/lpp/pages/EQUIPE/gendrot/page_web/scripts.htm) (consulté le 03/09/13)





**Figure 48:** Triangle vocalique de LOC1 d'après l'analyse formantique des stimuli du test AXB. Le premier formant (F1) est représenté en ordonnée et le deuxième formant (F2) en abscisse (en Hz).

Afin de prendre en compte des valeurs formantiques par rapport à des valeurs de références, nous avons décidé de les comparer aux données de :

- **Calliope** (Tubach, 1989 : 14) : formants mesurés sur deux répétition des onze voyelles françaises orales dans l'entourage [p-] pour les voyelles [e, o, u, y, ø] et [p-R] pour [i, a, ε, ɔ, œ].

	LOC1	Calliope	Ecart F1	LOC1	Calliope	Ecart F2	LOC1	Calliope	Ecart F3
	F1			F2			F3		
i	245	308	63	2259	2064	195	3221	3407	186
e	413	365	48	2142	1961	181	2667	2644	23
ε	649	530	119	1991	1718	273	2669	2558	111
a	751	248	503	1073	1256	183	2663	2503	160
y	256	300	44	1839	1750	89	2305	2120	185
ø	342	381	39	1316	1417	101	2550	2235	315
œ	591	517	74	1378	1391	13	2656	2379	277
u	284	315	31	593	764	171	2412	2027	385
o	350	383	33	644	793	149	2565	2283	282
ɔ	525	517	8	811	1391	580	2640	2379	261
		Moyenne	96,2		Moyenne	193,5		Moyenne	222,1
		Ecart type	146,0		Ecart type	152,9		Ecart type	106,3

**Tableau 48:** Comparaison des valeurs formantiques (en Hz), pour F1, F2 et F3, obtenues pour les voyelles enregistrées par LOC1 et les valeurs de référence de Calliope (Tubach, 1989). L'écart est obtenu par la valeur absolue de la différence des F1 de Loc1 et Calliope.

- **Gendrot & Adda** (2005 : 14) : formants mesurés par statistiques sur base de données du français à partir de corpus de parole continue issu d'enregistrement du journal à la radio « France Inter ».

	LOC1	Gendrot et al.	Ecart F1	LOC1	Gendrot et al.	Ecart F2	LOC1	Gendrot et al.	Ecart F3
	F1			F2			F3		
i	245	310	65	2259	2005	254	3221	2784	437
e	413	370	43	2142	1850	292	2667	2545	122
ɛ	649	438	211	1991	1717	274	2669	2490	179
a	751	557	194	1073	1444	371	2663	2438	225
y	256	336	80	1839	1803	36	2305	2425	120
ø	342	384	42	1316	1474	158	2550	2405	145
œ	591	400	191	1378	1445	67	2656	2440	216
u	284	371	87	593	1105	512	2412	2470	58
o	350	397	47	644	1041	397	2565	2477	88
ɔ	525	456	69	811	1203	392	2640	2420	220
		Moyenne	102,9		Moyenne	275,3		Moyenne	181
		Ecart type	64,4		Ecart type	144,5		Ecart type	101,2

**Tableau 49:** Comparaison des valeurs formantiques (en Hz), pour F1, F2 et F3, obtenues pour les voyelles enregistrées par LOC1 et les valeurs de référence de Gendrot & Adda (2005). L'écart est obtenu par la valeur absolue de la différence des F1 de Loc1 et Gendrot & Adda.

Nous constatons que de manière générale (Tableau 48 et Tableau 49), et ce pour les deux corpora de référence, que les données enregistrées par LOC1 pour les onze voyelles du français sont proches en ce qui concerne les trois premiers formants si l'on considère la marge d'erreur et les variations interindividuelles. Nous remarquons toujours que /a/ subit une variation importante en termes de F1 par rapport aux données de « Calliope » et « Gendrot & Adda » avec un F1 important ce qui le rapproche d'un [ɑ]. Cette caractéristique n'est pas pertinente dans notre étude puisque nous travaillons sur des oppositions et non de l'identification. Nous n'avons pas pu trouver de valeurs de référence « homme » prenant en compte F4. Afin de compléter cette analyse, les stimuli ont été testés par 8 locuteurs (3 hommes et 5 femmes) âgés entre 20 et 34 ans et dont le français est leur L1 à l'aide d'un test d'identification (Tableau 50) (réponses en annexe 31). Le test a été développé et passé sur Praat. Chaque son était présenté deux fois aux auditeurs.

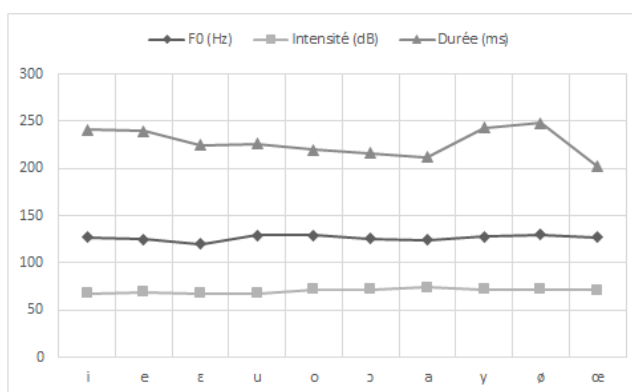
		Réponses données										
		i	e	ɛ	a	ɑ	u	o	ɔ	y	ø	œ
Stimuli	i	16 (5,0)										
	e		16 (4,9)									
	ɛ			16 (5,0)								
	a				14 (4,9)	2 (3,5)						
	u						16 (5,0)					
	o							16 (4,6)				
	ɔ								16 (4,5)			
	y									16 (5,0)		
	ø										16 (4,6)	
	œ											16 (4,5)

**Tableau 50:** Matrices de confusion par 8 locuteurs francophones natifs. En gris, les réponses au test en fonction du stimulus présenté (2 occurrences / phonème) et de la réponse donnée. Entre parenthèses, la moyenne des degrés de certitude sur la réponse entre 1 (pas du tout sûr) et 5 (tout à fait sûr).

Les résultats au test d'identification confirment l'analyse acoustique formantique sur le fait que l'ensemble des phonèmes vocaliques utilisés sont des prototypes pertinents en français standard. Cependant le phonème /a/ peut avoir un caractère ambigu entre un caractère antérieur ou postérieur même si l'auditeur n'a qu'un degré de confiance de 3/5. Toutefois, comme mentionné précédemment, cette distinction n'est pas pertinente dans notre travail expérimental.

Au regard des autres caractéristiques acoustiques, nous avons décidé de ne pas normaliser les stimuli naturels en termes de durée, intensité et F0 étant donné des valeurs stables entre les onze voyelles enregistrées (Tableau 51).

	F0 (Hz)	Intensité (dB)	Durée (ms)
i	127	68	241
e	125	69	240
ɛ	120	68	225
u	129	68	226
o	129	72	220
ɔ	126	72	216
a	124	74	212
y	128	72	243
ø	130	72	248
œ	127	71	202
Moyenne	126,5	70,6	227,3
Ecart type	2,95	2,17	15,22
Minimum	120	68	202
Maximum	130	74	248



**Tableau 51:** Valeurs de F0 (en Hz), d'intensité (en dB) et de durée (en ms) pour les onze voyelles enregistrées par Loc1. Calcul de la moyenne, de l'écart type et des valeurs minimum et maximum. Le graphique représente la moyenne des trois valeurs acoustiques.

Nous précisons également que notre décision de ne pas normaliser les stimuli a été motivée par la volonté dans une recherche en didactique de nous rapprocher, malgré un paradigme issu de la phonétique-phonologie, des unités les plus proches du vivant et d'un contexte naturel.

### 5.1.3 Les tâches

Les tâches ont lieu entre les tests AXB 1 et 2 (rappel Figure 36). Elles ont pour but d'induire un amorçage cognitif. Nous traiterons la notion d'amorçage cognitif dans la perspective de Lashley (1951 : 112) qui considère que : « *l'input n'est jamais dans un système calme et statique, mais toujours dans un système qui est déjà fortement stimulé et organisé [...]* le comportement est le résultat d'interactions entre cet état d'excitation avec l'input de

*n'importe quel stimulus* »<sup>76</sup>. Meineri<sup>77</sup> la définit comme « *orienter le traitement d'un stimulus « cible » (message, requête) par l'exposition préalable à un stimulus « amorce* » ». Notre hypothèse est que l'exposition à une musique instrumentale ira en faveur d'un traitement plus efficace des caractéristiques acoustiques et phonétiques ce qui impliquera, dans ce groupe, une plus grande progression intra-individuelle entre les scores des tests 1 et 2.

### 5.1.3.1 Caractéristiques des tâches

Trois tâches sont au centre de notre étude. La tâche principale de notre étude repose sur le document sonore de musique instrumentale. Le but est de mesurer l'influence d'une exposition musicale sur une tâche de discrimination à travers notre test AXB. Afin de contrôler l'influence d'un *input* auditif sur la tâche de discrimination, nous avons également sélectionné un document de parole continue en français et pour contrôler l'influence d'une modalité autre sur la même tâche, nous avons choisi un document visuel (Tableau 52).

Nom	Modalité de la tâche	Caractéristiques	Variable contrôlée
DV	Visuelle	Tableau d'un paysage calme et reposant	Contrôler l'influence d'une modalité autre sur la tâche de discrimination
DAP	Parole continue en français	Lecture à haute voix d'un texte en français. Débit lent. Voix calme.	Contrôler l'influence de la modalité auditive sur la tâche de discrimination
DAM	Musique instrumentale	Musique instrumentale peu ou pas connue. Tempo lent.	Mesurer l'influence d'une tâche d'amorçage musical sur une tâche de discrimination

**Tableau 52:** Caractéristiques des différentes tâches en fonction de la modalité et de la variable contrôlée. Les sigles représentent le document visuel (DV), le document audio de parole (DAP) et le document audio de musique (DAM).

Nous avons choisi de sélectionner des documents en lien à des émotions calme et reposantes afin de limiter le biais de l'influence d'un état psychologique modifié sur la réalisation de la tâche. En effet, nous avons vu, dans la section 3.3.2, que le tempo d'une musique pouvait

<sup>76</sup> "the input is never into a quiescent or static system, but always into a system which is already actively excited and organized. [...] behavior is the result of interaction of this background of excitation with input from any designated stimulus"

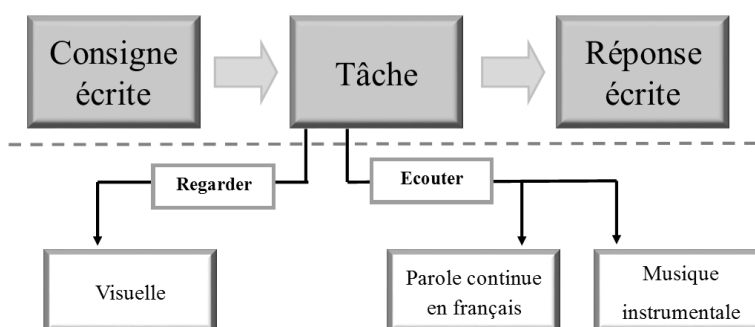
<sup>77</sup> Diapositive n°3 du TD 5 de Sébastien Meineri intitulé « Procédures d'influence & traitement cognitif » dans le cadre du cours « Psychologie de la persuasion, de l'influence et du changement » de la Licence 2 Action Sociale et de Santé à l'Université de Bretagne-Sud. En ligne :

[http://foad.univ-ubs.fr/pluginfile.php/111850/mod\\_resource/content/1/TD%205%20influence%20et%20traitement%20cognitif.pdf](http://foad.univ-ubs.fr/pluginfile.php/111850/mod_resource/content/1/TD%205%20influence%20et%20traitement%20cognitif.pdf) (consulté le 04/09/13).

influencer le rythme cardiaque et par conséquent l'état d'attention (Bernardi *et al.* 2006), par exemple. Ces états émotionnels peuvent influencer le traitement de la tâche de manière quantitative et qualitative. A ce titre, nous avons tenté de limiter les effets de ces mécanismes à travers le choix des documents et de deux pré-études.

### 5.1.3.2 Déroulement de la phase de tâches

La phase de tâche sera organisée en trois étapes : les consignes, la tâche elle-même et enfin une réponse écrite rapide (Figure 49).



**Figure 49:** Déroulement de la phase tâche en trois étapes.

Les consignes (rappel en annexe 29) ont été données à l'écrit (français et anglais) afin de limiter des échanges en français. Il en est de même concernant la réponse écrite qui consistera à définir le tempo du document et donner cinq mots inspirés par le document par une question ouverte (annexes 32 et 33). L'ensemble des auditeurs sera réparti dans l'un des trois groupes (Tableau 53) de manière aléatoire à l'aide des réponses au questionnaire afin d'obtenir le nombre voulu de musiciens et non-musiciens dans chaque groupe.

Groupe témoin 1	Groupe témoin 2	Groupe test		
X	X	X	Test de discrimination	Test AXB 1
X			Visuelle	Tâches d'amorçage
	X		Parole continue en français	
		X	Musique instrumentale	
X	X	X	Test de discrimination	Test AXB 2

**Tableau 53:** Répartition des différents groupes en fonction de la tâche.

Les trois tâches seront minutées et dureront le même laps de temps soit moins de 10 minutes : écouter/regarder (3 minutes) et enfin lecture des questions et réponses (3 minutes). Les deux documents sonores retenus sont disponibles dans un dossier partagé *Googledrive* en ligne (voir partie « documents sonores » de l'étude 2 en annexes).

### 5.1.3.3 Pré-étude 1 : sélection par des musiciens et non-musiciens

La pré-étude 1 a pour but de sélectionner les documents relatifs aux tâches visuelle, de parole en français et de musique instrumentale ayant comme implication émotionnelle, l'évocation du calme et du repos.

Cette pré-étude consiste en un questionnaire Googledocument© en ligne (disponible en annexes 34 et 35) qui questionne les sondés sur cinq documents : deux documents audio musicaux (DAM1 & DAM2), un document audio de parole en français (DAP3) et deux documents visuels (DV4 & DV5). DAP 3 a servi de points de référence pour sélectionner les documents musicaux et visuels. L'enquête a été largement diffusée à travers les réseaux sociaux et listes de diffusions en didactique des langues, phonétique-phonologie et groupes liés à la musique. Les 5 documents, présélectionnés par l'auteur en collaboration avec un batteur semi-professionnel, sont les suivants :

1) Document 1 audio musical (DAM1) :

**Titre:** Ljósið

**Auteur:** Ólafur Arnalds (artiste islandais)

**Source :** <https://www.youtube.com/watch?v=mYIfiQlfaas> (consulté le 31/07/14)

**Type de document :** une musique instrumentale minimaliste au croisement du post-rock et de la musique classique.

**Modification du document :** aucune

**Durées initiale et finale :** 3 minutes 37 secondes

2) Document 2 audio musical (DAM2) :

**Titre :** Metamorphosis 2

**Auteur :** Philip Glass (compositeur nord-américain)

**Source :** <http://www.youtube.com/watch?v=CcyTyLqkvoU> (consulté le 04/09/13)

**Type de document :** une œuvre pour piano minimaliste.

**Modification du document :** suppression d'une bande de 4 minutes 1 seconde (entre 2 minutes 31 secondes et 6 minutes 32 secondes) afin de réduire la longueur totale du document audio et supprimer le deuxième mouvement du morceau dont l'intensité émotionnelle change par rapport aux premier et troisième mouvements.

**Durée initiale :** 7 minutes 30 secondes

**Durée finale :** 3 minutes 29 secondes

3) Document 3 audio de parole (DAP3) :

**Titre** : Alexandre\_\_Dumas\_-\_Le\_Comte\_de\_Monte\_Cristo\_T2\_Chp\_01

**Donneur de voix**: Serge Dewulf

**Source** : [http://www.litteratureaudio.com/index.php?s=Le\\_Comte\\_de\\_Monte\\_Cristo\\_T2\\_Chp\\_01&sbtt=Ok](http://www.litteratureaudio.com/index.php?s=Le_Comte_de_Monte_Cristo_T2_Chp_01&sbtt=Ok)

(consulté le 04/09/13)

**Type de document** : une lecture à haute voix du début du chapitre XXII « Les contrebandiers » de de l'œuvre d'Alexandre Dumas : *Le Comte de Monte Cristo*. La transcription du document est disponible en annexe 36. Compte tenu de la syntaxe et du lexique employé, cet extrait nous semble adapté à un niveau C1/C2 du CECRL.

**Modification du document** : suppression des 13 minutes 57 secondes finales du chapitre afin de limiter la longueur totale du document audio. Nous avons coupé lors d'une pause naturelle dans la lecture.

**Durée initiale** : 17 minutes 12 secondes      **Durée finale** : 3 minutes 15 secondes

4) Document 4 visuel (DV4) :

**Titre** : *Autumn*

**Série** : *Future Memories*

**Artiste** : Catherine Nelson

**Source** : <http://catherinenelson.net/gallerij-futurememories.php#/2> (consulté le 04/09/13)

**Type de document** : une œuvre d'assemblages photographiques et de photo-manipulation.

**Modification du document** : aucune



*Figure 50* : Document 4 visuel (DV4)



## 5) Document 5 visuel (DV5) :

**Titre:** Landscape with a Long Arched Bridge

**Artiste :** Rembrandt

**Source :**

<http://arsmoderna.files.wordpress.com/2011/06/rembrandt-landscape-with-a-long-arched-bridge.jpg> (consulté le 04/09/13)

**Type de document :** une peinture sur toile baroque

**Modification du document :** aucune



**Figure 51 :** Document 5 visuel (DV5)

Le questionnaire se composait de deux questions par document adaptées pour les documents visuels et questionnant respectivement la perception du tempo et les émotions véhiculées par le document (Tableau 54).

Question	Documents	Questions
1	Audio musicaux et parole (DAM1, DAM2 & DAP3)	« Comment évalueriez-vous le tempo général du document ? »
	Visuels (DV4 & DV5)	« Si vous deviez associer une mélodie à cette image, quel tempo aurait-elle ? »
2	Audio musicaux et parole (DAM1, DAM2 & DAP3)	« Sélectionnez cinq mots différents caractérisant au mieux, selon vous, le document »
	Visuels (DV4 & DV5)	« Sélectionnez 5 mots différents caractérisant au mieux, selon vous, le document visuel et la mélodie qui vous lui associez »

**Tableau 54:** Questions du questionnaire de la pré-étude 1 portant sur la sélection des documents pour la phase tâche.

La question 1 à échelle d'attitude, portant sur l'évaluation du tempo, propose cinq gradients de réponse : très lent, lent, modéré, vif et très vif. Les termes ont été sélectionnés, selon un tableau disponible dans le « *Virginia Tech Multimedia Music Dictionary* » (Dictionnaire musical multimédia de Virginia Tech) qui offre une classification des termes courants en quatre langues permettant de décrire un tempo (disponible en annexe 37). Cette classification propose neuf termes mais nous n'en avons retenu que cinq afin que chacun d'eux puissent être pertinents à des non-musiciens.



La question 2 demande au répondant de sélectionner cinq mots dans une liste de 28 mots (Tableau 55) qui ont été sélectionnés par nous au préalable dans une fiche d'adjectif de « *vocabulaire pour caractériser une musique* » créée par une enseignante de musique au collège (annexe 38). Nous avons tenté garder une grande variété de termes des émotions positives (« dansant », « joyeux ») et négatives (« triste », « sombre »).

1	Dansant	8	Bucolique	15	Joyeux	22	Paisible
2	Dynamique	9	Mélancolique	16	Léger	23	Serein
3	Energique	10	Triste	17	Insouciant	24	Monotone
4	Animé	11	Sombre	18	Magique	25	Solennel
5	Militaire	12	Macabre	19	Rêveur	26	Elégant
6	Puissant	13	Violent	20	Enfantin	27	Majestueux
7	Triomphal	14	Calme	21	Mystérieux	28	Amusant

**Tableau 55:** Les 28 termes sélectionnés permettant de définir la projection émotionnelle des différents documents.

Notre pré-étude a permis de rassembler les réponses de 61 répondants dont le profil socio-culturel est le suivant :

- **39 sont des femmes et 22 des hommes,**
- **35 sont de nationalité française et 26 d'autres nationalités,**
- **la majorité est âgée entre 20 et 50 ans (51/61)** mais 4 ont entre 15-20 ans et 6 ont 50 ans et plus,
- **ils écoutent de la musique entre souvent (17/61) et tous les jours (33/61)** mais 5 écoutent peu souvent et 6 très souvent,
- **il y a 15 non-musicien-es et 46 musicien-nes** dont 25 musicien-nes amateur-riche, 8 musicien-nes semi-professionnel-les et 13 musicien-nes professionnel-les,
- **les musicien-nes jouent de la musique : rarement (10), peu souvent (7), souvent (15) ou tous les jours (14).**

Notre analyse portera de manière indépendante sur les réponses aux questions 1 et 2 afin de voir quels documents sont les plus pertinents pour notre étude :

- Question 1 : « *Comment évalueriez-vous le tempo général du document ?* » / « *Si vous deviez associer une mélodie à cette image, quel tempo aurait-elle ?* »

De manière générale, nous constatons que le document de référence soit DAP3 est majoritairement perçu comme d'un tempo modéré, les documents musicaux sont perçus comme lents ou modérés. Les réponses sont plus mitigées en ce qui concerne les documents visuels. Cependant, DV4 est plutôt perçu comme vif et DV5 comme lent (Tableau 56).

	DAM1	DAM2	DAP3	DV4	DV5	Total
Très lent	3	3	1	1	13	21
Lent	33	20	4	14	24	95
Modéré	25	31	46	13	8	123
Vif	0	7	10	26	11	54
Très vif	0	0	0	7	5	12
Total	61	61	61	61	61	

**Tableau 56:** Réponses relatives à la question portant sur le tempo perçu des cinq documents. Les réponses importantes (>20) sont sur fond gris.

Afin de nous rendre compte de quel document (musical et visuel) est le plus proche en termes de tempo de notre document de référence (DAP3), nous allons mesurer leur distance en nombre de réponses (Tableau 57).

	DAM1	DAM2	DAP3	DV4	DV5
Très lent	2	2		0	12
Lent	29	16		10	20
Modéré	21	15		33	38
Vif	10	3		16	1
Très vif	0	0		7	5
Total	62	36		66	76
Classement	2	1+		1	2

**Tableau 57:** Distance relative par rapport au nombre de réponses données pour DAP3. Le résultat est obtenu par la valeur absolue de la différence entre la réponse d'une case et celle correspondante à DAP3. Une valeur faible signifie que le nombre de répondants est proche de celui de DAP3. Le classement est donné pour les documents musicaux d'une part et les documents visuels d'autre part.

Nous pouvons constater que, parmi les documents musicaux, DAM2 est celui qui se rapproche le plus, de par les réponses données, à DAP3 avec une distance totale de 36 contre 66 pour DAM1. Concernant les documents visuels, le choix est moins tranché mais DV1 arrive en premier avec une distance de 66 contre 76 pour DV5. D'après ces premiers résultats, il semblerait que juger le tempo d'une musique associée à une image soit une tâche complexe qui ne permet pas d'obtenir de tendance tranchée.

**Conclusion à la question 1 :** DAM2 est le plus proche en termes de tempo de DAP3 pour les documents musicaux. DV4 se détache légèrement pour les documents visuels mais de manière moins tranchée.

- Question 2 : « Sélectionnez cinq mots différents caractérisant au mieux, selon vous, le document » / « Sélectionnez 5 mots différents caractérisant au mieux, selon vous, le document visuel et la mélodie que vous lui associez ».

Nous commencerons par analyser les premiers mots sélectionnés parmi les cinq demandés aux répondants. La totalité des réponses est donnée en annexe 39. Nous considérons que le premier mot choisi peut représenter le plus prototypique pour le répondant (Tableau 58).

	Dansant	Dynamique	Energique	Animé	Puissant	Bucolique	Mélancolique	Triste	Sombre	Calme	Paisible	Serein	Monotone	Solennel	Élégant	Rêveur	TOTAL REPOSE	TOTAL MAX
DAM1						6	15			9	13	4	4			4	55	
DAM2							18	10	4				6			5	43	
DAP3		6								14			10	5	5		40	61
DV4	9	10	8	7		11											45	
DV5					9	10	11	6	8								44	
Total	9	16	8	7	9	27	44	16	12	23	13	4	20	5	5	9		

**Tableau 58:** Réponses les plus importantes pour le premier mot donné parmi les cinq demandés à la question 2. Seuls les mots ayant obtenus 5 réponses ou plus sont affichés. En gras, les données qui correspondent aux réponses pour DAP3, notre document de référence. « Total réponses » correspond à la somme des réponses supérieures à 7 et « TOTAL MAX » au nombre total de réponses données pour chaque document.

Nous pouvons constater qu'en ce qui concerne le premier mot donné, DAP3/DAM1/DAM2/DV5 sont proches en termes de sensations provoquées soit « mélancolique, triste, sombre, calme, monotone ». DV4 se détache nettement puisque, contrairement aux autres, il est perçu comme « dansant, énergique et animé ». Nous allons maintenant comparer ces données avec l'ensemble des cinq mots choisis par chaque répondant (Tableau 59).

	Dansant	Dynamique	Energique	Animé	Puissant	Bucolique	Mélancolique	Triste	Sombre	Calme	Paisible	Serein	Monotone	Solennel	Élégant	Léger	Insouciant	Magique	Rêveur	Mystérieux	Total réponses
DAM1							32			45	55	31	19			15			38		235
DAM2							36	40	28	27	15		39						17	23	225
DAP3		16		15						40	22	24	36	26	19		15				213
DV4	26	25	21	21		18												29	15	15	170
DV5					26	19	26	19	44											30	164
Total	26	41	21	36	26	37	94	59	72	112	92	55	94	26	19	15	15	29	70	68	

**Tableau 59:** Réponses les plus importantes pour tous les mots à la question 2. Seuls les mots ayant obtenus 15 réponses ou plus sont affichés. En gras, les données qui correspondent aux réponses pour DAP3, notre document de référence. « Total réponses » correspond à la somme des réponses supérieures à 7. Le nombre total de réponses données est de 305.

Nous pouvons constater que le profil de réponses reste majoritairement similaire entre le mot 1 et les cinq mots au total. DAP3 est le plus proche de DAM1 et DAM2 en termes de réponses puisqu'ils sont tous les trois principalement vus comme « *calme, paisible, serein et monotone* ». DAM1 est légèrement plus proche de DAP3 selon les réponses données. Nous constatons que DV5 est vu comme « *puissant, bucolique, mélancolique, triste, sombre et mystérieux* » tandis que DV4 est perçu comme « *dansant, dynamique, énergique et animé* ». DAP3 est également perçu comme « *dynamique et animé* » mais cela ne représente respectivement que 16 et 15 réponses sur 305 au total.

**Conclusion à la question 2 :** DAM1 a une légère avance sur DAM2 dans leur proximité avec les réponses données pour DAP3. Ni DV4, ni DV5 ne semblent être proches en termes de réponses à DAP3, cependant, DV4 se distingue largement des autres par une intention « *dansant, dynamique, énergique et animé* » contre « *calme, paisible, serein et monotone* » pour tous les autres.

**Conclusion aux questions et justification du choix des documents :** nous avons vu qu'en ce qui concerne le tempo, DAM2 se détache largement par sa proximité de réponses avec DAP3. De plus, ni DAM1 ni DAM2 ne semblent montrer de distinction à la question 2. DAM2 apparaît donc comme le plus pertinent parmi les documents musicaux. Concernant les documents visuels, DV4 est sensiblement plus proche de DAP3 en termes de tempo perçu mais se détache complètement au sujet des mots choisis. En ce sens, nous choisirions DV5 comme document visuel. Cependant afin de confirmer ce choix, nous avons mené une pré-étude 2 afin de questionner des experts en art.

### 5.1.3.4 Pré-étude 2 : confirmation par des experts en art

La pré-étude 2 correspond, en tous points, à la pré-étude 1 à l'exception du public visé et du fait que seuls les documents visuels ont été questionnés (questionnaire disponible en annexes 40 et 41). L'enquête a été diffusée sur des groupes liés aux arts pictural et photographique sur les réseaux sociaux (*Facebook, LinkedIn, Viadeo*, etc.). Nous n'avons pu regrouper que 32 répondants ; cependant, nous considérons ce nombre suffisant au regard du but de cette pré-étude : la confirmation d'une des conclusions de la pré-étude 1.

Le profil socio-culturel des répondants est le suivant :

- **20 sont des femmes, 11 des hommes** et 1 non-spécifié,
- **20 sont de nationalité française** et **12 d'autres nationalités**,

- **la majorité est âgée entre 20 et 50 ans (26/32)** mais 2 ont entre 15-20 ans et 4 ont 50 ans et plus,
- **ils écoutent de la musique entre souvent (17/61) et tous les jours (33/61)** mais 5 écoutent peu souvent et 6 très souvent,
- il y a **2 non-connaisseurs et 30 connaisseurs** dont 11 connaisseurs amateurs, 13 grands connaisseurs ou étudiant-e-s en arts et 6 spécialistes ou professionnels en arts.
- **21 participent souvent (13) ou très souvent (8) à des événements en lien avec l'art pictural/photographique.** 9 déclarent y participer peu souvent et 2 rarement.

A la question 1, concernant le tempo imaginé pour une musique associée aux images, nous constatons le même schéma de réponses soit DV4 comme plutôt modéré/vif (22/32) et DV5 comme lent/modéré (20/32) que lors de la pré-étude 1. Même s'il n'y a pas de réponse tranchée car DV4 et DV5 sont majoritairement perçus comme lent/modéré/vif respectivement à 31/21 et 28/32 (Tableau 60).

	Très lent	Lent	Modéré	Vif	Très vif	Total
<b>DV4</b>	<b>0</b>	<b>9</b>	<b>12</b>	<b>10</b>	<b>1</b>	32
<b>DV5</b>	<b>4</b>	<b>14</b>	<b>7</b>	<b>7</b>	<b>0</b>	32
Total	4	23	19	17	1	

**Tableau 60** : Réponses relatives à la question portant sur le tempo perçu des deux documents visuels. Les tendances importantes sont sur fond gris.

En ce qui concerne le choix des mots à la question 2 (l'ensemble des réponses est disponible en annexe 42), nous constatons une fois encore, le même schéma de réponse que lors de la pré-étude 1 (Tableau 61).

	Dansant	Dynamique	Animé	Puissant	Bucolique	Mélancolique	Triste	Sombre	Calmé	Paisible	Solennel	Elégant	Léger	Insouciant	Magique	Rêveur	Mystérieux	Total Réponses	TOTAL MAX
DV4	14	12	9	7	17	8			7	9		7	7		7	15	9	128	160
DV5				13	7	20	16	25	9		7	7					19	123	160
Total	14	12	9	20	24	28	16	25	16	9	7	14	7	0	7	15	28		

**Tableau 61** : Réponses les plus importantes pour tous les mots à la question 2. Seuls les mots ayant obtenu 7 réponses ou plus sont affichés. « Total réponses » correspond à la somme des réponses supérieures à 7 et « TOTAL MAX » au nombre total de réponses données pour chaque document.

DV4 est principalement perçu comme « *dansant, dynamique, bucolique et rêveur* » tandis que DV5 comme « *bucolique, mélancolique, triste, sombre et mystérieux* ». Nous rappelons que dans la pré-étude 1, DAP3 (document de référence) avait été perçu comme « *calme, paisible, serein et monotone* ». Il semble donc que DV5 soit le plus pertinent pour notre étude car il se rapproche le plus en termes émotionnels de DAP3.

**Justification générale du choix des documents d'après les pré-études 1 et 2 :** au terme de ces deux pré-études, il semble pertinent de sélectionner les documents DAM2 et DV5 en référence à DAP3. Ce choix est justifié par le fait que DAM2 montre le tempo perçu le plus proche de DAP3. Il montre également une évocation émotionnelle proche avec des termes comme « *calme, paisible, serein et monotone* ». DV5 n'a pas pu être sélectionné sur le thème du tempo car imaginer un tempo à une musique reliée à une image semble ne pas permettre de dégager des tendances importantes. Cependant, DV4 semble avoir une évocation émotionnelle, démontrée dans la pré-étude 1 et confirmée par la 2, comme « *dynamique et dansant* » ce qui va à l'inverse des émotions perçues pour les quatre autres documents. DV5 montre un profil émotion proche de DAM2 et DAP3.

## 5.2 Résultats

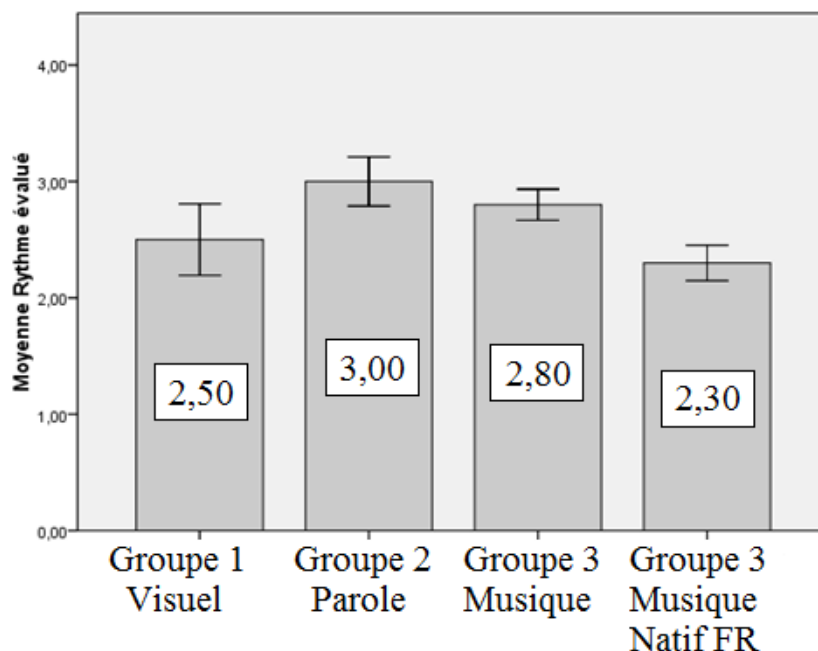
Deux analyses de variance ANOVA à mesures répétées ont été effectuées l'une par sujets (F1) et la seconde par items (F2) dans le but de tester l'influence du statut musical (musicien ou non-musicien) et du Groupe (groupe 1 visuel, groupe 2 parole en français, groupe 3 musique instrumentale) sur le taux de discrimination et sur les temps de réaction pour une tâche de discrimination AXB. Nous focaliserons notre analyse sur les facteurs montrant des tendances significatives à la fois pour F1 et F2 puisque elles seront généralisables au regard du fait que ces tendances prennent en compte à la fois les sujets et les items comme facteur aléatoire. Nous qualifierons de « marginalement significative » une tendance pour laquelle la valeur de p est comprise en 0,05 et 0,10 et si cette tendance va dans le sens de nos hypothèses (Gelman, 2013 : 69). L'ensemble des tableaux récapitulatifs des données et des mesures statistiques est donné en annexes de 43 à 53. Nous signalons également que notre protocole expérimental et les analyses statistiques ont été effectués avec l'aide d'une ingénieure en psycholinguistique rencontrée à plusieurs reprises tout au long de l'analyse des données (attestation en annexe 54).

### 5.2.1 Questionnaire relatif aux tâches

Nous commencerons par analyser les résultats obtenus lors du questionnaire permettant de mesurer pour chaque sujet leur perception du tempo du document proposé ainsi que cinq mots que leur inspirait le document. Nous rappelons que ce questionnaire était soumis à chaque sujet après exposition à un des trois documents en fonction du groupe.

- Le tempo

Au regard du tempo, nous constatons, pour les apprenants du français, une moyenne du tempo évalué autour de 2,65 sur 5 correspondant sur notre échelle à un tempo « modéré » avec une tendance vers « lent » (Figure 52). Nous rappelons que cette échelle allait de 1 (tempo très lent) à 5 (tempo très vif).



**Figure 52:** Moyennes et écart-types évalués en fonction de chaque groupe pour le tempo perçu après exposition à l'un des trois documents de la tâche. 1 = tempo très lent : 5 = tempo très vif. Les barres d'erreur indiquent l'erreur standard.

	Groupes	n=	moyenne	Test t	ddl	Valeur de p	<0,05 = *
App. FR	1	10	2,5 (0,97)	-1,342	18	0,196	ns
	2	10	3,0 (0,67)				
	1	10	2,5 (0,97)	0,802	18	0,433	ns
	3	10	2,8 (0,42)				
	2	10	3,0 (0,67)				
3	10	2,8 (0,42)	2,466	18	0,024	*	
App. FR	3	10					2,8 (0,42)
Natif FR	3	10	2,3 (0,48)				

**Tableau 62:** Récapitulatif de l'analyse statistique Test t pour échantillons indépendants concernant les différences entre les quatre groupes au regard du tempo évalué.

Au regard d'une analyse de type Test-t pour échantillons indépendants entre les différents groupes (Tableau 62), nous ne pouvons constater une différence significative qu'entre le groupe 3 d'apprenants et celui de natifs du français ( $t_{18} = 2,466$ ,  $p=0,024$ ) évalué respectivement à 2,8 et 2,3 sur 5. Cette différence reste néanmoins minimale est la significativité due probablement à des écart-types faibles.

- La projection émotionnelle

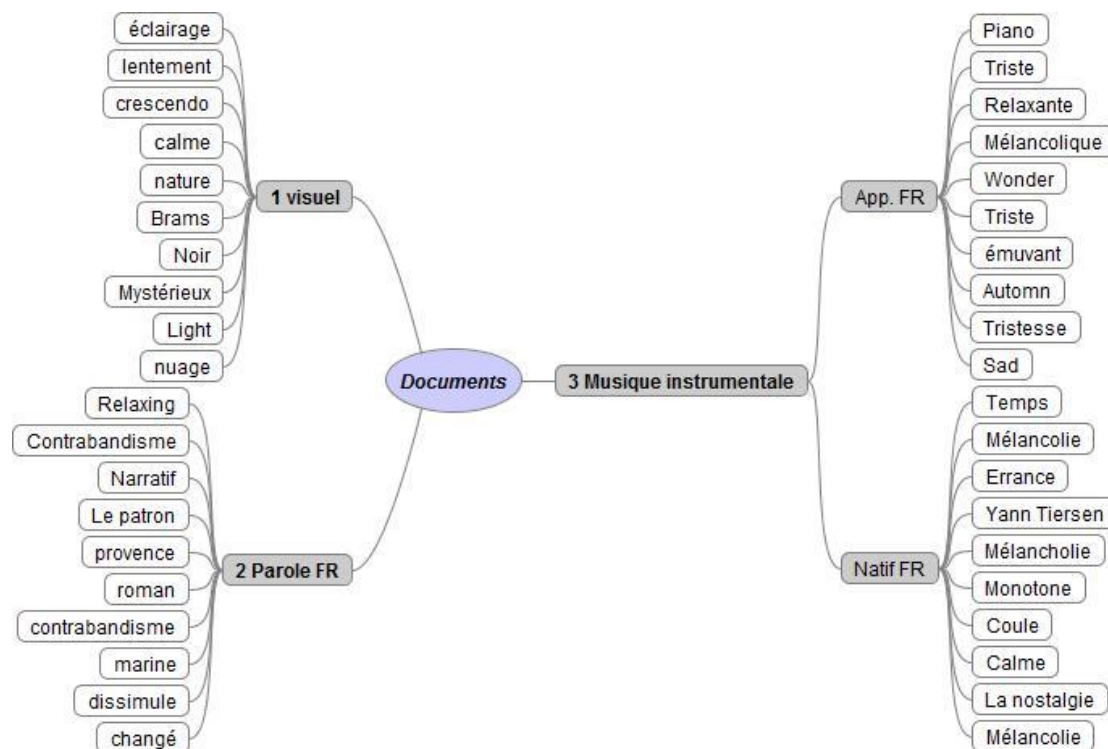
En ce qui concerne les cinq mots donnés, nous allons concentrer notre analyse sur une comparaison du nombre de mots donnés (les sujets n'étaient pas forcés de donner cinq mots) ainsi qu'une comparaison des premiers mots donnés par les 40 sujets et ce en fonction du document fourni (rappel : la liste complète des mots est donnée en annexe 43).

StatLang	Groupe	Musique	Mot 1	Mot 2	Mot 3	Mot 4	Mot 5	Total	Total
Appr.FR	1	1	5	5	5	5	4	24	46
		2	5	5	5	4	3	22	
	2	1	5	5	5	3	2	20	39
		2	5	4	4	3	3	19	
	3	1	5	5	5	5	4	24	47
		2	5	5	5	4	4	23	
Natif FR	3	1	5	5	5	4	3	22	44
		2	5	5	5	4	3	22	

**Tableau 63:** Nombre de mots donnés lors du questionnaire en fonction du groupe.

Au regard du nombre de mots donné (Tableau 63), nous constatons que les documents visuel et de musique instrumentale ont largement inspiré les sujets avec respectivement 46 et 47 mots sur 50 au total. Le document de parole continue en français a suscité moins de mots soit 39 sur 50 au total. Les francophones natifs du groupe 3 ont été un peu moins inspirés que les apprenants du français avec 44 mots contre 47 sur 50. Afin d'approfondir ces tendances, nous allons maintenant regarder plus en détail les premiers mots donnés par les sujets en fonction du type de document. Nous n'irons pas au-delà de cette analyse en raison du faible apport d'une analyse fine des différents mots fournis pour cette étude.





**Figure 53 :** Carte heuristique des premiers mots donnés par chaque sujet en fonction des quatre groupes (visuel, parole en français, musique instrumentale : apprenants et locuteurs natifs du français). Ce type de représentation permet mettre en lumière les liens qui existent entre un concept ou une idée, et les informations qui leur sont associées.

Les mots donnés (Figure 53) n'étant pas contrôlés, nous observons 8 occurrences de mots en anglais ou des mots en français soumis à des variations dues à l'activation d'autres langues comme l'espagnol par exemple : « contrabandisme ». Concernant le document 2 de parole continue en français, il est clair que les premiers mots donnés pas les apprenants ont été activés par leur présence dans le texte pour 7 occurrences sur 10 dont les mots apparaissent au moins une fois dans le document sonore (« contrebandisme », « le patron », « provence », « marine », « dissimule » et « changé ». Les trois autres mots font soit référence au type de document (« narratif » et « roman ») soit à l'impact émotionnel (« relaxing » (relaxant)). Pour le document 1 visuel, nous trouvons un ensemble de termes liés au caractère baroque de la peinture avec un univers « mystérieux », « noir », « calme », « nuage », « lentement », « nature » qui s'oppose à la lumière « éclairage », « light » (lumière) d'un univers en mouvement « crescendo » et relié au monde de la musique « Brahms ». Le document 3 de musique instrumentale se voit associé avec le même champ lexical par les apprenants et les natifs du français de la « mélancolie », « nostalgie », « triste(sse) », « errance », « monotone ». Pour les trois, nous retrouvons les termes de la pré-étude qui étaient « mélancolie » (44 occurrences), « calme » (23 occurrences) et « monotone » (20 occurrence).

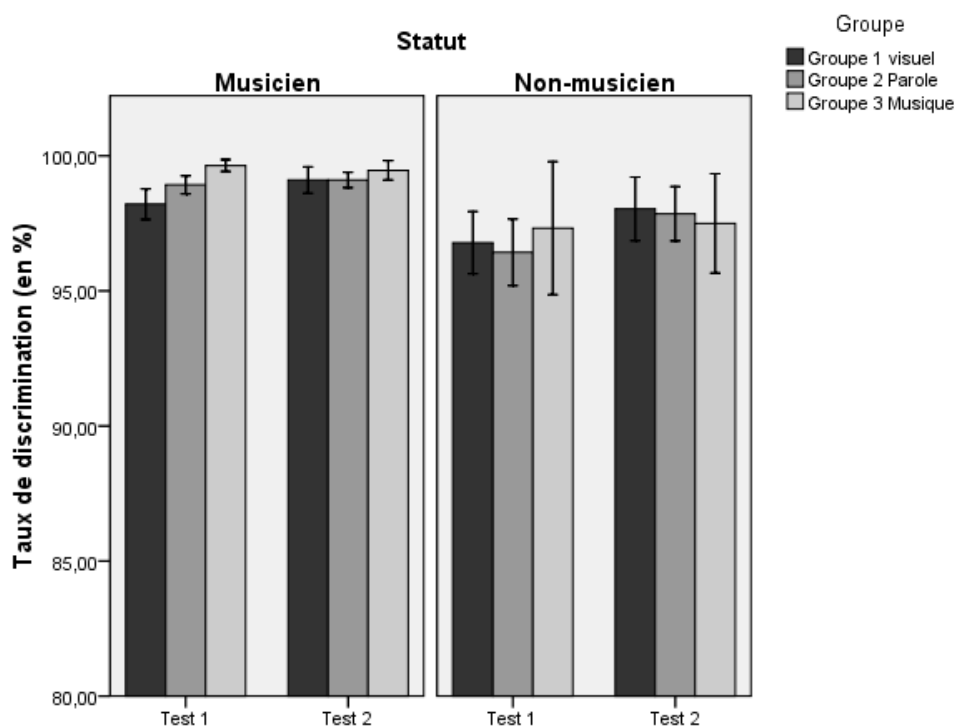
Les résultats obtenus lors de ces questionnaires semblent être en corrélation avec ceux obtenus lors de notre première étude. Nous rappelons que le but visé était de contrôler deux variables (le temps, la projection émotionnelle) qui auraient pu jouer comme des perturbateurs dans la réalisation de la tâche à travers la modification de constantes physiologiques (augmentation du rythme cardiaque, etc.). A ce titre, il semblerait que la pré-sélection effectuée par notre pré-étude ait été pertinente.

### 5.2.2 Les apprenants du français

Cette analyse prendra en compte les données des 30 apprenants répartis selon les trois groupes (10 sujets/groupe) et deux statuts musicaux (15 sujets/statut soit 5 sujets/groupe/statut musical).

- Taux de discrimination

Nous commencerons par analyser le taux de discrimination à l'aide d'une analyse de variances ANOVA à mesures répétées avec pour facteur intra-sujet le Test (Test1 vs. Test2) et pour facteurs inter-sujets le Groupe (1 visuel vs. 2 français vs. 3 musique instrumentale) et le statut musical (musicien vs. non-musicien).

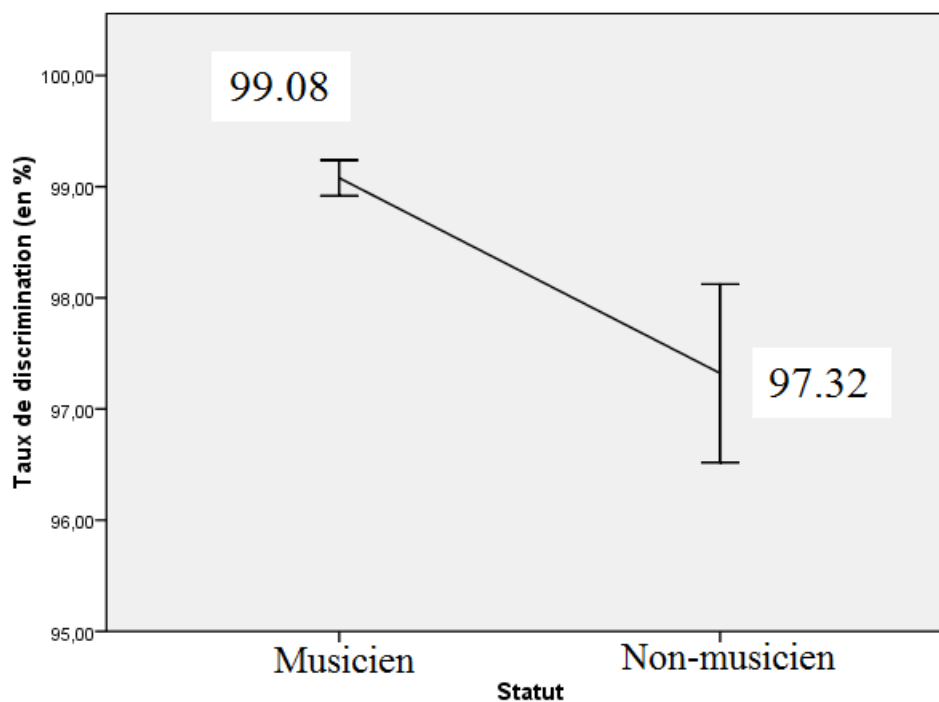


**Figure 54** : Diagramme du taux de discrimination moyen (en %) en fonction du groupe, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard.

Taux de discrimination (en %)		Test 1	Test 2
Groupe 1 Visuel	Musicien	98,21 (1,26)	99,11 (1,09)
	Non-musicien	96,79 (2,57)	98,04 (2,63)
Groupe 2 Parole	Musicien	98,93 (0,75)	99,11 (0,63)
	Non-musicien	96,43 (2,75)	97,86 (2,24)
Groupe 3 Musique	Musicien	99,64 (0,49)	99,46 (0,80)
	Non-musicien	97,32 (5,50)	97,50 (4,11)

**Tableau 64** : Récapitulatif du taux de discrimination (en %) en fonction du groupe, du test et du statut musical. Les écart-types sont donnés entre parenthèses.

Nous observons, tout d'abord et d'après la Figure 54 et le Tableau 64, que sur l'ensemble des trois groupes d'apprenants et ce quel que soit leur statut musical (musicien vs. non-musiciens) et le test (1 vs. 2), le taux de discrimination est, en moyenne, de 98,2% soit moins de 2% d'erreurs.



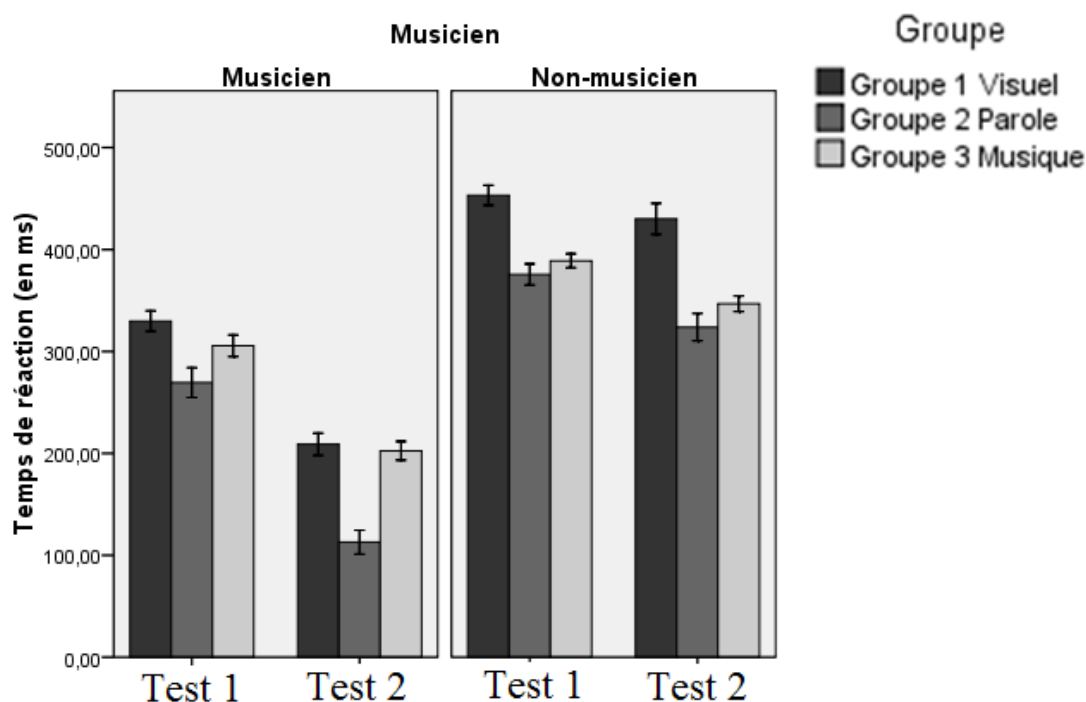
**Figure 55**: Impact du statut musical sur le taux de discrimination moyen (en %). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,24)=4,015$ ,  $p=0,057$  ;  $F(1,6)=64,751$ ,  $p=0,036$ ). Les barres d'erreur indiquent l'erreur standard.

Nous pouvons constater, d'après notre analyse de variance, qu'il n'y a qu'un effet marginalement significatif pour l'effet principal du statut ( $F(1,24)=4,015$ ,  $p=0,057$  ;  $F(1,6)=64,751$ ,  $p=0,036$ ) avec un meilleur score par les musiciens de 1,76% (Musiciens= 99,08% et Non-Musiciens= 97,32%) (Figure 55). Nous remarquons également un effet significatif pour F1 pour la mesure répétée du Test ( $F(1,24)=3,973$ ,  $p=0,058$  ;  $F(1,6)=8,203$ ,  $p=0,151$ ) avec une augmentation du taux de discrimination de 0,62% lors du deuxième test (Test 1= 97,89% et Test 2= 98,51%). Cependant, cet effet n'est pas généralisable puisque

marginalement significatif seulement pour F1. Les autres effets principaux et d'interaction ne montrent pas de différence significative pour F1 et F2.

- Temps de réaction

Nous poursuivons par l'analyse du temps de réaction à l'aide d'une analyse de variances ANOVA à mesures répétées avec pour facteur intra-sujet le Test (Test1 vs. Test2) et pour facteurs inter-sujets le Groupe (1 visuel vs. 2 français vs. 3 musique instrumentale) et le statut musical (musicien vs. non-musicien). Nous rappelons que les temps de réaction sont exprimés en ms à partir de la fin du stimulus. Seuls les temps de réaction des réponses correctes ont été analysés (Pisoni & Tash, 1974 ; Tantibundhit *et al.*, 2012). Cette étape a éliminé 1,80% (121 occurrences) des réponses totales. De plus, nous avons également éliminé les temps de réactions supérieurs à deux écart-types de la moyenne pour chaque sous-groupe d'après une procédure menée par Schneider *et al.* (2011) et reprise par Tantibundhit *et al.*, 2012). Cette étape a éliminé 5,43% (358 occurrences) du total pour F1 et 4,44% (293 occurrences) du total pour F2 de l'ensemble des réponses correctes.

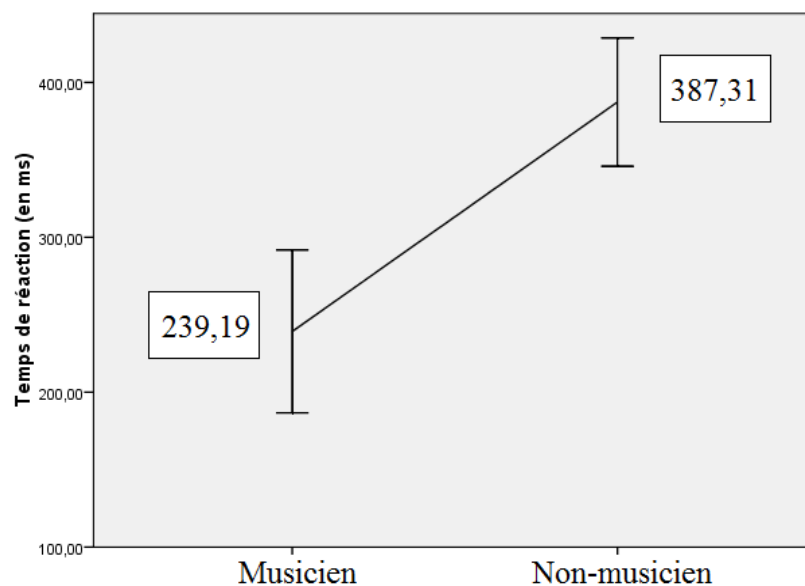


**Figure 56:** Diagramme du temps de réaction moyen (en ms) en fonction du groupe, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard.

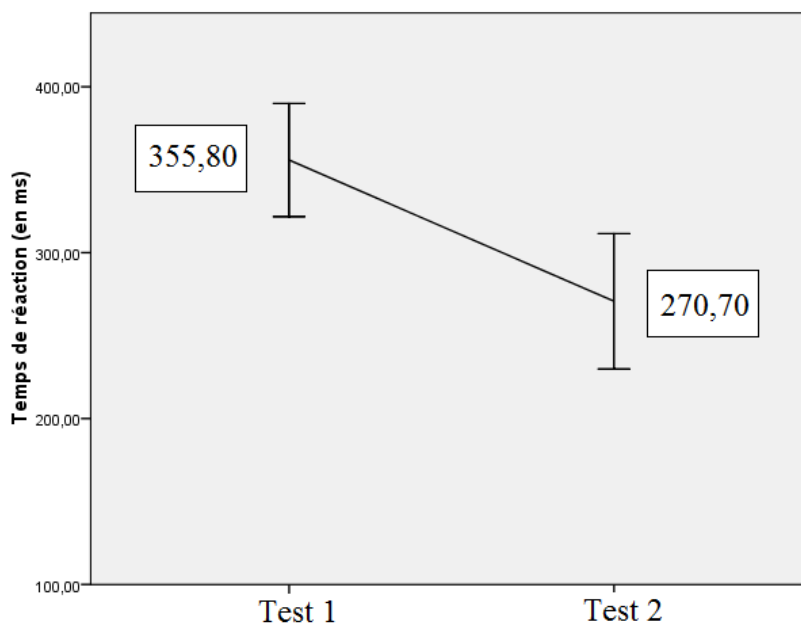
Temps de réaction (en ms)		Test 1	Test 2
Groupe 1 Visuel	Musicien	330,49 (159,35)	207,67 (213,66)
	Non-musicien	455,71 (175,33)	427,69 (279,55)
Groupe 2 Parole	Musicien	273,06 (316,25)	113,18 (251,58)
	Non-musicien	379,50 (140,47)	326,18 (240,16)
Groupe 3 Musique	Musicien	307,80 (206,66)	202,92 (157,29)
	Non-musicien	388,24 (85,10)	346,53 (91,94)

**Tableau 65:** Récapitulatif du temps de réaction moyen (en ms) en fonction du groupe, du test et du statut musical. Les écart-types sont donnés entre parenthèses.

Nous pouvons observer, de manière générale et d'après la Figure 56 et le Tableau 65, que l'ensemble des sous-groupes ont répondu en moyenne plus rapidement lors du test 2 de 85,10 ms (Test 1=355,8 ms et Test 2=270,69 ms). Nous pouvons constater, d'après notre analyse de variance qu'il n'y a un effet significatif que pour la mesure répétée du Test ( $F(1,24)=12,284$ ,  $p=0,002$  ;  $F(1,6)=233,315$ ,  $p<0,0001$ ) avec une diminution du temps de réaction de 85,11 ms lors du deuxième test (Test 1= 355,80 ms et Test 2= 270,69) et de l'effet principal du Statut ( $F(1,24)=4,381$ ,  $p=0,047$  ;  $F(1,6)=324,149$ ,  $p<0,0001$ ) avec un meilleur score par les musiciens de 148,12 ms (Musiciens= 239,19 ms et Non-Musiciens= 387,31 ms) (Figure 57 et Figure 58).

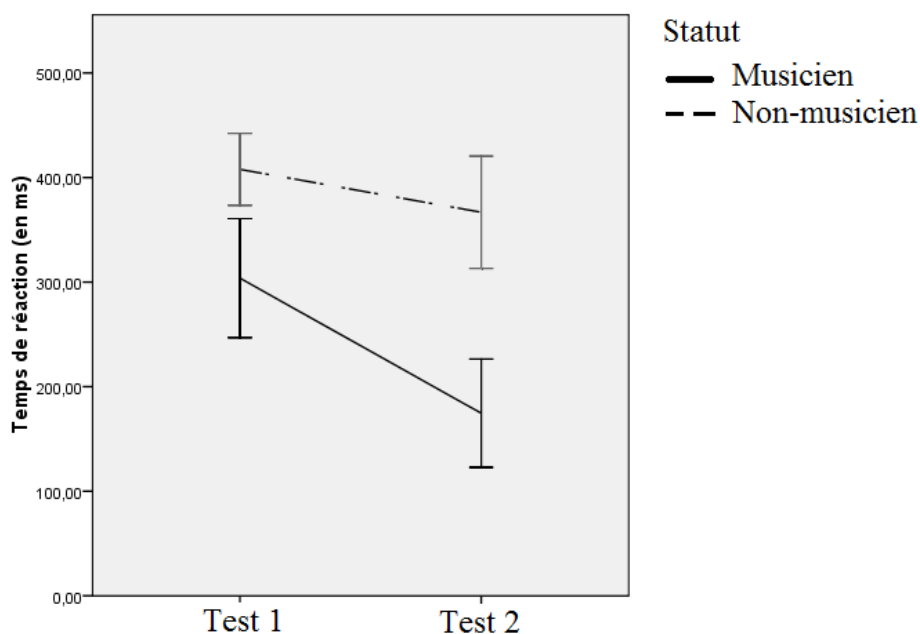


**Figure 57:** Impact du statut musical sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,24)=4,381$ ,  $p=0,047$  ;  $F(1,6)=324,149$ ,  $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard.



**Figure 58** : Impact du Test sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,24)=12,284$ ,  $p=0,002$  ;  $F(1,6)=233,315$ ,  $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard.

Nous constatons également une interaction marginalement significative Test X Statut ( $F(1,24)=3,297$ ,  $p=0,082$  ;  $F(1,6)=16,637$ ,  $p=0,007$ ) (Figure 59). En ce qui concerne cette interaction, nous observons une interaction ordonnée (Tableau 66).



**Figure 59** : Impact de l'interaction Statut X Test sur le temps de réaction moyen (en ms). La différence est marginalement significative d'après une analyse de variance à mesures répétées ( $F(1,24)=3,297$ ,  $p=0,082$  ;  $F(1,6)=16,637$ ,  $p=0,007$ ). Les barres d'erreur indiquent l'erreur standard.

	Test 1	Test 2	Effet intra du statut
Musicien	303,78	174,59	-129,19
Non-musicien	407,82	366,80	- 41.02
Effet intra du test	104,04	192,21	57,3

**Tableau 66** : Interaction ordonnés entre Test et Statut. Les moyennes sont données en ms.

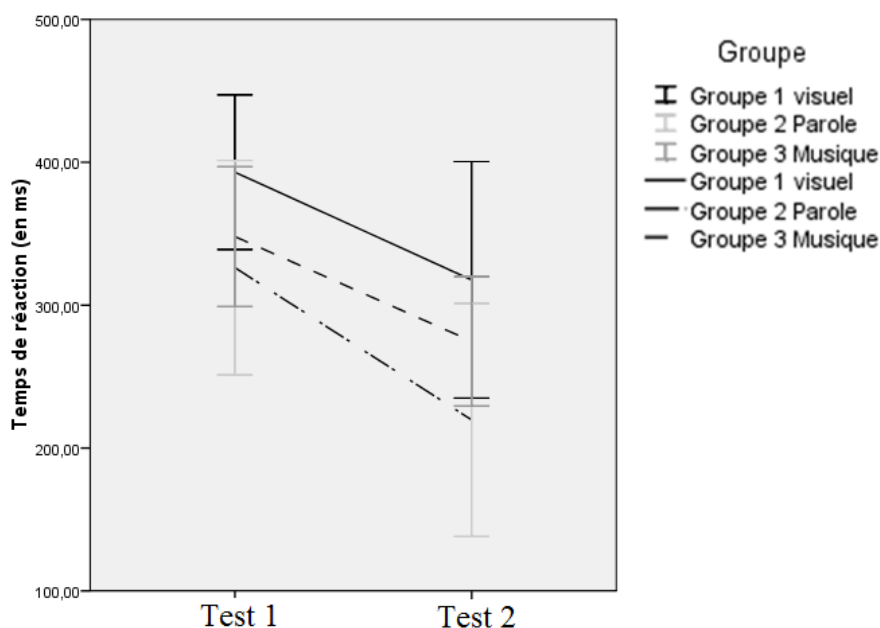
On observe une interaction entre les facteurs Statut et Test puisque les effets intra du Statut sont différents, en valeur absolue, pour les musiciens et pour les non-musiciens (respectivement, 121,28 et 63,98 ms) et pour les test 1 et 2 (respectivement 107,14 ms et 164,44 ms). Cette interaction est ordonnée puisque les effets intra sont de même signe. L'effet du facteur Statut est plus important pour les musiciens (-121,28 ms) que pour les non-musiciens (-63,98 ms). L'effet du Test est plus important pour le test 2 (+164,44) que pour le test 1 (+107,14).

	Statut	n=	moyenne	Test t	ddl	Valeur de p	<0,05 = *
Test 1	Musicien	15	303,78 (220,52)	-1,564	28	0,129	ns
	Non-Musicien	15	407,82 (133,16)				
Test 2	Musicien	15	174,59 (200,55)	-2,576	28	0,016	*
	Non-Musicien	15	366,80 (208,05)				

**Tableau 67** : Récapitulatif de l'analyse statistique test t pour échantillons indépendants concernant les différences entre musiciens et non-musiciens respectivement pour les tests 1 et 2.

Un test t pour échantillons indépendant (Tableau 67) a permis de mettre en évidence que la différence entre musiciens et non-musiciens n'est statistiquement pas significative lors du test 1 ( $t_{28} = -1,564$ ,  $p=0,129$ ) mais elle l'est lors du test 2 ( $t_{28} = -2,576$ ,  $p=0,016$ ).

D'autres résultats significatifs sont mesurés pour F2 mais non pour F1 et ne sont par conséquent pas généralisables du Groupe (F2 (2,12)=130,724,  $p<0,0001$ ), Groupe X Statut (F2 (2,12)=4,080,  $p=0,044$ ) et Groupe X Test (F2 (1,307 ; 7,84)=14,198,  $p=0,004$ ). Une analyse PostHoc a été menée afin de mesurer les différences significatives entre les trois groupes pour l'effet principal du Groupe et dans les interactions Groupe X Statut et Groupe X Test. Les résultats complets de cette analyse Posthoc sont donnés en annexes 49. Les trois mesures ont été effectuées de manière indépendante afin de limiter l'introduction de biais.



**Figure 60:** Impact de l'interaction Groupe X Test sur le temps de réaction moyen (en ms). La différence n'est pas significative pour F1 d'après une analyse de variance à mesures répétées ( $F1(2,24)=0,197$ ,  $p=0,823$  ;  $F2(1,307 ; 7,84)=14,198$ ,  $p=0,004$ ). Les barres d'erreur indiquent l'erreur standard.

Nous nous sommes focalisé sur l'interaction Groupe X Test (Figure 60) qui n'est significative qu'entre les groupes 1 et 2 ( $F2(1,6)=6,381$ ,  $p=0,045$ ) et entre les groupes 2 et 3 ( $F2(1,6)=161,699$ ,  $p<0,0001$ ) sachant que la différence entre les deux tests est d'autant plus importante (116,15 ms) pour le groupe 2 que le groupe 1 (90,26 ms) et que le groupe 3 (71,48 ms). De plus, lors du test 1, le groupe 2 a répondu plus rapidement que le groupe 1 (75,34 ms) et que le groupe 3 (31,49 ms) et il en va de même lors du test 2 avec respectivement 101,22 ms et 76,17 ms. Cependant, nous pouvons suggérer que cette tendance a été influencée par le fait que les musiciens du groupe 2 et lors du test 2 ont répondu plus rapidement que les autres groupes soit 112,35 ms contre 324,77 en moyenne pour les autres. Par conséquent, cet effet mesuré par items ( $F2(1,307 ; 7,84)=14,198$ ,  $p=0,004$ ) ne peut pas être généralisé à d'autres populations ( $F1(2,24)=0,197$ ,  $p=0,823$ ).

En résumé, nous avons pu voir que seuls trois effets sont significatifs et généralisables pour les items (F2) et les sujets (F1) pour le temps de réaction. Ces derniers sont l'effet principal du Statut (musiciens vs. non-musiciens) ( $F1(1,24)=4,381$ ,  $p=0,047$ ) ( $F2(1,6)=324,149$ ,  $p<0,0001$ ), celui du Test (Test 1 vs. Test 2) ( $F1(1,24)=12,284$ ,  $p=0,002$ ) ( $F2(1,6)=233,315$ ,  $p<0,0001$ ) et l'interaction Test X Statut ( $F1(1,24)=3,297$ ,  $p=0,082$ ) ( $F2(1,6)=16,637$ ,  $p=0,007$ ) même si l'on observe une significativité marginale pour F1. Cela signifie que les non musiciens comme les musiciens répondent plus rapidement au test 2 qu'au test 1, mais à un moindre degré pour les



non musiciens dont les temps de réaction sont supérieurs à ceux des sujets qui pratiquent la musique.

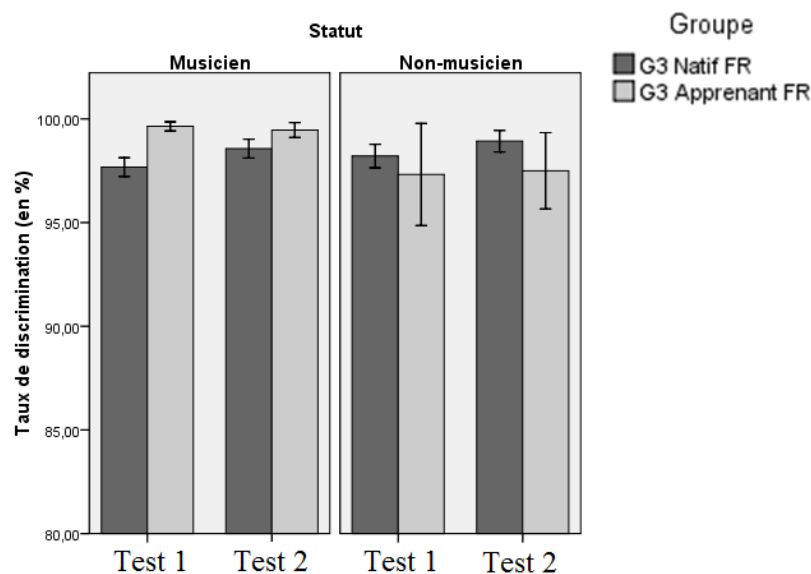
### 5.2.3 Apprenants vs. locuteurs natifs du français

Cette analyse prend en compte les données des 20 sujets répartis selon deux statuts langagiers (apprenant vs. natif du français : 10 sujets/statut langagier) et deux statuts musicaux (musiciens vs. non-musiciens : 10 sujets/statut musical soit 5 sujets/groupe/statut musical).

Afin de comparer les résultats obtenus pour notre groupe test (Groupe : 3 musique instrumentale) nous avons fait passer le même test à dix locuteurs natifs du français. Nous suivrons la même méthodologie d'analyse statistique que celle effectuée précédemment. Deux analyses de variance ANOVA à mesures répétées ont été effectuées l'une par sujets (F1) et la seconde par items (F2) dans le but de tester l'influence du statut musical (musicien vs. non-musicien) et du statut langagier (apprenants vs. natifs du français) sur le pourcentage de réponses correctes et sur les temps de réaction en tâche de discrimination AXB.

- Taux de discrimination

Nous commencerons par analyser le taux de discrimination à l'aide d'une analyse de variances ANOVA à mesures répétées avec pour facteur intra-sujet le Test (Test1 vs. Test2) et pour facteurs inter-sujets le Statut Langagier (apprenant vs. francophone natif) et le Statut Musical (musicien vs. non-musicien).



**Figure 61:** Diagramme du taux de discrimination moyen (en %) en fonction du statut langagier, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard.

Taux de discrimination (en %)		Test 1	Test 2
Natif FR	Musicien	97,68 (1,02)	98,57 (1,02)
	Non-musicien	98,21 (1,26)	98,93 (1,17)
App. FR	Musicien	99,64 (0,49)	99,46 (0,80)
	Non-musicien	97,32 (5,50)	97,5 (4,11)

**Tableau 68** : Récapitulatif du taux de discrimination (en %) en fonction du Statut langagier, du test et du statut musical. Les écart-types sont donnés entre parenthèses.

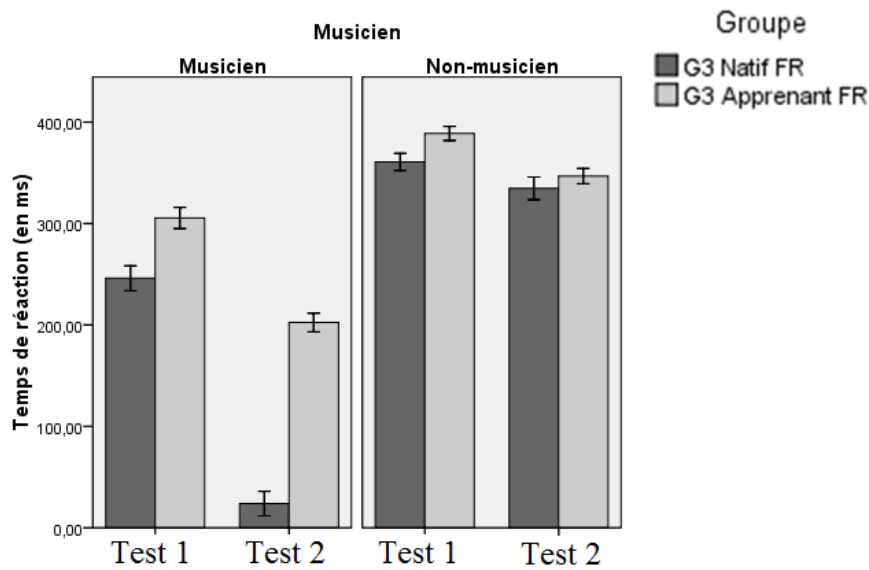
Nous observons, tout d'abord et d'après la Figure 61 et du Tableau 68, que sur l'ensemble des deux groupes d'apprenants et ce quel que soit leur statut musical (musicien vs. non-musiciens) et langagier (apprenants vs. natifs du français), nous observons un taux de discrimination, en moyenne, de 98,4% soit moins de 2% d'erreurs. Nous pouvons, cependant, constater que d'après notre analyse de variance qu'il n'y a aucun effet significatif pour les facteurs étudiés et ce pour F1 et F2.

- Temps de réaction

Nous poursuivons par l'analyser du temps de réaction à l'aide d'une analyse de variances ANOVA à mesures répétées avec pour facteur intra-sujet le Test (Test1 vs. Test2) et pour facteurs inter-sujets le Groupe (1 visuel vs. 2 français vs. 3 musique instrumentale) et le statut musical (musicien vs. non-musicien). Nous avons suivi la même procédure que pour la section précédente. Seuls les temps de réaction des réponses correctes ont été analysés. Cette étape a éliminé 1,6% (71 occurrences) des données du total. De plus, nous avons également éliminé les temps de réactions supérieurs à deux écart-types de la moyenne pour chaque sous-groupe. Cette étape a éliminé 4,6% (202 occurrences) du total pour F1 et 4,5% (200 occurrences) du total pour F2 de l'ensemble des réponses correctes.

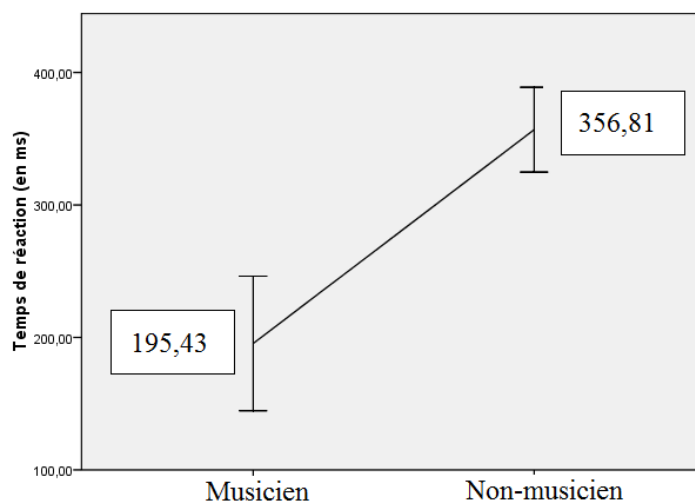
Temps de réaction (en ms)		Test 1	Test 2
Natif FR	Musicien	246,05 (169,12)	24,93 (125,99)
	Non-musicien	359,61 (99,80)	332,87 (152,95)
Apprenant FR	Musicien	307,80 (206,66)	202,92 (157,29)
	Non-musicien	388,24 (85,10)	346,53 (91,94)

**Tableau 69** : Récapitulatif du temps de réaction (en %) en fonction du Statut langagier, du test et du Statut musical. Les écart-types sont donnés entre parenthèses.

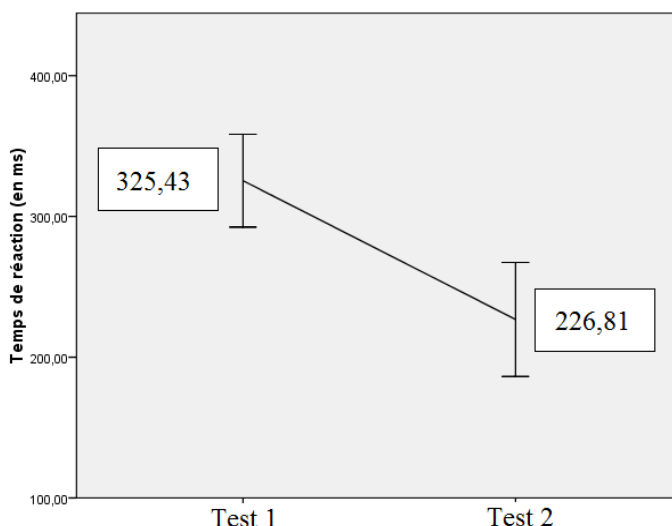


**Figure 62:** Diagramme du temps de réaction moyen (en %) en fonction du statut langagier, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard.

Les données sont récapitulées dans le Tableau 69 et la Figure 62. Nous pouvons constater, d'après notre analyse de variance qu'il n'y a un effet significatif que pour la mesure répétée du Test ( $F(1,16)=22,732$ ,  $p<0,0001$  ;  $F(1,6)=204,656$ ,  $p<0,0001$ ) avec une diminution du temps de réaction de 98,61 ms lors du deuxième test (Test 1 = 325,43 ms et Test 2 = 226,81 ms) et de l'effet principal du statut musical ( $F(1,16)=7,242$ ,  $p=0,016$  ;  $F(1,6)=324,990$ ,  $p<0,0001$ ) avec un meilleur score par les musiciens de 161,39 ms (Musiciens = 195,43 ms et Non-Musiciens = 356,81ms) (Figure 63 et Figure 64).

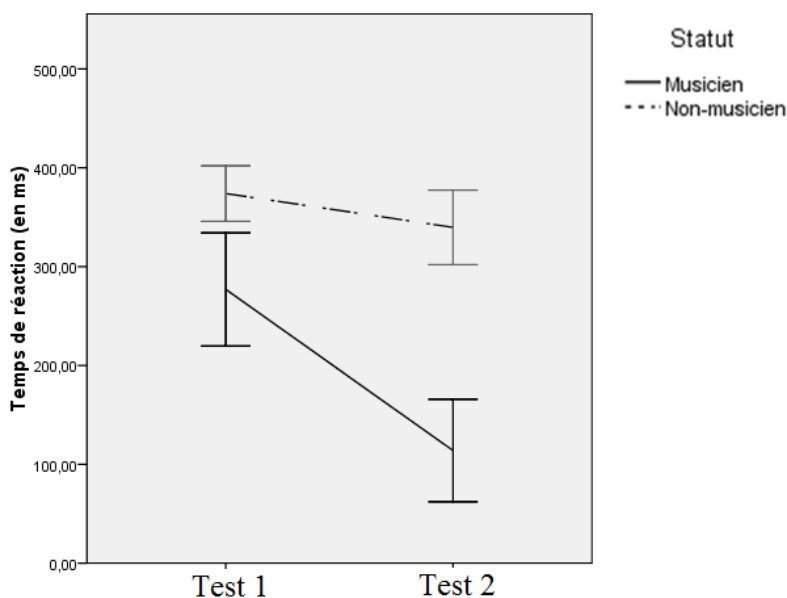


**Figure 63:** Impact du statut sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures ( $F(1,16)=7,242$ ,  $p=0,016$  ;  $F(1,6)=324,990$ ,  $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard.



**Figure 64 :** Impact du Test (à gauche) sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,16)=22,732$ ,  $p<0,0001$  ;  $F(1,6)=204,656$ ,  $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard.

Nous constatons également une interaction significative Test X Statut musical ( $F(1,16)=9,691$ ,  $p=0,007$  ;  $F(1,6)=38,221$ ,  $p=0,001$ ). D'autres résultats significatifs sont mesurés pour F2 mais non pour F1 et ne sont par conséquent pas généralisables : Statut Langagier ( $F(1,6)=220,406$ ,  $p<0,0001$ ), Statut Langagier X Statut Musical ( $F(1,6)=27,405$ ,  $p=0,002$ ) et Statut Langagier X Test ( $F(1,6)=70,966$ ,  $p<0,0001$ ).



**Figure 65 :** Impact de l'interaction Statut X Test sur le temps de réaction moyen (en %). La différence est marginalement significative d'après une analyse de variance à mesures répétées ( $F(1,16)=9,691$ ,  $p=0,007$  ;  $F(1,6)=38,221$ ,  $p=0,001$ ). Les barres d'erreur indiquent l'erreur standard.

Nous signalons que les musiciens du groupe de natifs montrent une tendance très forte vers un temps de réaction court comparé aux autres groupes. Cela pourrait s'expliquer par une stratégie d'écoute vers une comparaison AX seule et donc sans prendre en considération le stimulus B.

	Test 1	Test 2	Effet intra du statut
Musicien	276,93	113,93	-163,00
Non-musicien	373,93	339,70	-34,23
Effet intra du test	97,00	225,77	128,77

**Tableau 70:** Interaction ordonnés entre Test et Statut. Les moyennes des musiciens, non-musiciens, test 1 et test 2 sont données en ms.

A l'aide de la Figure 65 et du Tableau 70, on observe une interaction entre les facteurs Statut et Test puisque les effets intra du Statut sont différents, en valeur absolue, pour les musiciens et pour les non-musiciens (respectivement, 163,00 et 34,23 ms) et pour les test 1 et 2 (respectivement 97,00 ms et 225,7 ms). Cette interaction est ordonnée puisque les effets intra sont de même signe. L'effet du facteur Statut est plus important pour les musiciens (-163,00 ms) que pour les non-musiciens (-34,23 ms). L'effet du Test est plus important pour le test 2 (+225,77 ms) que pour le test 1 (+97,00 ms).

En résumé, nous avons pu constater qu'il n'y a pas d'effet significatif en ce qui concerne le taux de discrimination selon les différents facteurs étudiés. Nous constatons notamment que le taux moyen d'erreur est de moins de 2% et qu'il n'y a pas de différence significative entre les apprenants et les locuteurs natifs du français. Au regard du temps de réaction, nous avons retrouvé les effets significatifs démontrés lors de l'étude avec les apprenants seuls soit un effet significatif pour la mesure répétée du Test ( $F(1,16)=22,732$ ,  $p<0,0001$  ;  $F(1,6)=204,656$ ,  $p<0,0001$ ) avec une diminution du temps de réaction de 98,61 ms lors du deuxième test (Test 1 = 325,43 ms et Test 2 = 226,81 ms) et de l'effet principal du statut musical ( $F(1,16)=7,242$ ,  $p=0,016$  ;  $F(1,6)=324,990$ ,  $p<0,0001$ ) avec un meilleur score par les musiciens de 161,39 ms (Musiciens = 195,43 ms et Non-Musiciens = 356,81ms). Nous avons également pu remarquer une interaction significative Test X Statut musical ( $F(1,16)=9,691$ ,  $p=0,007$  ;  $F(1,6)=38,221$ ,  $p=0,001$ ) ce qui montre que les musiciens sont d'autant plus meilleur lors du test 2 qu'ils ne l'étaient lors du test 1 par rapport aux non-musiciens.

### 5.3 Analyses et discussions

Dans cette partie, nous discuterons des résultats obtenus dans cette étude et ce au regard de nos questionnements de départ. Nous noterons, dès à présent, qu'il n'a pas été possible de comparer nos résultats avec la littérature empirique existante en raison du faible nombre de travaux dans

le domaine. Les protocoles empiriques sur lesquels nous nous sommes appuyés portaient sur la perception des consonnes (Hallé, Chéreau & Segui, 2000 ; Hallé & Best, 2007 ; Tantibundhit *et al.*, 2012 ; parmi d'autres). D'autres travaux portent sur les sons vocaliques mais ne prennent en considération que le taux de discrimination (Tsukada, 2012 ; Chan, 2013 ; Best *et al.*, 2003 ; parmi d'autres). De plus, aucun document publié ne porte sur des groupes linguistiques hétérogènes.

### 5.3.1 Taux de discrimination et traitement acoustique

Nous allons commencer par traiter le taux de discrimination, soit le nombre de réponses correctes données. Nous avons pu constater que, quel que soit le groupe d'apprenants ou bien de locuteurs natifs du français, nous avons pu mesurer un très grand nombre de réponses correctes, soit 98,3% ce qui équivaut à moins de 2% d'erreurs. Nous n'avons pu montrer aucun effet significatif à l'exception d'un effet marginalement significatif du statut musical pour les apprenants du français ( $F(1,24)=4,015$ ,  $p=0,057$  ;  $F(1,6)=64,751$ ,  $p=0,036$ ) avec les musiciens ayant un meilleur taux de discrimination que les non-musiciens de 0,62%. Ces premiers résultats nous amènent à plusieurs considérations. Tout d'abord, ils confirment notre éclairage théorique émergentiste et socioconstructiviste car il montre que les apprenants de français n'ont significativement pas plus de difficulté, en moyenne, que les francophones natifs à discriminer les phonèmes du français lorsque nous testons leur capacité à comparer acoustiquement deux sons à l'aide d'un ISI court. Cela signifie que le traitement ascendant des informations sonores n'est pas directement altéré par le processus phonologique lié que l'on connaît déjà. Etant donné que l'input est inférieur à 2 secondes, les informations acoustiques sont toujours présentes dans la mémoire de travail et il est donc possible de les comparer entre elles. La mémoire à long terme correspondant aux représentations phonologiques est peu activée ou du moins ces représentations ne sont pas prédominantes dans le choix de la réponse. Il est donc clair que la notion de « surdité » n'est pas pertinente puisque les apprenants ne sont pas sourds à une différence mais que c'est la nativisation/assimilation phonologique, soit une habitude préférentielle de traitement phonologique (descendant), qui prédominerait sur un traitement acoustique voire phonétique (ascendant). Il semblerait cependant que les musiciens aient un score légèrement meilleur que les non-musiciens ce qui pourrait aller en faveur de l'hypothèse que les capacités des musiciens à traiter l'objet sonore seraient un avantage dans une tâche de discrimination de sons d'une langue étrangère. Il serait dès lors intéressant de mener une étude complémentaire afin de comparer les scores d'apprenants à des tests AXB en

faisant varier l'ISI dans le but de montrer des parallèles ou des asymétries entre les traitements acoustique et phonologique. Cela permettrait notamment de montrer des types de profil d'individus sur leurs capacités de discrimination sur le plan acoustique d'une part, et sur le plan phonologique d'autre part. Nous pourrions, dès lors, mieux comprendre comment le processus de perception des sons de langues étrangères émerge entre des pressions psychoacoustiques et phonologiques. Sur le plan didactique, ce pourrait devenir un outil dans l'évaluation initiale des apprenants afin de prendre en compte des profils individuels différents.

### **5.3.2 Effets sur le temps de réaction**

L'analyse des temps de réaction semble permettre d'aller plus loin dans la réflexion. Concernant les apprenants du français, nous avons pu remarquer trois effets significatifs dont celui du Test ( $F(1,24)=12,284$ ,  $p=0,002$  ;  $F(1,6)=233,315$ ,  $p<0,0001$ ), du statut ( $F(1,24)=4,381$ ,  $p=0,047$  ;  $F(1,6)=324,149$ ,  $p<0,0001$ ) et une interaction marginalement significative Test X Statut ( $F(1,24)=3,297$ ,  $p=0,082$  ;  $F(1,6)=16,637$ ,  $p=0,007$ ). Les musiciens ont répondu plus rapidement que les non-musiciens de 148,12 ms, les trente sujets ont répondu plus rapidement lors du test 2 de 85,11 ms et nous avons pu mettre en avant une interaction ordonnée (Statut X Test) montrant que les musiciens répondent plus rapidement lors du test 2 que lors du test 1 (-121,28 ms) par rapport au non-musiciens (63,98 ms) et que les musiciens répondent plus rapidement que les non-musiciens au test 2 qu'au test 1 (respectivement 164,44ms et 107,14 ms). Ces résultats confirment la tendance vue lors de l'analyse du taux de discrimination, à savoir que les musiciens répondraient à ce genre de tâche de manière plus précise mais également plus rapide en moyenne. Cela sous-tendrait une capacité à prendre une décision plus rapidement que les non-musiciens et que donc leur traitement acoustique de l'information sonore serait plus efficace.

Une nouvelle donnée nous permet de constater qu'en général, les apprenants répondent significativement plus rapidement lors du test 2 que lors du test 1 de 85,10 ms. Cette tendance va dans le sens de nombreux travaux dont deux de Bygate (2001) et Bei (2013) qui révèlent que la répétition rapide d'une même tâche montre toujours une progression sur le plan qualitatif et/ou quantitatif. Pour notre étude, nous avons pu montrer une progression significative pour le temps de réaction. Cette tendance pourrait s'expliquer par plusieurs facteurs certainement combinés dont une habitude cognitive à la tâche, une habitude à la voix du locuteur des stimuli et une habitude aux contrastes phonémiques proposés. Cette habitude serait d'autant plus forte que le laps de temps entre les deux tests était faible (moins de 10 minutes).

Cependant, nous avons également pu montrer que la différence entre les musiciens et non-musiciens augmente lors du test 2 par rapport au test 1 de manière significative. Nous pouvons donc penser que les musiciens seraient plus sensibles à cette habitude par rapport aux non-musiciens pour plusieurs raisons : une endurance cognitive auditive plus importante, efficace et une appétence pour ce type d'exercice auditif. Cela irait également dans le sens de la littérature scientifique actuelle pour laquelle l'entraînement musical a des effets sur d'autres capacités auditives que les capacités musicales (Kraus & Chandrasekaran 2010 : section 1.2.1).

En comparant les résultats du groupe 3 avec ceux des locuteurs natifs du français en ce qui concerne le temps de réaction, nous constatons les mêmes résultats que ceux obtenus pour les apprenants du français soit un effet significatif du Test ( $F(1,16)=22,732, p<0,0001$  ;  $F(1,6)=204,656, p<0,0001$ ) et de l'effet principal du statut musical ( $F(1,16)=7,242, p=0,016$  ;  $F(1,6)=324,990, p<0,0001$ ) ainsi qu'une interaction significative Test X Statut musical ( $F(1,16)=9,691, p=0,007$  ;  $F(1,6)=38,221, p=0,001$ ). Une fois de plus, les musiciens répondent plus rapidement que les non-musiciens (161,39 ms) : dans l'ensemble les sujets répondent plus rapidement au test 2 qu'au test 1 (98,61 ms) et les musiciens répondent d'autant plus rapidement lors du test 2 que du test 1 (163 ms) par rapport au non-musiciens (34,23 ms) et que les musiciens répondent d'autant plus rapidement que les non-musiciens au test 2 qu'au test 1 (respectivement 225,77 ms et 97 ms). Cela vient donc renforcer la tendance montrée pour les apprenants du français seuls. Cependant, il n'a pas été possible de montrer de différences significatives entre le groupe d'apprenants et celui de locuteurs natifs du français que ce soit pour le taux de discrimination ou le temps de réaction. Il semblerait donc, une fois de plus, que la capacité à comparer acoustiquement deux sons de parole ne dépende pas de l'existence ou non de ces sons dans la/les langue(s) première(s) mais d'autres facteurs comme la répétition de la tâche ou le statut musical. Cela vient appuyer les travaux de Werker et Tee (1984a : 1876) sur un double traitement (simultané) de la parole au niveau phonétique et phonémique et seul le niveau phonémique serait impacté par la/les L1(s).

### 5.3.3 Impact de la tâche

Nous n'avons pu constater également aucune interaction significative Groupe X Test pour F1 et F2 ce qui confirme le fait qu'aucune des tâches n'a eu d'impact significatif sur la progression entre le test 1 et le test 2. Il se pourrait donc que l'impact de la musique instrumentale ne se réalise qu'à travers un entraînement musical (notre cohorte de musiciens) et non à travers une exposition de courte durée. Cela irait dans le sens de la notion d'expertise (pratique) qui



nécessite du temps afin d'enclencher des modifications importantes cérébrales. Gladwell (2008 : 40) propose que cette expertise ne serait atteinte qu'après au moins 10 000 heures de pratique. Nous noterons, cependant, que ce chiffre est ouvertement critiqué notamment par K. Anders Ericsson, professeur de psychologie à l'université d'état de Floride, dans une lettre qu'il adresse en 2012 à l'*APS Observer*<sup>78</sup> dans laquelle il dénonce une « généralisation provocatrice » de Gladwell. Selon lui, la notion d'expertise est bien plus complexe et intègre des prédispositions physiologiques voire cognitives, ainsi que la pratique délibérée, soit la pratique en autonomie. Néanmoins, et au-delà de la notion d'expertise, comme aucun effet d'interaction lié au groupe n'a pu être mis en avant, il est possible que les conditions expérimentales aient limité des effets potentiels. Nous pouvons penser, d'une part, que la durée d'exposition aux documents n'était pas assez longue (environ 3 minutes) pour enclencher des effets cognitifs qui auraient pu se traduire sur le plan comportemental. D'autre part, afin de contrôler la perception des documents, et ce en écho aux pré-études 1 et 2, nous avons mis en place un court questionnaire à la suite de l'exposition au document. Or ce dernier, proposé sous forme écrite, activait forcément un traitement langagier à travers les consignes en français / anglais et le fait de devoir trouver cinq mots en lien au document. Cette activation aurait pu limiter l'impact des activations en œuvre lors de l'exposition aux documents visuels ou sonores. Il serait intéressant dès lors de reproduire cette expérience en augmentant la durée d'exposition à l'input et en supprimant le questionnaire afin d'augmenter l'influence de l'exposition sur la tâche de discrimination.

Nous souhaitons également signaler, pour terminer, que nous ne négligeons pas le fait que les auditeurs sont principalement des individus avec un fort potentiel sur le plan métalinguistique de par leur profil académique, socio-professionnel et langagier. Ceci a pu fortement influencer le taux de discrimination important et le temps de réaction. Nous ne pouvons donc pas prédire si les tendances mesurées seraient reproductibles avec des apprenants en difficulté sociale comme dans des publics en alphabétisation, illettrisme ou en contexte de français d'intégration. Ces résultats sont nouveaux dans le domaine et, nous l'espérons, pourront être complétés par d'autres recherches empiriques.

---

78 <http://www.psy.fsu.edu/faculty/ericsson/2012%20Ericssons%20reply%20to%20APS%20Observer%20article%20Oct%2028%20on%20web.doc> (consulté le 05/08/14)

## 5.4 Conclusion générale de l'expérimentation 2

Au terme de cette étude psychoacoustique, nous avons pu répondre à nos questionnements de départ, à savoir (i) les différences entre musiciens et non-musiciens sur un test de discrimination AXB se focalisant sur le traitement acoustique, (ii) l'impact d'une exposition à un input de musique instrumentale sur la progression entre la répétition d'un test de discrimination AXB et finalement (iii) les différences entre apprenants et natifs du français à ce genre de test. Nous avons pu constater que l'ensemble des sujets montre un fort taux de discrimination ce qui confirmerait, si les résultats étaient confirmés dans d'autres études, que la nativisation/assimilation est un processus purement phonologique et non acoustique. Nous avons également pu montrer à travers le temps de réaction que de manière globale, les sujets ont progressé entre les deux tests et que les musiciens ont répondu plus rapidement que les non-musiciens. Ce point est d'autant plus intéressant que nous avons mesuré un effet significatif d'interaction entre le Statut musical (musicien vs. non-musicien) et le Test (test 1 vs. 2). Cela signifie que lors du test 2, la différence entre les musiciens et les non-musiciens était plus importante qu'elle ne l'était lors du test 1. Ce point viendrait confirmer sur le plan comportemental que l'entraînement musical aurait des effets sur d'autres capacités auditives que celles liées à la musique, et en ce qui nous concerne, la parole. Cette étude nous a également permis de montrer qu'une exposition courte à une musique instrumentale n'a pas d'effet significatif sur la progression à la répétition d'un test de discrimination AXB. Si un effet était probable, il aurait pu être minoré par certaines caractéristiques empiriques ce qui nous amène à souhaiter reproduire cette étude en modifiant la durée d'exposition aux documents. Pour terminer et dans le prolongement du premier résultat, nous n'avons pu démontrer de différences significatives entre les apprenants et les locuteurs natifs du français. Ce point vient renforcer le fait que la notion de « surdité » n'est sans doute pas pertinente car sur le plan acoustique (les apprenants ayant des scores comparables à des locuteurs natifs du français) mais que c'est le niveau phonologique, soit un traitement préférentiel selon les langues que l'on connaît/utilise qui vient perturber ce traitement. Le symbolisme sous-jacent à la notion de « surdité » ne correspond pas au processus phonologique en œuvre qui perturbe le traitement acoustique mais qui « ne rend pas sourd ». Ce symbolisme tend souvent vers la représentation de difficultés « d'oreille » plutôt que de remettre en cause les automatismes cognitifs. Cependant, les différences entre musiciens et non-musiciens sont persistantes mêmes chez les locuteurs natifs du français ce qui peut être une piste prometteuse en faveur de l'entraînement musical comme moyen facilitateur du traitement acoustique. Il reste cependant à définir précisément par de

nouvelles recherches en didactique comment cet aspect « facilitateur » pourrait être intégré dans des formations.

**Synthèse de positionnement :**

Cette étude a été un catalyseur de recul épistémique quant à l'un des points centraux de ma thèse, soit la pluri-inter-transdisciplinarité. En effet, concevoir cette étude a nécessité un lent processus de formation personnelle (acoustique, phonétique, psycholinguistique, statistiques) mais également de confrontation entre deux domaines : la didactique des langues et la phonétique. En tant que chercheur, j'ai fait le choix, pour cette étude, de mettre le curseur vers les exigences empiriques de la phonétique au détriment de la conception d'implications didactiques des résultats obtenus. Néanmoins, j'ai pu prendre conscience des exigences particulières de chaque paradigme qui nécessitent de faire des choix expérimentaux. C'est certainement la partie de ma thèse qui a été à l'origine du plus grand nombre de questionnements et de remises en question à travers les obstacles et incompréhensions de mes choix pluri-inter-transdisciplinaires par des pairs. Néanmoins, les résultats de cette étude vont dans le sens de mon cadre théorique émergentiste et socio-constructiviste au regard du taux de discrimination des apprenants du français. Cela renforce l'idée de processus ascendants et descendants et la prévalence de l'un sur l'autre dans certains cas (la nativisation/assimilation par exemple). J'ai pu également montrer que les résultats en neurocognition sur la différence entre musiciens et non-musiciens se retrouvent sur le plan comportemental. Il reste cependant à faire le lien avec l'application didactique et notamment sur les possibilités pour un développement langagier. C'est à ce titre que l'étude 3 de notre thèse portera sur la mise en place d'un dispositif de médiation de perception et de production des sons du français en contexte FLE/S à l'aide d'outils de la musique instrumentale. Il est difficile en l'état actuel des connaissances et du peu de travaux en didactique des langues de proposer des applications concrètes. Il faudrait pour cela montrer que les différences entre musiciens et non-musiciens se retrouvent avec un test de discrimination avec un ISI long (supérieur à 1 seconde) et que cela peut impacter sur la production. Cependant, cela nécessitera de mieux comprendre les asymétries entre perception et production. Nous faisons l'hypothèse ici que cette relation est multiple et qu'elle pourrait faire émerger différents profils d'individus.

## 6. Etude 3 : didactique quasi-expérimentale et interventionniste

### Résumé

Cette étude a pour but de tester une approche pédagogique concrète de l'utilisation de la musique instrumentale dans un dispositif d'apprentissage de la prononciation du français en contexte de langue étrangère ou seconde. Elle permet notamment d'apporter au domaine une approche scientifique et empirique à des pratiques pédagogiques souvent peu questionnées ou remises en question (cf. Speelingua.com). Nous tenterons notamment de déterminer le caractère potentiellement développemental de ce dispositif tout en nous focalisant plus précisément sur l'intérêt d'une association entre note de musique et aperture dans la médiation de la production des phonèmes vocaliques du français. Cette étude repose notamment sur la notion de point d'abstraction (Barth, 2004) qui permettrait une meilleure abstraction des points saillants acoustiques des phonèmes vocaliques du français (Gendrot et *al.*, 2008).

Cette étude en didactique interventionniste et quasi-expérimentale s'est déroulée dans le cadre d'un cours de prononciation au diplôme universitaire d'études française (DUEF1) à l'Université Sorbonne-Nouvelle Paris 3 dans lequel nous avons retenu les données de 27 apprenant-es de profils hétérogènes correspondant à la diversité du public d'apprenants de français à Paris. La formation s'est déroulée sur douze semaines dont quatre pour la partie quasi-expérimentale. L'étude comprend trois questionnaires et deux enregistrements. Les 27 apprenants étaient répartis en trois groupes pour la partie quasi-expérimentale. Dans le groupe 1, les apprenants devaient prononcer les deux continua soit « i e ε a » et « u o ɔ a » de manière spontanée avec pour seul support le trapèze vocalique. Les groupes 2 et 3 devaient prononcer les mêmes continua mais les apprenants disposaient d'un *input* auditif des continua vocalique respectivement avec une F0 plate et une F0 corrélée à l'aperture et utilisation d'un piano. Les productions obtenues ont été analysées à travers un test d'identification par dix locuteurs natifs du français.

Les trois questionnaires montrent de manière générale une tendance légèrement positive sur l'impact du dispositif sur la prononciation du français des apprenants. Cette tendance est d'autant plus forte lors du questionnaire 3 final dans lequel l'ensemble des 27 apprenants déclarent qu'ils ont perçu des changements dans leur prononciation depuis le début du cours de prononciation alors qu'ils n'étaient que 23 lors du questionnaire 2. Une analyse statistique (test t pour échantillons appariés) révèle, en moyenne, une progression perçue significative entre les questionnaires 1 et 2 au regard de la prononciation générale mais aussi de la prononciation des voyelles, des consonnes et de la mélodie du français. Il semblerait également que les apprenants aient été plus sensibles au travail de

médiation de la production en groupe plutôt que la partie théorique et ce à la fois pour la compréhension et l'amélioration de la prononciation du français. Ils déclarent aussi être plutôt d'accord sur le fait de se sentir plus à l'aise avec les symboles phonétiques et que ces derniers les ont aidé à mieux percevoir et mieux prononcer les voyelles du français. Concernant notre groupe test, soit le groupe 3, nous avons pu observer que l'association entre notes de piano et voyelles affiche dans les représentations des apprenants une tendance bimodale lors du questionnaire 2 et qui se retrouve dans le questionnaire 3.

Une étude acoustique a pu permettre de montrer que statistiquement les phonèmes produits par les groupes 2 et 3, ayant eu un stimulus auditif avant la production, sont plus longs que ceux produits, en moyenne, par le groupe 1. Cela tend à démontrer que les apprenants des groupes 2 et 3 ont bien pris en compte les stimuli qui leur ont été proposés. Cependant, au regard du groupe 3, il n'est pas possible de montrer une tendance unique vers la production des phonèmes avec la F0 attendue et ce certainement pour des raisons psychoaffectives qui ont été mises en évidence à travers les questionnaires.

D'après l'analyse perceptive des productions des apprenants par dix auditeurs natifs du français, il a été possible de mettre en évidence plusieurs phénomènes. Premièrement, nous avons pu montrer, de manière significative, que la grille choisie influence le nombre moyen d'identifications correctes, quelle que soit la modalité de production. Le taux d'identification correcte pour la grille « archiphonème » est en moyenne supérieur d'environ 13% à celui de la grille « extrême ». De manière générale et d'après la grille « archiphonème », nous remarquons que la plupart des phonèmes sont reconnus au-delà de 80% à l'exception du /u/ qui est reconnu à 63,7% et confondu à 26,5% avec le /y/. De plus, d'après cette analyse, la production en continuum n'a pas été une aide articulatoire ou cognitive dans la réalisation canonique des phonèmes vocaliques du français standard : par conséquent, cette réalisation pourrait induire des confusions supplémentaires à celles qui pourraient apparaître à l'isolée.

Enfin, il semblerait dès lors, d'après notre protocole expérimental et notre dispositif d'apprentissage, qu'il y ait une potentialité acquisitionnelle de l'association proposée entre phonèmes et notes de musique ; néanmoins cette dernière montre une tendance bimodale et peut ne pas convenir à tous les profils d'apprenants. De plus, d'après notre protocole, il n'a pas été possible de montrer que cette association avait plus de potentialité développementale que la simple présentation du trapèze vocalique.

Pour terminer, une analyse restreinte, pour des raisons de problèmes techniques, de la progression entre les pré- et post-tests de deux apprenants par groupe n'a pas permis de montrer statistiquement

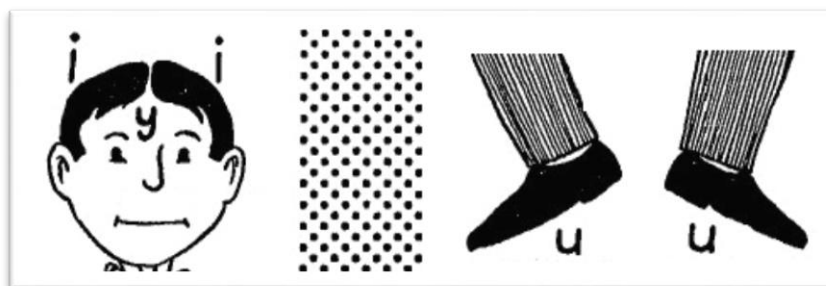
une progression entre les deux tests. Cependant, nous avons pu constater qu'en moyenne, les phonèmes ont été mieux identifiés lors du deuxième enregistrement à l'isolée (+7,1%) et en continuum (+7,5%) pour la grille « extrême », et respectivement +5% et 4% pour la grille « archiphonème ». Ceci est d'autant plus intéressant que les apprenants qui montrent une tendance de progression (+10 identifications correctes en moyenne) est, de manière absolue, plus forte que la tendance de ceux qui affichent une régression (-2 identifications correctes en moyenne).

En conclusion, cette étude a permis de questionner le dispositif d'apprentissage testé. D'une part, nous avons vu que sur le plan des représentations, les apprenants ont perçu une progression et cette dernière a pu être légèrement montrée à travers le sous-groupe d'apprenants dont les enregistrements pré- et post-tests étaient analysables. De manière plus intéressante et au regard du groupe 3, notre groupe test, nous avons pu montrer que l'association proposée entre note de musique et aperture des phonèmes vocaliques du français montre une tendance bimodale dans son utilité dans les représentations des apprenants. Cette tendance pourrait être expliquée par deux éléments : l'un par le sentiment perçu d'utilité pédagogique et l'autre dû à des pressions psychoaffectives (stress, etc.) qui peuvent être vécues dans un exercice de l'intime qu'est la voix chantée liée à la production d'une F0 particulière. Il semblerait, au regard de l'ensemble de ces éléments, que cette association apparaîtrait, dès lors, comme un outil supplémentaire à l'enseignant dans la prise en compte de la pluralité des profils d'apprentissage.

### 6.1 Méthodologie et questions de recherche

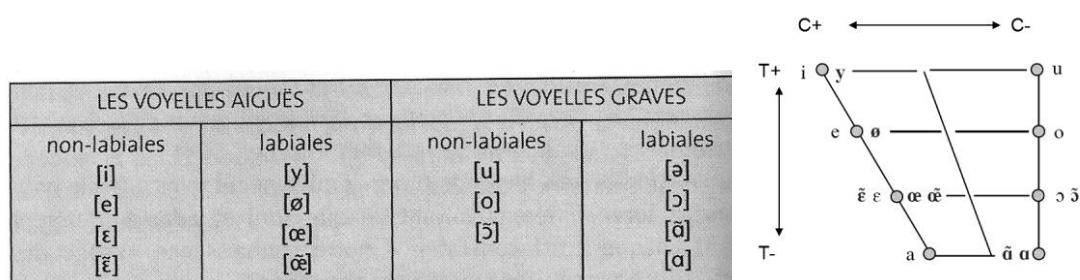
Depuis l'approche communicative à la fin des années 70 (section 3.1.1), le changement de paradigme dans la didactique des langues, considérant le langage du point de vue de son utilisation uniquement, a induit la réduction programmée des études sur l'enseignement / apprentissage de la prononciation en contexte FLE/S. Actuellement, nous proposons de placer les pratiques pédagogiques dans le domaine sur un continuum entre les approches « traditionnelles » issues de la méthode audio-orale et structuro-globale audiovisuelle proposant des exercices (drills) articulatoires ou auditifs ; et les approches « en marge » qui proposent un travail basé sur la prise de conscience du corps (yoga), la distanciation avec sa production sonore (théâtre, chanson). Certaines de ces dernières ont tenté de faire du lien entre les phonèmes vocaliques du français et des items non-linguistiques. Pour exemple, le bonhomme sonore (Guimbretière & Kaneman-Pougatch, 1991 : 70) associe les phonèmes vocaliques (oraux et nasals) et la spatialité corporelle (tête, front, pieds, etc.) : un /i/ est représenté au niveau

du haut du crâne car décrit comme « aérien » contrairement au /u/ représenté au niveau des pieds car « terrestre ». Ces associations sont pour la plupart basées non pas sur des corrélats acoustiques mais sur une approche psycholinguistique de la représentation possible de ces sons par les individus. Certains des choix sont reliés à des caractéristiques acoustiques /i/ versus /u/ avec respectivement un renforcement d'énergie dans les hautes et basses fréquences mais d'autres sont plus difficilement défendables sur le plan acoustique tel que le /y/ au niveau du front (Figure 66).



**Figure 66:** Schéma simplifié du Bonhomme Sonore pour les voyelles /i/, /y/ et /u/ (d'après Guimbretière & Kaneman-Pougatch, 1991 : 70).

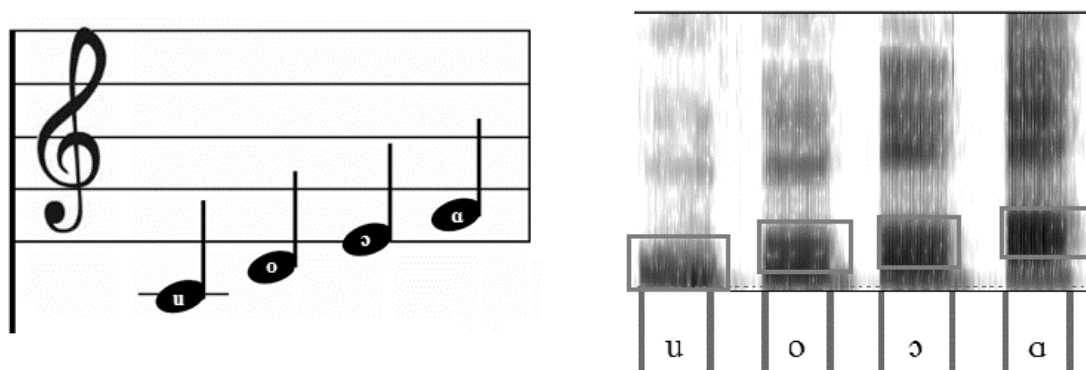
Dans cette perspective, Callamand (1981) parle d'acuité comme un trait perceptuel (aigu versus grave) qui dépendrait des deux autres traits fondamentaux que sont la labialité et la tension. De ce postulat, il serait possible de classer les différents sons vocaliques du français en deux classes : les aigus et les graves (Figure 67). Une même typologie a été proposée pour les sons consonantiques.



**Figure 67:** Acuité des sons vocaliques du français standard (Lauret, 2011 : 110 ; d'après Guimbretière & Kaneman-Pougatch, 1991:8) à gauche et (Billières, 2005), à droite. « C » correspond à la caractéristique « clair(C+)/sombre(C-) » et « T » à la tension.

La notion d'acuité fait référence à celle de hauteur spectrale qui est liée à l'enveloppe spectrale d'un son soit la quantité et la répartition relative des harmoniques de ce dernier. Cette répartition est le produit des résonateurs qui agissent comme un filtre dans la théorie source-filtre (section 1.3.2 ; Fant, 1960 : 15). De ce point de vue-là, il est possible de proposer une classification en fonction de la répartition de l'énergie des sons vocaliques du français standard et notamment

des voyelles dites focales (section 1.3.2) comme ci suit : (aigu) /i/ > /y/ > /ɑ/ > /ɔ/ > /o/ > /u/ > (grave). Ce point est généralement donné par le centre de gravité spectral des voyelles (CGS) et qui correspond au centre de gravité du spectre du signal. Le CGS est communément associé à la notion de « brillance » (Gordon & Grey, 1978) du son sur le plan perceptif. Plus le CGS d'un son est élevé plus le son est qualifié de « brillant » (Kim, Moreau & Sikora, 2005). Il ne faut néanmoins pas confondre la hauteur spectrale avec la hauteur tonale qui est la conséquence de la vibration des plis vocaux et qui détermine ce que l'on appelle communément la prosodie (ou l'intonation) mais dont la notion regroupe également d'autres éléments comme le rythme et l'accentuation. Nous noterons, cependant, que les voyelles ont une hauteur mélodique intrinsèque qui dépend principalement du degré d'ouverture (Hombert *et al.*, 1979 : 42; Whalen et Levitt, 1995 : 349-66). Les voyelles focales du français (section 1.3.3) auraient donc une particularité acoustique qui rendrait de manière saillante une région fréquentielle en termes de perception. Il est admis que « *les formants proches sont intégrés d'un point de vue perceptif en un pic simple* » (dans Gendrot et al, 2008 : 1 d'après Chistovich & Lublinskaja, 1979). Comme nous l'avons vu dans la section 3.3.2 et rappelé ci-dessus, de nombreuses approches pédagogiques lient les objets sonores de la parole à d'autres modalités (corps, chant, etc.) et ce que nous proposons ici est une association entre notes de musique et phonèmes vocaliques, où chacun de ces derniers seraient accompagnés d'une note selon son degré d'aperture. Notre hypothèse de travail est donc que corrélérer la perception de la hauteur spectrale à celle tonale pourrait servir de point d'appui à l'appropriation de certains sons vocaliques (cf. les voyelles focales postérieures arrondies du français standard) (Figure 68).



**Figure 68:** Corrélation arbitraire proposée entre la hauteur tonale des notes de musique et l'acuité des sons vocaliques postérieurs arrondis du français standard. A gauche : notation musicale ; à droite : spectrogramme (en abscisse : le temps et en ordonnée : la fréquence en Hz (min =0 Hz ; max = 5000 Hz) des quatre voyelles postérieures du français.



Nous tenons à préciser que cette hypothèse s'inscrit dans la proposition du bonhomme sonore de Guimbretière, c'est-à-dire une association arbitraire mais néanmoins motivée par une réflexion théorique et pédagogique. Cette réflexion repose sur la notion de « point d'appui à l'apprentissage » (Barth, 2004 ; section 3.1.3) qui explique la nécessité de faire des ponts conceptuels afin de favoriser l'abstraction des notions développées. Dans notre cas, le point d'appui serait d'ordre perceptif. Nous pouvons signaler également que cette association peut rappeler la stratégie opérée dans les langues sifflées. Rialland (2005) a notamment pu montrer chez un locuteur de langue sifflée qu'il montrait la réalisation d'un « contraste fondamental entre les voyelles « aigues » et « grave » correspondant à une opposition entre les voyelles antérieures *versus* centrales et postérieures en espagnol » (*Ibidem* : 241).

Dans notre étude et à travers ce cadre, nous tenterons de mettre en corrélation deux éléments acoustiques (hauteur spectrale et tonale) en supposant que cette association permettra une meilleure abstraction des particularités acoustico-phonétiques et favorisera donc une progression positive en production. Cette recherche s'intégrera dans un paradigme de didactique quasi-expérimentale : trois groupes expérimentaux mais dont le profil des participants n'a pas été contrôlé ; et interventionniste, puisque je serai à la fois concepteur, médiateur, évaluateur et chercheur.

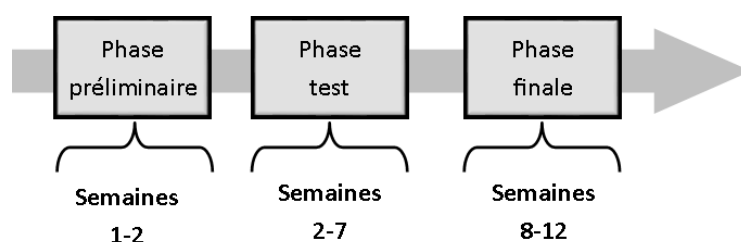
### **6.1.1 Matériel et Méthodes**

Dans cette sous-section, nous présenterons le déroulement général du dispositif d'apprentissage depuis sa mise en place et ses acteurs jusqu'au recueil des données. Nous mentionnons dès à présent que l'ensemble des documents sonores sont disponibles dans un dossier partagé *Googledrive* en ligne (voir partie « documents sonores » de l'étude 3 en annexes)

#### **6.1.1.1 Le dispositif général**

Le dispositif général d'apprentissage que nous avons mis en place a pour but mettre en place une médiation la perception et la production des sons du français chez des apprenants en contexte FLE/S. Ce dispositif d'apprentissage s'intègre dans l'approche par tâche (section 3.1.1) comme une micro-tâche d'entraînement puisque ce cours vise un travail spécifique sur la prononciation qui pourra être réintégré dans la pratique générale du français dans des cours non plus axés sur « la langue » mais sur des contenus disciplinaires comme la littérature ou la civilisation. Ce dispositif se déroule pendant le premier semestre de l'année universitaire 2013-2014 dans le cadre du Diplôme Universitaire d'Etudes Française (DUEF1) proposé à

l'université Sorbonne Nouvelle Paris 3. Cette formation consiste d'une part en un renforcement des éléments langagiers du français (écrit, oral, grammaire, prononciation) et d'autre part en une initiation aux différents aspects culturels, littéraires et historiques du monde francophone. Notre dispositif d'apprentissage était le « cours de prononciation » réparti sur une séance hebdomadaire de deux heures pendant douze semaines. Le volume horaire total représente donc 24 heures en présentiel. Ce cours était un des quatre choix facultatifs parmi les huit cours optionnels proposés. Le dispositif comportait trois phases principales : préliminaire, de test et finale (Figure 69).



**Figure 69:** Les trois phases chronologiques du dispositif d'apprentissage

Le dispositif d'apprentissage se composait de trois phases distinctes (Figure 69) :

- **La phase préliminaire** (une séance et demie soit 3h) : elle comporte une réflexion métalinguistique et théorique sur la phonétique-phonologie du français et sur les processus de production des sons.
- **La phase de test** (cinq séances et demie soit 11h) : elle comporte deux enregistrements (une séance et demie ; pré-test et post-test) ainsi qu'un travail en groupe portant sur une sensibilisation au trapèze vocalique et des exercices de type drills sur cinq principales oppositions phonémiques (quatre séances soit 8h).
- **La phase finale** (5 séances soit 10h) : elle comporte trois évaluations dont deux écrites visant à mesurer respectivement les connaissances théoriques vues pendant le cours et le recul pris durant l'année sur l'évolution de sa propre prononciation. La troisième évaluation tente d'évaluer les performances de production des dix voyelles principales du français standard à l'isolée et la lecture d'un court texte préparé pendant l'une des séances de la phase finale.

Lors de la phase de test, les étudiants étaient répartis en deux groupes de douze personnes et un groupe de 11. Les trois groupes faisaient strictement les mêmes exercices dans le temps imparti. Seule la manière de travailler avec le trapèze vocalique changeait (Tableau 71).

Groupe	1	2	3
Trapèze vocalique	F0 non contrôlée	F0 plate	F0 montante et piano

**Tableau 71:** Mode de présentation du trapèze vocalique aux apprenants selon les trois groupes.

Nous développerons plus en détails la phase préliminaire et de test dans la section 6.1.2. Nous ne développerons pas plus précisément la phrase finale car les évaluations proposées tiennent compte des critères imposés par cette formation soit : du contrôle continu, une réflexion linguistique et un travail langagier.

Ce dispositif aura permis aux apprenants de travailler de manière synchrone mais également asynchrone pour l'une des séances ; de manière individuelle, en groupe restreint (9 personnes) ou en groupe élargi (35 personnes). L'ensemble de ces caractéristiques est repris dans un schéma global en annexe 55.

### 6.1.1.2 Les apprenants

Les apprenants sont des adultes inscrits au DUEF1. Les cours et le diplôme correspondent aux niveaux B2 et C1 définis par le Cadre européen commun de référence du Conseil de l'Europe. Pour être admissibles, les étudiants doivent attester d'un niveau équivalent au DELF B2. Le profil des apprenants du cours était très hétéroclite en termes de niveau général en français, de niveau en prononciation du français, de profils langagiers, d'âge, etc. Au total, 35 personnes ont été inscrites administrativement mais en réalité 34 ont réellement suivi ce cours. Parmi ces derniers, nous avons choisi de ne garder pour notre étude que les données de 27 d'entre eux sur la base de leur présence à au moins 4 sur les 5 séances en groupe restreint et aux deux enregistrements (annexe 56). Le profil précis de ce groupe de 27 apprenants est le suivant :

- **Nationalités** : le groupe est hétérogène avec 18 nationalités différentes dont quatre personnes de nationalité russe ainsi que quatre de nationalité chinoise, deux personnes de nationalité taiwanaise ainsi que deux roumains et deux coréennes du sud. Les autres nationalités représentées ne comportent qu'un individu et sont péruvienne, iranienne, indonésienne, paraguayenne, géorgienne, chilienne, suédoise, française (anciennement vietnamienne), indienne, thaïlandaise, japonaise, colombienne et brésilienne.
- **Genre** : la majorité est féminine (23 femmes) mais nous comptons également 4 hommes
- **Age** : la plupart ont entre 20 et 30 ans (18 personnes), 2 ont moins de 20 ans, 6 entre 30-40 ans et 1 personne de 50 ans et plus.
- **Age d'apprentissage du français** : la majorité a appris le français de manière tardive soit 15 ans et plus (24 personnes) ; deux l'ont appris en 10-15 ans et un entre 5-10 ans.
- **Durée d'apprentissage du français** : la plupart apprend le français depuis une période de moins de 5 ans (23 personnes) ; un l'apprend depuis 5-10 ans et trois depuis plus de 15 ans.

- **Vivre en France** : une grande partie du groupe habite en France depuis moins de 5 ans (22 personnes), 2 entre 5-10 ans, 1 entre 10-15 ans et deux depuis plus de 15 ans.
- **Cours de prononciation** : 19 apprenants n'ont jamais suivi de cours de phonétique ou de prononciation du français avant le cours de prononciation du DUEF1 contre 8 qui en ont déjà suivi.
- **Intérêt** : la totalité des apprenants (27 personnes) affirme que la prononciation du français est importante dans leur apprentissage et leur pratique de cette langue.

En résumé, nous pouvons dire que les apprenants de ce dispositif d'apprentissage sont de nationalités hétérogènes (18 différentes / 27 apprenants) correspondant à des profils langagiers variés. Ce sont majoritairement des femmes (23/27) jeunes adultes (20/27) qui ont commencées tardivement l'apprentissage du français (15 ans et plus : 24/27), qui apprennent le français (moins de 5 ans : 23/27) et habitent en France (moins de 5 ans : 22/27) depuis une courte période. Plus de la moitié n'a jamais suivi de cours de prononciation avant celui-ci (19/27).

Afin de préciser cette analyse, nous avons posés des questions supplémentaires à notre groupe test soit le troisième groupe sur leur rapport à la musique :

- **Musicien-nes** : un peu moins de la moitié déclare se considérer comme musicien-ne (4/9) et l'autre non-musicien-ne (5/9).
- **Le chant** : une part importante affirme aimer chanter en général (7/9) cependant la même proportion avoue ne pas se sentir à l'aise s'ils doivent chanter.
- **Pratique de la musique** : une grande majorité ne pratique ou n'a pas pratiqué la musique sur une longue période (8/9).
- **Auditeur** : une grande partie écoute très souvent (4/9) ou souvent (3/9) de la musique contre une minorité peu souvent (1/9) ou rarement (1/9).

En résumé, il nous est donc possible de dire que notre sous-panel qui forme le groupe test n'est pas tranché en termes de musicien/non-musiciens. Cependant, les musiciens pratiqueraient la musique de manière plutôt occasionnelle. Ils aiment plutôt chanter mais ne se sentent pas à l'aise avec cette pratique. Ce sont plutôt des auditeurs réguliers de musique.

### **6.1.1.3 Les enregistrements**

Les enregistrements ont été effectués par l'auteur dans une salle plutôt calme affectée à ce cours. Ils ont été faits à l'aide d'un ordinateur portable HP Pavillon g6 Notebook PC fonctionnant sous Windows 7 64 bits, d'une carte son EdiroL UA-25EX et d'un micro-casque AKG C520L. Le

taux d'échantillonnage était de 44,1 kHz (16 bits). Les enregistrements des apprenants sont disponibles dans un dossier partagé *Googledrive* en ligne (voir partie « documents sonores » de l'étude 3 en annexes). Chaque apprenant devait prononcer :

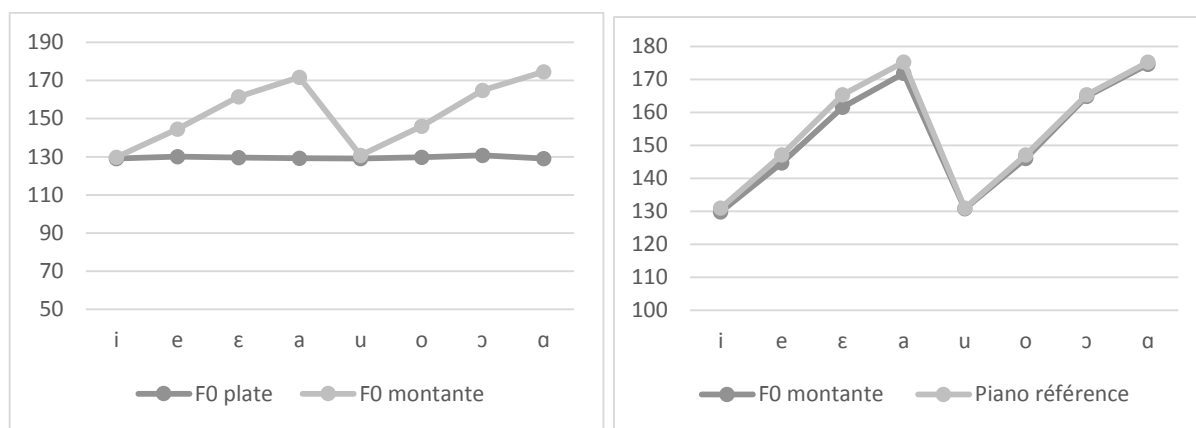
1. deux fois onze voyelles orales du français standards à l'aide de la transcription phonétique et d'une phrase-cadre (22 occurrences par apprenant),
2. deux fois deux continua de quatre voyelles soit les voyelles antérieures non-arrondies et postérieures arrondies (16 occurrences par apprenant).

Au total, nous avons donc 38 occurrences par apprenant. Pour l'enregistrement des voyelles à l'isolée, les apprenants des trois groupes visualisaient, à l'aide d'un diaporama (*Powerpoint*) (annexe 57), chaque son de manière individuelle dans une phrase cadre permettant aux apprenants de contextualiser si besoin le son dans un mot. Les apprenants ne devaient prononcer que le son vocalique à l'isolée car en raison du grand nombre d'apprenants et des contraintes de temps que nous avons, cela permettait de réduire le temps d'enregistrement pour chaque individu. Pour des raisons similaires, nous n'avons pas pu tester les performances en perception. Les phrases-cadre ont été tirées de Kamiyama (2010 : 138) ; par exemple « je dis [i] comme dans « lit » ». L'ordre des voyelles n'avait pas de logique articulatoire ou acoustique particulière : « i y e a ɔ ø u ε o α œ ». Pour l'enregistrement des voyelles dans les continua, nous avons mis en place trois situations différentes (Tableau 72).

<b>Groupe</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b><i>Input visuel</i></b>	Trapèze vocalique		
<b><i>Input auditif vocalique</i></b>	Aucun	F0 plate	F0 corrélée à l'aperture et utilisation d'un piano

**Tableau 72:** Types d'input visuel et auditif en fonction des trois groupes d'apprenants.

Dans le groupe 1, les apprenants devaient prononcer les deux continua soit « i e ε a » et « u o ɔ α » de manière spontanée avec pour seul support le trapèze vocalique. Les groupes 2 et 3 devaient prononcer les mêmes continua mais les apprenants disposaient d'un *input* auditif des continua vocalique respectivement avec une F0 plate et une F0 corrélée à l'aperture et utilisation d'un piano. Chaque continuum était joué quatre fois à l'apprenant avant qu'il ne les prononce deux fois lui-même. Les continua ont été enregistrés par l'auteur avec le même dispositif que celui utilisé pour les apprenants. Les continua sont disponibles dans un dossier partagé *Googledrive* en ligne (voir partie « documents sonores » de l'étude 3 en annexes). Il y a donc une différence de tâche (production pour le groupe 1 et répétition pour les groupes 2 et 3) entre les groupes.

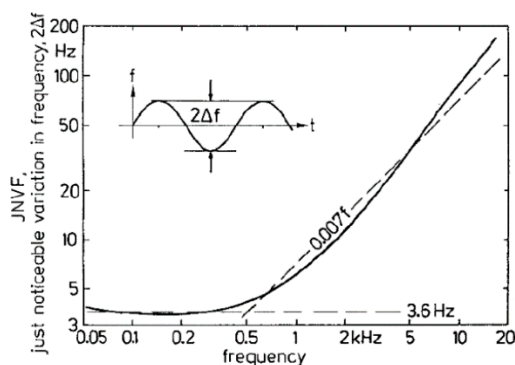


**Figure 70:** Fréquence (en Hz) de la F0 des continua d'input auditif pour les voyelles prononcées avec une F0 plate ou montante (à gauche) puis comparée avec des valeurs de référence jouées au piano (à droite).

			F0 plate								F0 montante							
	Référence	Piano midi	i	e	ε	a	u	o	ɔ	α	i	e	ε	a	u	o	ɔ	α
do	131	0,1	1,9	0,8	1,3	1,7	1,9	1,1	0,1	1,9	1,1				0,2			
ré	147	0										2,4				1		
mi	165	0,3											3,8				0,5	
fa	175	0,2												3,4				0,7

**Tableau 73:** Tableau comparatif (en Hz) de la différence absolue de la valeur de f0 des voyelles et la hauteur tonale de référence des notes de musique « do, ré, mi et fa » pour un diapason à 440 Hz ainsi que les notes jouées au piano midi.

A l'aide de la Figure 70 et du Tableau 73, nous pouvons constater qu'en termes de F0, les phonèmes vocaliques avec une F0 constante sont homogènes puisque proche d'un « do » (131 Hz) et avec une distance maximale de 1,9 Hz. De la même manière, pour les voyelles prononcées avec une F0 corrélée à l'aperture soit /i/ - /u/ avec « do », /e/ - /o/ avec « ré », /ε/ - /ɔ/ avec « mi » et /a/ - /α/ avec « fa », la distance maximale est de 3,8 Hz.

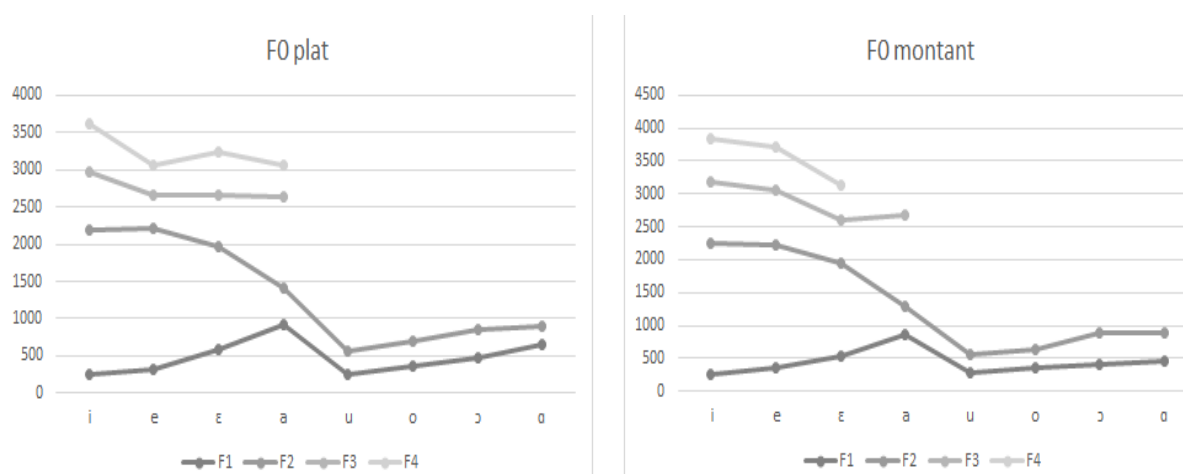


**Figure 71:** Modulation juste perceptible de fréquence en fonction de la fréquence pour des modulations de fréquences sinusoïdales à une fréquence de modulation de 4 Hz. Les lignes pointillées sont des approximations affines de la courbe (d'après Zwicker & Fastl, 2003 :183 dans Charlemagne, 2010 : 35).

Au regard de la Figure 71 et la notion de « modulation juste perceptible de fréquence » (« Just noticeable variation in frequency » : JNVF), nous pouvons constater qu'entre 100-200 Hz, la variation minimale perceptible en terme de fréquence sur des sons purs est de l'ordre de 3-4 Hz. Nous pouvons donc dire que les variations observées en termes de F0 des stimuli enregistrés ne sont pas significatives sur le plan perceptif puisque inférieure à 4Hz et donc peu perceptible. Sur le plan formantique, les stimuli montrent les caractéristiques acoustiques prototypiques du français mises en évidence dans des grands corpora (cf. section 1.3.3 : Tubach, 1989 ; Gendrot & Adda, 2005 ; Georgetown *et al.*, 2012) et ils ont été identifiés à l'aveugle par deux locuteurs francophones natifs (Tableau 74 et Figure 72).

	F0 plate								F0 montante							
	i	e	ɛ	a	u	o	ɔ	ɑ	i	e	ɛ	a	u	o	ɔ	ɑ
F1	260	329	591	926	258	361	486	643	251	354	540	856	274	366	400	453
F2	2180	221	1973	1414	555	695	864	897	2254	2219	1956	1299	572	637	883	888
F3	2970	2648	2667	2629					3184	3059	2591	2684				
F4	3604	3050	3230	3054					3848	3712	3136					

**Tableau 74:** Valeurs formantiques (en Hz) des stimuli vocaliques utilisés pour les inputs des groupes 2 et 3 et mesurées avec Praat d'après la valeur moyenne de la totalité du phonème.



**Figure 72:** Représentation graphique des valeurs formantiques (en Hz) des stimuli vocaliques utilisés pour les inputs des groupes 2 et 3 et mesurées avec Praat d'après la valeur moyenne de la totalité du phonème. En abscisse les phonèmes et en ordonnée la fréquence (en Hz).

Nous remarquons dans les deux cas (F0 plate ou montante) que les voyelles postérieures sont focales (F1F2) et les deux premiers formants augmentent selon l'aperture (fermé -> ouvert). Le /i/ a le F4 le plus élevé et le F1 le plus bas des quatre phonèmes du continuum antérieur étiré. Cependant dans le deux cas, le /i/ ne montre pas de focalité (F3F4). La stratégie acoustique d'opposition entre /e/-/ɛ/ pour ce locuteur est une augmentation de F1 et une diminution de F2 dans les deux productions (F0 plate et montante).

Au regard de la durée de l'input pour les groupes 2 et 3, nous constatons qu'en moyenne, les phonèmes qui ont servi de stimuli ont une durée comprise en une et deux secondes (Tableau 75).

F0 plate						
Phonème	i	e	ɛ	a	Total	Temps total de l'input (3 répétitions)
Durée	1,2	1,3	1,4	1,1	5,1	27,3
Phonème	u	o	ɔ	ɑ	Total	Temps total de l'input (3 répétitions)
Durée	1,2	1,3	1,3	1,2	5,0	25,9
F0 montante						
Phonème	i	e	ɛ	a	Total	Temps total de l'input (3 répétitions)
Durée	1,5	1,3	1,3	1,4	5,6	26
Phonème	u	o	ɔ	ɑ	Total	Temps total de l'input (3 répétitions)
Durée	1,5	1,3	1,5	1,4	5,7	25,8

**Tableau 75** : Durée des phonèmes et de l'input total en fonction du continuum et de la F0 produite. Les données sont en seconde.

Afin de former les stimuli finals, nous avons concaténé, l'aide d'un script de Cédric Gendrot, chaque phonème par continuum espacé par un silence de 200 ms. Chaque continua été répété trois fois espacé de 1500 ms et précédé par un bip d'une durée de 500 ms.

D'après cette analyse acoustique, il apparaît que les stimuli utilisés pour l'étude sont pertinents sur le plan acoustique puisqu'ils montrent un écart non significatif à la cible en ce qui concerne la F0. Au niveau formantique, les deux continua produits dans deux modalités différentes montrent les caractéristiques acoustiques formantiques prototypiques du français standard.

## 6.1.2 Médiation de la perception et production des sons de la parole

Dans cette section, nous présenterons les choix liés à la progression du curriculum, les outils pédagogiques ainsi que les documents utilisés.

### 6.1.2.1 La réflexion théorique

La première étape du dispositif d'apprentissage mis en place, pendant 3,5 heures lors des séances 1 et 2, se concentre autour d'éléments théoriques de la prononciation du français. Cette étape a pour but d'inciter les apprenants à prendre du recul, dans une perspective épistémique, sur leurs représentations de la/leur prononciation (du français) à travers les caractéristiques de base du système phonético-phonologique du français. Cette étape a également pour but de



sensibiliser les apprenants sur les variations du français que ce soit en termes de registre (académique versus relâché), régional (accent méridional, du nord, etc.) ou international (français d'Afrique, du Canada, etc.) mais aussi en termes du rôle des émotions dans ce processus. Le but assumé du cours, tout au long du semestre, n'était pas d'imposer une prononciation normée même si ce cours reposait majoritairement sur le français standard. Notre but était basé sur une médiation conduisant à se sentir à l'aise dans la prononciation d'une langue étrangère : le français. Cette partie théorique est également un choix pédagogique global dans le sens où le DUEF1 ne se compose pas seulement de contenus langagiers mais aussi de savoirs savants sur les plans linguistique, littéraire, culturel et historique en vue d'une intégration possible à l'université en licence voire master (ou équivalent). Cette étape contribue au développement d'une réflexion métacognitive sur la pratique du français. Comme mentionné précédemment, nous nous sommes largement appuyé sur les caractéristiques du français standard car elles semblent être les plus complètes parmi les possibles variations du français et car elles se rapprochent le plus du registre du français oral académique. Cependant, nous avons souligné, à chaque fois, les variations possibles par rapport à ce cadre. La transcription en API a été souvent utilisée afin de permettre aux apprenants de travailler en autonomie grâce aux transcriptions du dictionnaire. Les points traités sont les suivants (50 diapositives) :

**a) Réflexion générale :**

- Les définitions de la prononciation
- Les différents aspects de la prononciation (linguistique et paralinguistique)
- La relation transductive « langage/langue/parole-discours »
- La prononciation et le CECRL
- Les différences entre phonétique et phonologie

**b) Phonétique articulatoire et phonèmes du français :**

- L'appareil phonatoire
- Les phonèmes vocaliques
- Les phonèmes consonantiques
- Les semi-voyelles/consonnes
- La transcription API

**c) Eléments prosodiques :**

- La syllabe
- Les enchainements et liaisons
- Le « e » muet
- Les groupes rythmiques
- L'accentuation du français
- L'intonation

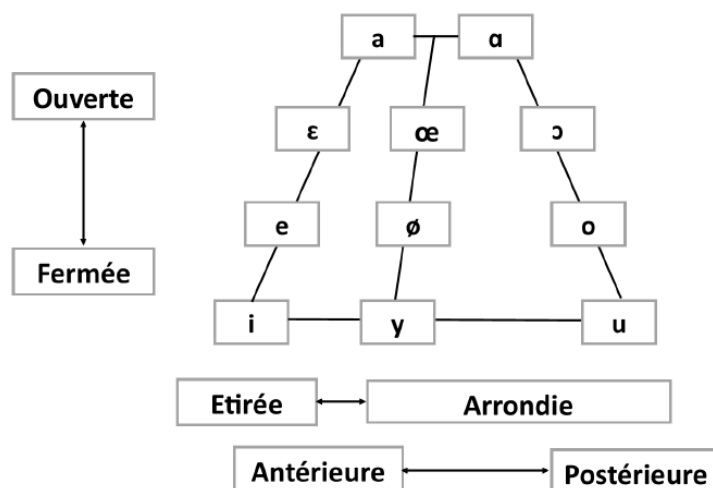
Les éléments mentionnés ci-dessus reposent donc sur le postulat que l'appropriation de savoirs savants chez des apprenants ayant un niveau de type universitaire leur permettra de prendre de la distance sur leurs représentations de la/leur prononciation à travers la compréhension des

mécanismes phonétique et phonologique du français. Ce processus pourrait se mettre en place à travers une réflexion métacognitive.

### **6.1.2.2 Le trapèze vocalique**

Lors des quatre séances de deux heures en groupe, nous avons systématiquement commencé les séances par des exercices avec le trapèze vocalique maximal du français standard soit onze phonèmes vocaliques. Nous n'avons pas présenté les « voyelles » nasales car leur intégration dans le trapèze vocalique est sujet à polémique étant donné que les recherches montrent un éloignement articulaire et acoustique avec leurs correspondantes phonologiques (Montagu, 2007 ; Carignan, 2013) et que nous ne pouvions traiter dans le temps imparti tous les phonèmes du français. Néanmoins dans la partie théorique, nous avons mentionné les particularités de ces phonèmes nasals du français. Nous avons décidé de présenter également des oppositions fines telles que les voyelles moyennes ou les deux réalisations phonétiques antérieure et postérieure du /a/ car nous rappelons que notre but est d'amener les apprenants à être sensibles à des différences phonétiques tout en développant de manière implicite de nouvelles représentations mentales des sons de la parole du français soit les phonèmes. Nous souhaitons préciser également que, dans la perspective des théories émergentistes, travailler sur des unités minimales (phonétique) a pour but d'initier une réorganisation du système (phonologique) en favorisant sa déconstruction. Cependant, l'émergence d'une telle rupture dans le système ne peut être prédite ni quantitativement, ni qualitativement ni qu'elle s'effectuera de manière homogène parmi les apprenants (section 2.3.2).

Nous postulons qu'un lien pédagogique entre aperture des voyelles postérieures et F0 pourrait favoriser la prise de conscience perceptive des voyelles et donc hypothétiquement leur production. En termes de hauteur spectrale, le /u/ est plus « grave » que le /a/ et c'est à ce titre que nous les avons respectivement associés à un « do » et « fa » de la même gamme. Afin de rendre cette association plus logique sur le plan visuel pour notre continuum postérieur test, nous avons décidé de renverser le trapèze vocalique « traditionnel » en mettant les voyelles fermées en bas (Figure 73). On peut retrouver cette représentation inversée notamment dans le « triangle de voyelles » sur le site [jaidumalachanter.fr](http://jaidumalachanter.fr) (consulté le 13/10/2014).



**Figure 73:** Trapèze vocalique renversé utilisé pour l'étude.

Nous avons travaillé sur le trapèze vocalique de plusieurs manières :

- isolée avec un ordre logique ou randomisé,
- en continua,
- par paires d'opposition.

Les apprenants étaient entraînés à la fois dans leur capacité à produire le phonème à partir du symbole (avec ou sans *input* audio de l'enseignant) mais également à discriminer les phonèmes produits par l'enseignant ou d'autres apprenants. Le but était de limiter le phénomène d'nativisation/assimilation à l'aide d'un *input* visuel et ainsi stimuler la construction de nouvelles représentations phonémiques vocaliques.

### 6.1.2.3 Les exercices d'entraînement

Les exercices d'entraînement se sont déroulés lors des quatre séances de travail en groupe. Leur but, toujours centré sur les voyelles orales du français standard, s'inscrivait dans la perspective de créer des automatismes articulatoires et perceptifs à travers, nous le supposons, la création de nouvelles représentations phonologiques internes. Comme nous nous situons dans une étude de type didactique quasi-expérimentale interventionniste, nous avons fait le choix d'exercices de type « *drills* » à partir du manuel : Abry & Chalaron (2011) « *Les 500 exercices de phonétique B1/B2* » (Tableau 76) ; afin de pouvoir aborder le même nombre d'exercices et donc contrôler l'*input* pédagogique fourni aux apprenants en termes quantitatif et qualitatif.

Opposition phonémique	Type d'exercices	Numéro et pages du manuel
[i] – [y] – [u]	Discrimination	1 p30
	Entraînement articulatoire	4-5-6 p32
	Rythme et intonation	11-13 p35
/E/ : [e] – [ɛ]	Discrimination	1 p43
	Entraînement articulatoire	2-3 p44
	Rythme et intonation	12-14 p47, 18 p48
/OE/ : [ø] – [œ]	Discrimination	1 p51
	Entraînement articulatoire	2-3 p52
	Rythme et intonation	10 p54
/O/ : [o] – [ɔ]	Discrimination	1 p58
	Entraînement articulatoire	3 p59
	Rythme et intonation	15-16 p63
/E/ - /OE/ - /O/	Discrimination	1 p66, 2-3 p67
	Entraînement articulatoire	4 p67, 8-9 p69, 10-11 p70, 12-13 p70
	Rythme et intonation	16 p71, 18 p72

**Tableau 76** : Exercices utilisés lors du travail en groupe avec les oppositions phonémiques travaillées, le type d'exercice selon la typologie proposée par le manuel et les références du manuel.

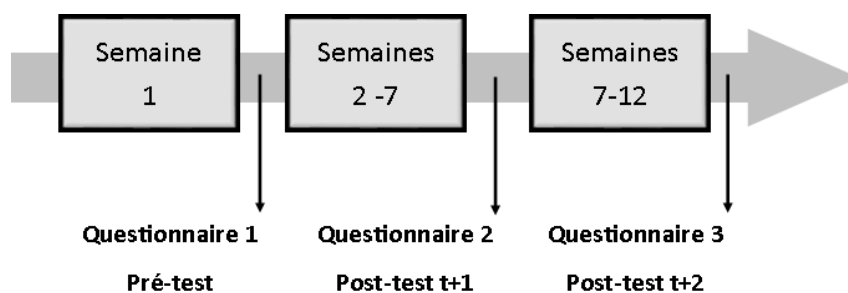
Nous avons ciblé les exercices intitulés dans le manuel « discrimination », « entraînement articulatoire » et « rythme et intonation » car étant donné le temps court sur lequel se déroulait l'expérimentation, nous souhaitions focaliser sur des exercices d'entraînement portant sur des processus de bas niveau tout en insistant sur la perception et la production. Les exercices « rythme et intonation » nous permettait d'introduire des exercices sur des énoncés longs.

## 6.2 Résultats

Dans cette partie, nous traiterons les différentes méthodes et modalités de recueil des données qui constituent ce terrain de recherche. Nous tenons à préciser que cette partie n'est que la première étape dans le processus d'analyse, interprétation et conclusion. L'interprétation se fera de manière thématique une fois que l'analyse des trois types de données sera achevée dans la partie « Discussions ». Nous analyserons dans l'ordre les données obtenues par questionnaire à trois moments du dispositif d'apprentissage, puis nous nous intéresserons aux résultats des enregistrements sous le prisme d'analyses acoustiques d'une part et d'un test d'identification passé par dix locuteurs natifs du français, d'autre part.

### 6.2.1 L'enquête par questionnaire

L'enquête par questionnaire a été créée sur la base d'un formulaire GoogleDocument© et comporte différents types de questions : à choix multiples, échelle de Likert et questions ouvertes. Les trois questionnaires sont disponibles en annexes 58, 59 et 60. Le but de ce questionnaire est de montrer des tendances positives ou négatives qui reflèteraient un dispositif d'apprentissage favorisant ou limitant le développement langagier à différents moments du semestre. Les données ont été traitées de manière manuelle ou automatique grâce aux différentes formules de calcul Excel. Le recueil des données à travers l'envoi d'un courriel s'est fait à trois étapes clés durant le semestre. Etant donné qu'il s'agit d'une étude en didactique quasi-expérimentale interventionniste, nous avons informé au préalable les étudiants que ces questionnaires ne seraient pas évalués et qu'ils permettraient de juger la progression des étudiants. Ils étaient également informés qu'il n'y avait ni « bonne » ni « mauvaise » réponse aux différentes questions. Nous avons ainsi tenté de limiter le biais qui peut influencer certaines réponses des apprenants dans le but de maintenir l'enseignant et l'enseignement dispensé dans un espace qui ne peut être critiqué et critiquable. Trois questionnaires ont permis de prélever l'état des représentations des apprenants sur leur apprentissage, et l'évolution de leur prononciation en français (Figure 74).



**Figure 74:** Chronologie des trois questionnaires visant à mesurer les représentations des apprenants pendant le dispositif d'apprentissage.

Les trois questionnaires ont été mis en place afin de mesurer différentes informations :

- **Questionnaire 1** : ce questionnaire fait office de pré-test car il a été mis en place après la première session de cours théorique. Il a permis de définir le profil socio-culturel des apprenants (section « Les apprenants ») mais également de mesurer les représentations sur leurs niveaux en français à l'oral, en prononciation générale, prononciation dissociée des voyelles, des consonnes et de la prosodie du français. Nous avons également cherché à savoir quelle place avait la prononciation dans leur apprentissage et s'ils avaient déjà travaillé avec le triangle/trapèze vocalique.

- **Questionnaire 2** : ce questionnaire est un post-test t+1. Il a été mis en place après le premier enregistrement suivi de quatre séances de travail en groupe et du deuxième enregistrement. Il a permis d'une part de mesurer l'évolution des représentations des apprenants sur les points du questionnaire 1 mais aussi de réfléchir sur les différences qu'ils ont ressenties entre les enregistrements 1 et 2 (ex : « lors du deuxième enregistrement, je me suis senti plus stressé-e/confiant-e/à l'aise/libre ») . Il nous a également été permis de récupérer des données plus spécifiques au groupe 3 : notre groupe test.
- **Questionnaire 3** : ce questionnaire est un post-test t+2. Il a eu lieu à la fin du semestre soit après la séance 12. Il nous a permis d'avoir un regard plus large sur le sentiment de développement langagier qui a eu lieu tout au long du semestre.

Nous nous intéresserons, tout d'abord, à analyser l'évolution entre le pré-test 1 et le post-test t+1 puis nous comparerons ces résultats à ceux obtenus au post-test t+2 afin de voir s'ils persistent à la fin du semestre où s'ils ont été atténués. Nous terminerons sur une comparaison plus spécifique en groupe.

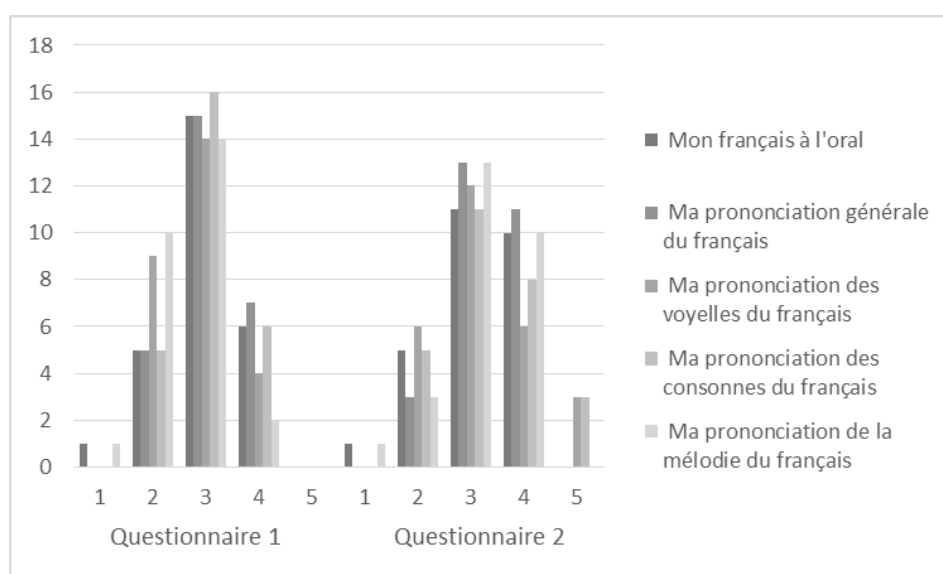
### 6.2.1.1 Tendances entre le pré-test t-1 et le post-test t+1

Nous nous sommes avant tout intéressé aux représentations des 27 apprenants concernant leur niveau sur plusieurs plans : leur français à l'oral, leur prononciation générale du français et leur prononciation spécifique des voyelles, des consonnes et de la mélodie. Pour cela, nous avons mis en place une question à échelle d'évaluation de type Likert avec cinq gradients (1 à 5) dont 1 correspond à « peu satisfait-e » et 5 « très satisfait-e ». Nous noterons que les apprenants n'avaient pas de moyen de mémoriser ni de se rappeler des réponses données lors questionnaire 1 avant ou pendant qu'ils remplissaient le questionnaire 2. Ils donnaient donc à chaque fois un sentiment à un moment donné non-corrélé aux réponses données antérieurement.

Indicateurs	Tests	n	Moyenne (écart-type)	Test t	Valeur de p	<0,05 = *
Mon français à l'oral	Q1	27	3 (0,8)	-1,162	0,256	ns
	Q2	27	3,1 (0,8)			
Ma prononciation générale du français	Q1	27	3,1 (0,7)	-1,363	0,185	ns
	Q2	27	3,3 (0,7)			
Ma prononciation des voyelles du français	Q1	27	2,8 (0,7)	-2,383	0,25	ns
	Q2	27	3,2 (0,9)			
Ma prononciation des consonnes du français	Q1	27	3,0 (0,6)	-1,615	0,118	ns
	Q2	27	3,3 (0,9)			
Ma prononciation de la mélodie du français	Q1	27	2,6 (0,7)	-3,844	0,001	**
	Q2	27	3,2 (0,8)			

**Tableau 77:** Comparaison à l'aide d'un Test t pour échantillons appariés et moyennes entre les questionnaires 1 et 2 pour les cinq caractéristiques évaluées, tous groupes confondus.

De prime abord et au regard des moyennes (Tableau 77), nous pouvons constater que selon les représentations des apprenants, il n'y a pas de grande progression entre les questionnaires 1 et 2. Toutefois, nous remarquons que la moyenne au questionnaire 2 est toujours supérieure à celle du questionnaire 1 entre 0,1 et 0,6 points supérieurs. Par conséquent, même s'il n'y a pas de progression importante, il est possible de constater une évolution positive. Les progressions les plus importantes sont celles de la prononciation des voyelles (0,4) et de la mélodie (0,6) du français, dont les taux au questionnaire 1 sont les plus bas. Sur le plan statistique, un test t pour échantillons appariés nous informe que la progression entre le test 1 et 2 n'est significative que pour la mélodie ( $p=0,001$ ). Afin de préciser cette première tendance, nous verrons plus en détail différentes caractéristiques.



**Figure 75:** Graphique représentant les réponses à l'échelle d'évaluation portant sur l'auto-évaluation du niveau de français des apprenants pour les questionnaires 1 et 2. En ordonnée, le nombre de réponses données (max = 27).

	Questionnaire	Note	Total des 5 questions	Tendance Q2-Q1
Notes extrêmes	Q1	1	2	=
	Q2		2	
	Q1	5	0	+
	Q2		6	
Notes centrales	Q1	2	34	-
	Q2		22	
	Q1	3	74	-
	Q2		60	
	Q1	4	25	+
	Q2		45	

**Tableau 78:** Tableau montrant l'évolution positive (+), négative (-) ou neutre (=) entre les questionnaires 1 et 2 pour les notes extrêmes (1 et 5) et centrales (2, 3 et 4) dont 1 correspond à « peu satisfait-e » et 5 « très satisfait-e ».

Nous pouvons voir sur la Figure 75 et le Tableau 78 que, même s'il n'y a pas de tendance moyenne majeure entre les représentations mesurées dans le questionnaire 1 puis le 2, il est clair qu'une tendance positive est mesurable. Tout d'abord, dans les notes extrêmes, même si le nombre de réponse « 1 » reste stable entre les deux questionnaires, en revanche six réponses maximales « 5 » apparaissent dans le questionnaire 2 contre zéro au questionnaire 1. De la même manière, entre les deux questionnaires, les réponses « 2 » et « 3 » diminuent tandis que les réponses « 4 » augmentent. Il semblerait donc qu'en moyenne, la progression ne soit pas importante mais dans une analyse détaillée des résultats, une progression vers des réponses « 4 » et « 5 » est à noter. Afin de prolonger cette réflexion, nous nous intéressons maintenant à la progression non plus générale du groupe des 27 apprenants mais à la progression individuelle. Nous avons pour cela soustrait l'évaluation donnée par chaque apprenant au questionnaire 2 à celle donnée au questionnaire 1.

Q2-Q1	Mon français à l'oral	Ma prononciation générale du français	Ma prononciation des voyelles du français	Ma prononciation des consonnes du français	Ma prononciation de la mélodie du français	Total
-2	0	1	0	2	0	16
-1	4	3	3	1	2	
0	15	13	13	13	10	64
1	8	9	9	9	13	55
2	0	1	1	2	2	
3	0	0	1	0	0	
Total	27	27	27	27	27	

**Tableau 79:** Effectifs de la soustraction entre les réponses données aux questionnaires 2 et 1. Une valeur positive (1, 2 3) signifie qu'un apprenant a mis une meilleure évaluation au questionnaire 2 qu'au 1. Une valeur négative (-1, -2) signifie l'inverse. Nous avons comptabilisé pour chaque question le nombre d'occurrence par tendance (n=27).

Grâce au Tableau 79, nous constatons qu'une part importante des apprenants pour les cinq questions (64/135 soit 47%) ne montre aucune évolution ni de manière positive ni négative. Une autre part importante (55/135 soit 41%) montre une progression positive par une augmentation de 1 à 3 points. Cependant, nous ne pouvons nier l'existence d'une minorité qui montre une diminution de niveau (16/135 soit 12%). En moyenne, la progression perçue est de 0,3 points donc une tendance légèrement positive sur l'ensemble du groupe.

Q2-Q1	Total des 5 questions			Ma prononciation des voyelles		
	Groupe 1	Groupe 2	Groupe 3	Groupe 1	Groupe 2	Groupe 3
-	3	8	5	1	2	0
=	24	15	25	5	1	7
+	18	22	15	3	6	2

**Tableau 80:** Tableau comparatif de la progression perçue par les apprenants en fonction des trois groupes de manière globale (à gauche ; n=45) et concernant la prononciation des voyelles (à droite ; n=9).



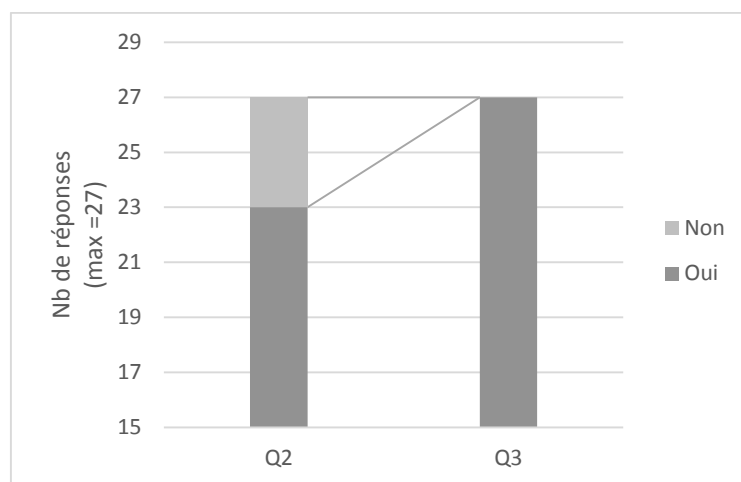
Le but initial de notre étude est de comparer la progression d'une cohorte d'apprenants répartis dans trois groupes différents soumis à des différences pédagogiques. Nous allons donc voir si des différences apparaissent dans ces trois groupes. La différence pédagogique entre les trois groupes portait principalement sur la production des sons vocaliques du français, nous nous intéressons donc plus particulièrement à ce point. D'après le Tableau 80 (à droite), nous pouvons constater le même profil de progression perçue parmi les trois groupes soit une majorité de stagnation à l'exception du groupe 2 qui marque une tendance majoritairement positive. Nous pouvons remarquer cependant que dans les groupes 1 et 2, au total trois apprenants marquent une régression dans leur auto-évaluation concernant leur capacité à produire les sons vocaliques du français.

D'après ces premières données, il semblerait qu'au regard d'un travail de deux heures hebdomadaires de travail mixte pendant sept semaines (théorie, pratique en groupe, enregistrements, évaluation), le groupe, en général, montre une légère tendance à une progression positive dans les représentations des apprenants dans leurs pratiques orale du français. Il ne semble pas y avoir de tendances particulières quand on compare les trois groupes. Toutefois, nous tenons à signaler que dans les deux questionnaires pré-test et post-test t+1, l'intégralité des apprenants (27/27) affirment puis maintiennent que la prononciation du français est importante dans leur apprentissage et leur pratique de cette langue.

### **6.2.1.2 Tendances au post-test t+2**

Après nous être intéressé à la progression perçue par les apprenants de leur prononciation du français entre le pré-test et le post test t+1, nous allons analyser et comparer ces résultats avec ceux obtenus lors du post-test t+2 mis en place à la fin du dispositif soit au bout de 12 semaines de travail dont 11 en présentiel et 1 à distance. Le questionnaire a été mis en place dans les mêmes conditions que les deux précédents. Nous ne dénombrons aucune perte dans les effectifs et par conséquent nous restons sur la base de 27 apprenants répartis en 3 groupes de 9 personnes. Nous reprendrons l'ordre d'analyse de la section précédente.

Nous noterons, dès à présent, que l'ensemble des 27 personnes déclare que la prononciation du français est importante dans leur apprentissage, leur pratique de cette langue et ce tout au long des trois questionnaires. Ils répondent également tous positivement, lors du troisième questionnaire, à l'affirmation : « j'ai perçu des changements dans ma prononciation depuis le début du cours de prononciation » avec une progression par rapport au questionnaire 2 (Figure 76). Lors du questionnaire 2, quatre apprenants répondaient négativement à cette affirmation.



**Figure 76:** Graphique représentant les réponses aux questionnaires 2 et 3 à la question « j'ai perçu des changements dans ma prononciation depuis le début du cours de prononciation » ( $n=27$ ). En abscisse, les questionnaires 2 et 3 et en ordonnée le nombre de réponses.

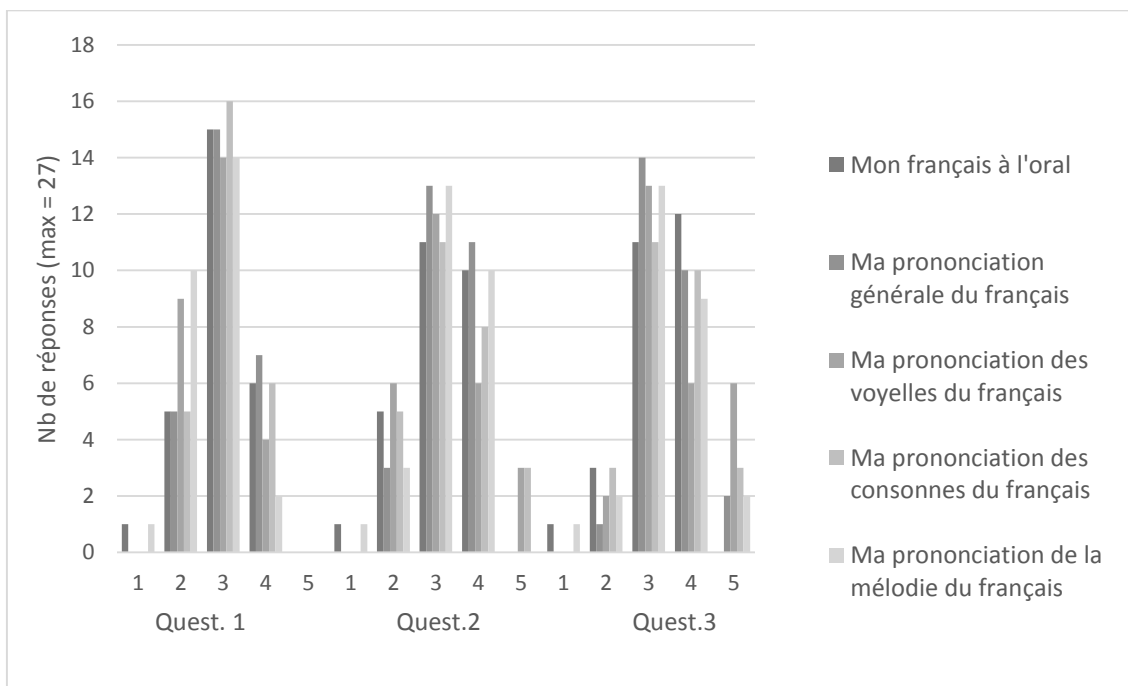
Nous nous intéresserons, tout d'abord, à leurs représentations de leur progression, sur plusieurs plans, à l'aide d'une échelle de Likert à cinq gradients (1 correspond à "peu satisfait-e" et 5 "très satisfait-e") : leur français à l'oral, leur prononciation générale du français et leur prononciation spécifique des voyelles, des consonnes et de la mélodie (résultats en annexe 61). Nous noterons que les apprenants n'avaient toujours aucun moyen de mémoriser leurs réponses aux deux questionnaires précédents.

Indicateurs	Tests	n	Moyenne (écart-type)	Test t	Valeur de p	<0,05 = *
Mon français à l'oral	Q1	27	3 (0,8)	-1,870	0,073	ns
	Q3	27	3,3 (0,8)			
Ma prononciation générale du français	Q1	27	3,1 (0,7)	-2,833	0,009	**
	Q3	27	3,5 (0,7)			
Ma prononciation des voyelles du français	Q1	27	2,8 (0,7)	-3,991	<0,0001	***
	Q3	27	3,6 (0,9)			
Ma prononciation des consonnes du français	Q1	27	3,0 (0,6)	-2,726	0,011	*
	Q3	27	3,5 (0,8)			
Ma prononciation de la mélodie du français	Q1	27	2,6 (0,7)	-4,208	<0,0001	***
	Q3	27	3,3 (0,9)			

**Tableau 81:** Comparaison à l'aide d'un test t pour échantillons appariés et moyennes entre les questionnaires 1 et 3 pour les cinq caractéristiques évaluées.

Au vu du Tableau 81, il est clair que les différentes composantes qui composent l'oral et la prononciation du français montrent, dans les représentations des apprenants, une progression. Nous noterons que la progression mesurée la plus importante touche, tout d'abord, la prononciation des voyelles du français (0,8 points) puis la mélodie du français (0,7 points).

Nous pouvons expliquer la tendance touchant les voyelles par un travail ciblé sur celles-ci tout au long du semestre et notamment à travers le travail de groupe. La découverte des caractéristiques prosodiques du français (l'allongement de la dernière syllabe de groupe rythmique ou les règles de liaison/enchaînement), sont souvent inconnues mais rapidement assimilées, ce qui a pu influencer la deuxième tendance. Nous remarquerons également, dans le tableau susmentionné, que la progression moyenne perçue des composantes de la prononciation (voyelles/consonnes/mélodie) est supérieure à celle de la prononciation générale soit respectivement 0,6 et 0,4 points. Selon eux, leur progression de la prononciation générale, tout au long du semestre, sur la même échelle de Likert a été jugée en moyenne de 3,4 points (écart type : 0,8) et la prononciation spécifique des voyelles de 3,6 (écart type : 0,9). Cependant, nous retrouvons le même ordre de grandeur. La progression moyenne mesurée est quasiment le double si l'on considère la progression entre Q1/Q2 et Q1/Q3. Cependant, la tendance est plus faible lorsque l'on considère l'impact d'un travail sur les voyelles sur la prononciation générale avec une moyenne des réponses de 2,1 points (écart type : 1,2).



**Figure 77:** Graphique représentant les réponses à l'échelle d'évaluation portant sur l'auto-évaluation du niveau de français des apprenants pour les trois questionnaires. En ordonnée, le nombre de réponses données (max = 27).

	Questionnaire	Note	Total des 5 questions	Tendance Q2-Q1	Tendance Q3-Q2	Tendance Q3-Q1
Notes extrêmes	Q1	1	2	=	=	=
	Q2		2			
	Q3		2			
	Q1	5	0	+	+	+
	Q2		6			
	Q3		13			

Notes centrales	Q1	2	34	-	-	-
	Q2		22			
	Q3		11			
	Q1	3	74	-	+	-
	Q2		60			
	Q3		62			
	Q1	4	25	+	+	+
	Q2		45			
	Q3		47			

**Tableau 82:** Tableau montrant l'évolution positive (+), négative (-) ou neutre (=) entre les questionnaires 1, 2 et 3 pour les notes extrêmes (1 et 5) et centrales (2, 3 et 4).

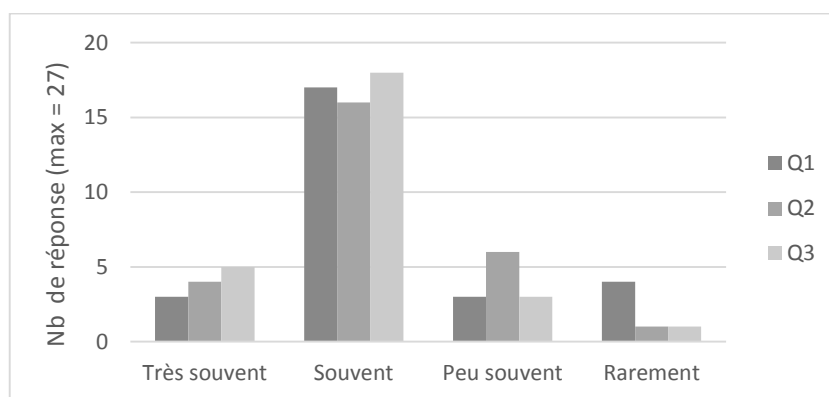
Nous constatons de manière générale (Tableau 82 et Figure 77) que sur les cinq niveaux étudiés, la note minimale (1) stagne tandis que la note maximale augmente de 13 occurrences. En ce qui concerne les notes centrales, les notes 2 et 3 tendent à diminuer respectivement de 23 et 12 occurrences tandis que les réponses 4 augmentent de 22 occurrences. Il apparaît donc que, même si en terme de moyennes, nous pouvons constater une progression perçue qui augmente de l'ordre de la décimale, il y a entre les questionnaires 1 et 3, une augmentation des évaluations supérieures (4 et 5) et une diminution des réponses inférieures et centrales (1, 2 et 3). Cette tendance est confirmée au regard de la progression mesurée pour chaque apprenant aux cinq questions entre deux questionnaires (Tableau 83).

Progression mesurée	Q2-Q1		Q3-Q1	
-2	16	12%	7	5%
-1				
0	64	47%	69	51%
1	55	41%	59	44%
2				
3				

**Tableau 83 :** Comparaison de la progression mesurée entre les réponses données aux questionnaires 1 et 2 ainsi que 1 et 3. Une valeur positive (1, 2 3) signifie qu'un apprenant à mis une meilleur évaluation au questionnaire 2 qu'au 1. Une valeur négative (-1, -2) signifie l'inverse. Nous avons comptabilisé pour chaque question le nombre d'occurrence par tendance ( $n=135$ ).

Nous pouvons voir grâce au Tableau 83, qu'entre les réponses données aux questionnaires 1 et 3, une majorité des apprenants, et ce pour les 5 niveaux concernés (oral, prononciation générale, voyelles, consonnes et mélodie du français), fait état soit d'une stabilisation (69 occurrences ou 51%) soit d'une progression allant de 1 à 3 points (59 occurrences ou 44%). Une très faible minorité (7 occurrences ou 5%) fait preuve d'une diminution. Nous remarquons que ces tendances se sont accentuées si l'on compare les résultats de la comparaison entre Q2-Q1.

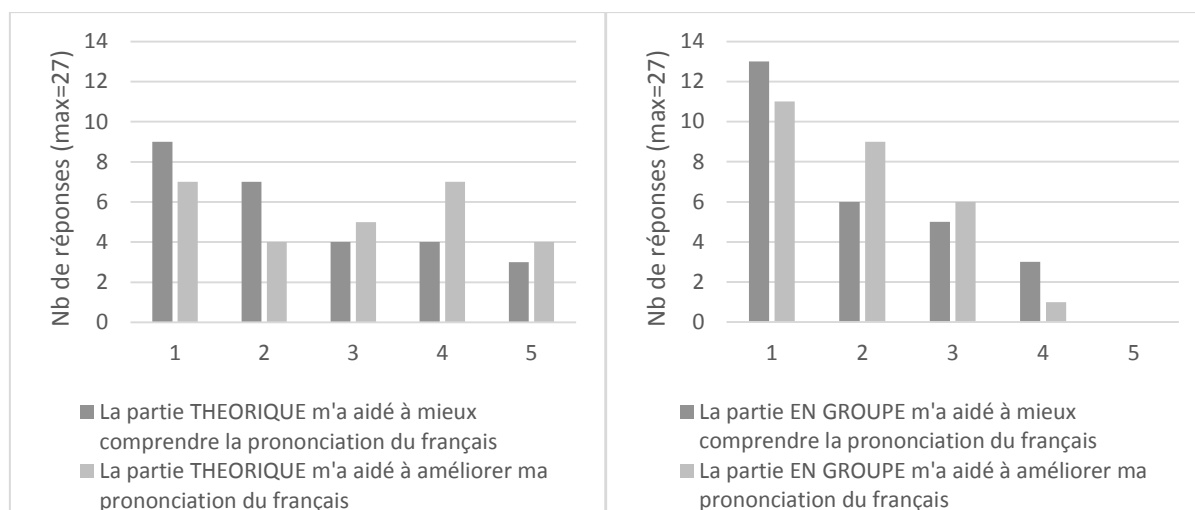
La progression peut être expliquée, en plus du travail lors des sessions de travail, par l'influence des pratiques en dehors du « cours de prononciation » puisque le dispositif d'apprentissage n'est, d'après notre cadre théorique, pas un univers clos. En outre, la majorité des apprenants déclare avoir souvent pratiqué le français à l'oral en dehors des cours. Cette tendance est maintenue voire en progression positive entre les questionnaires 1, 2 et 3 (Figure 78).



**Figure 78:** Graphique des réponses aux questionnaires 2 et 3 à la question « je pense avoir pratiqué le français à l'oral (en dehors des cours) » ( $n=27/\text{questionnaire}$ ). En abscisse, la fréquence et en ordonnée, le nombre de réponses.

De la même manière, lors du questionnaire 3, les apprenants déclarent, sur une échelle de 1 (« pas du tout ») à 5 (« beaucoup »), à la question « avez-vous travaillé sur votre prononciation en dehors du cours de prononciation? », avec une moyenne de 3,3 (écart type : 0,9) mais avec une représentation des réponses « 4 et 5 » de 10/27 contre 3/27 pour les réponses « 1 et 2 ». Cette tendance est légèrement en hausse par rapport au questionnaire 2 pour lequel la moyenne était de 3 (écart type = 1). Ils avaient également à disposition, à partir du questionnaire 2, une vidéo en ligne de l'enseignant prononçant les voyelles orales et nasales du français. Toutefois, cette vidéo n'a pas été utilisée par tous et de la même manière car sur une échelle allant de 1 (plusieurs fois) à 5 (jamais), leur réponse moyenne est de 2,8 (écart type : 1,4). Il est donc clair que la progression mesurée dans les représentations des apprenants sera le produit du travail proposé pendant les sessions en présentiel mais également du travail effectué en dehors de la « classe ». Nous devons prendre cet aspect en compte dans notre interprétation et nos conclusions.

Nous avons, par la suite, voulu préciser la tendance positive susmentionnée en réfléchissant sur l'impact des différentes étapes didactiques sur l'avantage acquisitionnel et sur le développement langagier. Les apprenants ont du évaluer sur une échelle de Likert de 1 à 5 respectivement « tout à fait d'accord » et « pas du tout d'accord », l'impact des parties théorique et pratique en groupe sur leur prononciation générale du français (Figure 79).



**Figure 79:** Graphiques représentant les réponses concernant l'impact des parties théorique (à gauche) et pratique en groupe (à droite) sur leur prononciation générale du français (n=27). En abscisse, le degré de réponse (1 correspond à « tout à fait d'accord » et 5 « pas du tout d'accord ») et en ordonnée, le nombre de réponse.

Nous pouvons constater, tout d'abord, grâce à la Figure 79 que les tendances générales, en ce qui concerne l'impact de la partie théorique d'une part et à la partie pratique en groupe, sont différentes. Il apparaît nettement, dans les deux graphiques, que l'impact perçu, de la partie pratique en groupe, est notée comme plus positive que celui de la partie théorique. Sur les deux questions portant sur la partie théorique et celle pratique (n=54), la moyenne des réponses aux premières est de 2,7 (écart type : 1,4) contre 1,9 (écart type : 1) pour les deuxièmes. Les apprenants ont donc perçu, de manière générale, que la partie pratique en groupe les a plus aidés à comprendre et améliorer leur prononciation du français que la partie théorique. De manière plus spécifique, il semblerait que la partie théorique ait plus aidé les apprenants à comprendre la prononciation du français (moyenne : 2,4 ; écart type : 1,4) qu'à les aider à améliorer leur prononciation (moyenne : 2,9 ; écart type : 1,4). Il en va de même concernant la partie pratique en groupe mais avec une différence moins marquée : « comprendre la prononciation du français » (moyenne : 1,93 ; écart type : 1) et « améliorer la prononciation du français » (moyenne : 1,88 ; écart type : 0,9).

Pour compléter cette réflexion, nous nous sommes questionné sur l'impact de l'utilisation du trapèze vocalique. Lors du questionnaire 1, une majorité d'apprenants (19/27) déclarait ne pas le connaître avant de commencer le « cours de prononciation ». Lors du questionnaire 2, ils déclaraient que, par rapport au début du cours, ils se sentaient plus à l'aise avec les symboles phonétiques et que ces derniers les avaient aidés à mieux percevoir/prononcer les voyelles du français (Tableau 84).

	Par rapport au début du cours, je me sens plus à l'aise avec les symboles phonétiques	Les symboles phonétiques m'ont aidé à mieux <i>percevoir</i> les voyelles du français	Les symboles phonétiques m'ont aidé à mieux <i>prononcer</i> les voyelles du français
<b>1 (tout à fait)</b>	7	8	6
<b>2</b>	13	10	9
<b>3</b>	3	3	7
<b>4</b>	3	5	4
<b>5 (pas du tout)</b>	1	1	1
<b>Moyenne (écart type)</b>	2,2 (1,1)	2,3 (1,2)	2,4 (1,1)

**Tableau 84:** Tableau des effectifs, moyennes et écart types ( $n=27$ ) des réponses à trois questions portant sur le rapport aux symboles phonétiques dans le questionnaire 2.

Ces tendances sont confirmées voire amplifiées lors du questionnaire 3 à la question « le triangle vocalique m'a aidé à mieux prononcer les voyelles du français » pour laquelle la moyenne des réponses est de 1,7 (écart type : 1,1).

D'après cette deuxième approche des données, il semblerait que la potentialité acquisitionnelle du dispositif d'apprentissage soit confirmée voire amplifiée par rapport au questionnaire 2. Ces tendances restent à une échelle modérée en raison d'une tendance vers une gaussienne autour de la note centrale : 3 ce qui reste l'un des biais de ce type de questionnaire. Néanmoins, il est important de signaler que cette progression positive se mesure à plusieurs niveaux des différents questionnaires sur un nombre plutôt faible de sujets (27) et de manière prolongée dans le temps. Cette tendance est mesurable même dans le cas où les apprenants ne pouvaient pas se rappeler de l'évaluation donnée aux questionnaires précédents.

### 6.2.1.3 Tendances spécifiques au travail en groupe

Nous traiterons, dans cette sous-partie, les questions portant sur les représentations et leurs évolutions entre le début et la fin du travail en groupe, soit quatre séances de 40 minutes par groupe d'une dizaine de personnes. Nous commencerons par une réflexion comparée entre la progression mesurée des trois groupes, pour les cinq niveaux mentionnés précédemment : oral général, prononciation générale, les consonnes, les voyelles et la mélodie du français (Tableau 85).

		Oral			Prononciation			Voyelles			Consonnes			Mélodie		
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
<b>Groupe 1</b>	Moyenne	3,00	3,11	3,33	3,00	3,11	3,56	2,67	3,11	3,22	2,78	3,22	3,11	2,33	3,11	2,78
	Ecart-T	1,00	1,05	0,71	0,71	0,60	0,73	0,71	1,17	0,97	0,67	0,97	1,05	0,71	0,78	0,83
<b>Groupe 2</b>	Moyenne	2,89	3,22	3,22	3,22	3,67	3,44	2,89	3,44	3,78	3,22	3,33	3,44	2,56	3,00	3,22
	Ecart-T	0,60	0,83	0,83	0,83	0,50	0,53	0,78	0,88	0,83	0,67	1,00	0,53	0,73	1,00	0,67
<b>Groupe 3</b>	Moyenne	3,00	3,00	3,22	3,00	3,11	3,44	2,89	3,11	3,78	3,11	3,44	3,89	3,00	3,44	4,00
	Ecart-T	0,71	0,71	0,97	0,50	0,78	0,88	0,60	0,78	0,97	0,60	0,88	0,78	0,50	0,53	0,71

**Tableau 85** : Moyennes et écarts-type (Ecart-T) des réponses au questionnaire d'attitude concernant les 5 niveaux évalués en fonction des trois groupes et des trois questionnaires.

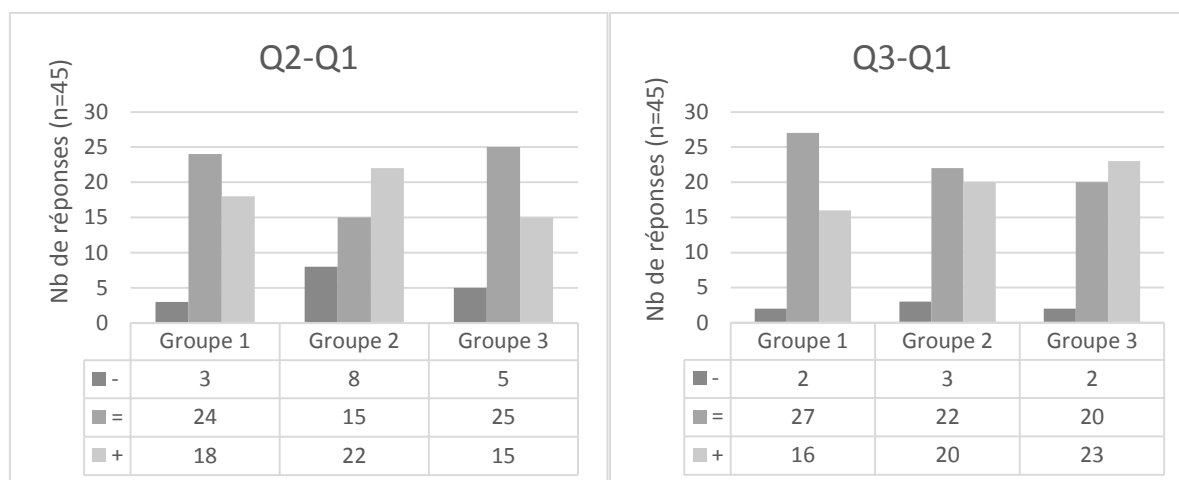
Afin de mesurer de manière statistique, l'effet du groupe sur la progression entre les Q1-Q2 et Q1-Q3, nous avons procédé à une analyse de variance (ANOVA à mesures répétées) avec comme facteur intra-sujet l'évaluation aux trois questionnaires pour chaque critères (oral, prononciation générale, voyelles, consonnes et mélodie) et comme facteur inter-sujet le groupe. Nous nous sommes intéressés à au facteur d'interaction entre le groupe et la progression Q1-Q2 d'une part et Q1-Q3 d'autre part (Tableau 86).

	f	erreur	ddl	sig	<0,05 = *
Oral (Q1-Q2)*groupe	0,571	24	2	0,572	ns
Oral (Q1-Q3)*groupe	0,051	24	2	0,951	ns
Prononciation (Q1-Q2)*groupe	0,444	24	2	0,646	ns
Prononciation (Q1-Q3)*groupe	0,444	24	2	0,646	ns
Voyelles (Q1-Q2)*groupe	0,311	24	2	0,736	ns
Voyelles (Q1-Q3)*groupe	0,308	24	2	0,738	ns
Consonnes (Q1-Q2)*groupe	0,269	24	2	0,766	ns
Consonnes (Q1-Q3)*groupe	1,091	24	2	0,352	ns
Mélodie (Q1-Q2)*groupe	0,571	24	2	0,572	ns
Mélodie (Q1-Q3)*groupe	0,927	24	2	0,410	ns

**Tableau 86** : Résultats de l'analyse ANOVA à mesures répétées pour l'interaction entre le groupe et les auto-évaluations entre les questionnaires 1 et 2 ainsi que 1 et 3 pour les cinq niveaux interrogés.

Au regard de cette analyse statistique, nous constatons qu'il n'y a pas d'effet d'interaction entre le groupe et la progression quelle que soit. Nous avons donc tenté d'analyser nos résultats de manière détaillée afin de mettre en évidence des profils ou tendances.





**Figure 80:** Graphiques et tableaux (n=45) récapitulatifs de la progression globale perçue par les apprenants en fonction des trois groupes concernant les cinq niveaux questionnés au regard des questionnaires 1 & 2 (à gauche) et 1 & 3 (à droite).

Globalement, nous pouvons constater que les trois groupes présentent les mêmes tendances : une minorité montre une régression, une majorité une stagnation et une partie importante une progression positive. Ces tendances sont constatées que ce soit entre les questionnaires 1 et 2 qu'entre les questionnaires 1 et 3. Nous remarquons également, grâce à la Figure 80, que la progression perçue entre le questionnaire 1 et les questionnaires 2 et 3, propose une diminution des régressions, et une augmentation des stagnations et progressions.

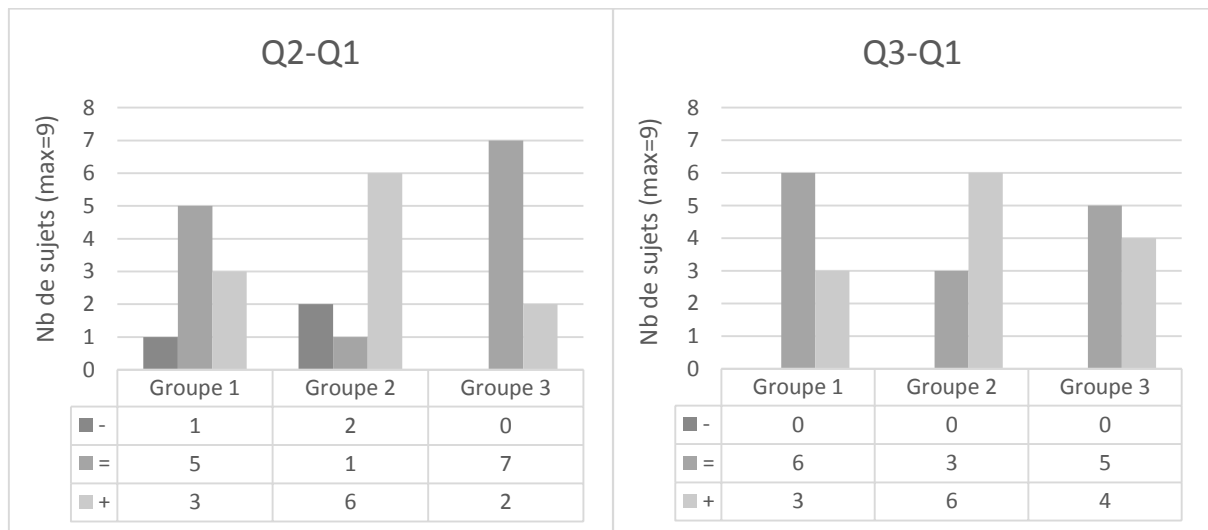
	Moyenne (écart type)		Progression Q3-Q2		
	Q2	Q3	-	=	+
<b>Groupe 1</b>	3,3 (1)	3,2 (0,7)	2	6	1
<b>Groupe 2</b>	3,8 (0,7)	3,3 (0,9)	3	6	0
<b>Groupe 3</b>	3,2 (1)	3,7 (0,9)	0	6	3
<b>Général</b>	3,4 (0,9)	3,4 (0,9)	5	18	4

**Tableau 87:** Tableau représentant (à gauche) la moyenne et l'écart type des réponses données en fonction de chaque groupe à la question « Selon vous, comment votre prononciation générale en français a évolué pendant le semestre 1 ? » et (à droite) la progression mesurée à la même question entre les questionnaires 2 et 3. (-) représente une différence négative donc une régression, (=) neutre donc une stagnation et (+) positive donc une progression entre les évaluations 2 et 3.

Afin de pouvoir comparer avec le point précédent, nous avons demandé aux apprenants d'évaluer sur une échelle de Likert allant de 1 (« pas de changement ») à 5 (« beaucoup de changements ») comment leur prononciation générale en français a évolué pendant le semestre 1. A travers le Tableau 87 (à gauche), nous pouvons constater une certaine homogénéité de la moyenne des réponses et ce entre les trois groupes. Nous remarquons de légères tendances différentes entre Q2 et Q3 pour chaque groupe mais ces dernières sont compensées sur l'ensemble des apprenants avec une moyenne de 3,4 (écart type = 0,9). En comparant les réponses de chaque apprenant au questionnaire 2 puis au 3, nous avons pu déterminer leur

progression mesurée, reportée dans le Tableau 87 (à droite). Il est possible de constater, dans les trois groupes, une majorité de stagnation et une homogénéité entre les régressions et progressions sur l'ensemble des 27 apprenants. Nous noterons simplement que le groupe 3 est le seul à ne présenter aucune régression perçue.

Regardons maintenant plus précisément les tendances intergroupes en ce qui concerne la prononciation des voyelles, plus spécifiquement, puisqu'elles constituent une approche particulière dans le travail de groupe et notre travail de thèse.



**Figure 81** : Graphiques et tableaux ( $n=9/\text{groupe}$ ) récapitulatifs de la progression perçue par les apprenants en fonction des trois groupes concernant la prononciation des voyelles au regard des questionnaires 1 & 2 (à gauche) et 1 & 3 (à droite). (-) représente une différence négative donc une régression, (=) neutre donc une stagnation et (+) positive donc une progression entre les évaluations 2 et 3.

A ce sujet, nous observons les mêmes tendances que pour la progression générale soit une minorité de régression, une majorité de stagnation, une partie importante et une augmentation de stagnation et progression positive perçue au regard de la progression perçue entre 1 & 2 et 1 & 3 (Figure 81). Nous pouvons également constater qu'au terme des douze semaines de travail, nous ne notons plus aucune régression perçue.

Etant donné que notre corpus intègre des enregistrements, nous avons voulu en connaître plus sur l'impact de ces derniers sur la performance elle-même à travers des éléments psychoaffectifs.

Lors du 2 <sup>ème</sup> enregistrement, je me suis senti	Plus stressé-e				Plus confiant-e				Plus à l'aise				Plus libre				Je n'ai pas perçu de différence			
	G1	G2	G3	TT	G1	G2	G3	TT	G1	G2	G3	TT	G1	G2	G3	TT	G1	G2	G3	TT
Tout à fait d'accord					1	4	3	<b>8</b>	3	4	2	<b>9</b>	2	2	2	<b>6</b>				
Plutôt d'accord	2	1		3	6	4	5	<b>15</b>	4	4	6	<b>14</b>	7	7	6	<b>20</b>			1	1
Moyennement d'accord																				
Plutôt pas d'accord	4	2	5	<b>11</b>	2	1		3	2	1		3					5	4	3	<b>12</b>
Pas du tout d'accord	3	6	4	<b>13</b>			1	1			1	1			1	1	4	5	5	<b>14</b>

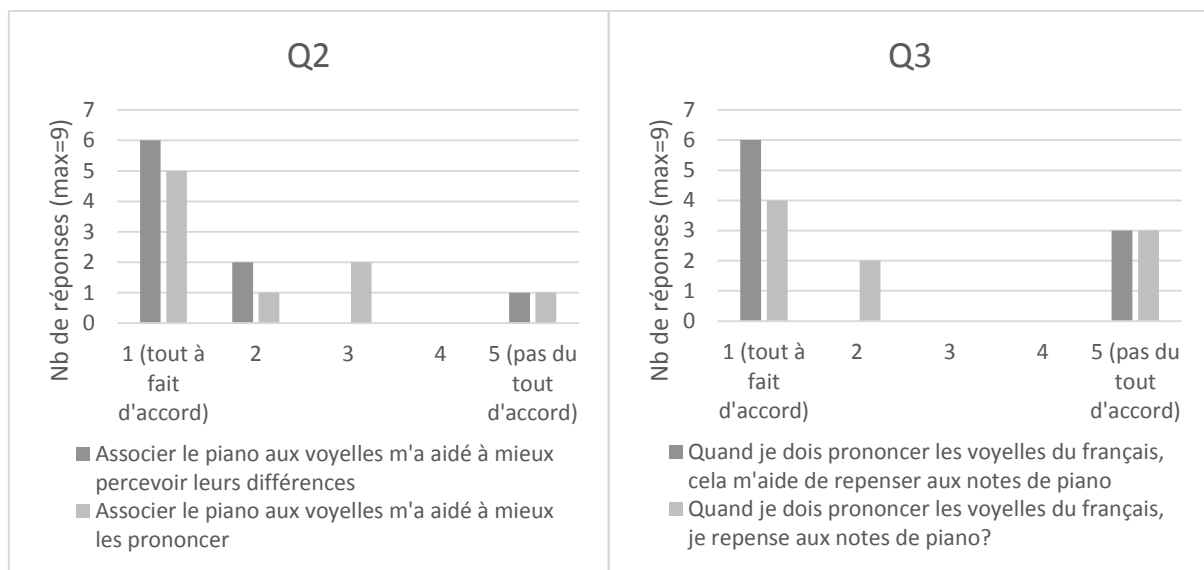
**Tableau 88:** Tableau représentant les réponses sur les différences ressenties entre les deux enregistrements. Les données sont présentées pour chaque affirmation en fonction de chaque groupe (G1, G2, G3, n=9) puis le total pour chaque réponse possible (TT, n=27). Les réponses nulles ont été représentées par une case vide pour plus de clarté. En gras, les données les plus marquantes pour le total de chaque réponse possible.

Le Tableau 88 met clairement en évidence le fait que, de manière générale, au deuxième enregistrement, les apprenants ne se sont pas senti plus stressé-es mais plus confiant-es, plus à l'aise et plus libres. Une grande majorité n'est pas d'accord avec le fait qu'ils n'ont pas perçu de différence lors du deuxième enregistrement. Nous pouvons également voir que les réponses se font principalement dans les extrêmes car sur les cinq affirmations, aucun apprenant n'a mentionné la réponse « moyennement d'accord ». Les trois groupes montrent des tendances similaires. Dans le but de compléter cette vision, nous avons proposé une question ouverte du type « si vous avez perçu un changement, expliquez brièvement, selon vous, ce qui a changé dans votre manière d'être lors du 2<sup>ème</sup> enregistrement par rapport au 1<sup>er</sup> ». Seize étudiants sur 27 ont commenté de manière plus spécifique mais une réponse n'a pas été prise en compte car trop vague et l'ensemble des quinze réponses est disponible en annexe 62. Nous avons pu grâce à cette question ouverte repérer quatre grandes thématiques :

- 1) **La prise de conscience articulatoire et l'entraînement** : sept occurrences mentionnent l'impact de l'entraînement sur la prise de conscience sur leur manière de produire les sons du français.
- 2) **La prise de confiance** : sept apprenant-es déclarent s'être senti-es plus sûr-es, plus confiant-es et plus à l'aise lors du deuxième enregistrement.
- 3) **La connaissance de l'activité** : trois occurrences font état du fait que connaître la nature et le déroulement de la tâche leur permettait d'être plus à l'aise dans ce qui était demandé.
- 4) **La connaissance des outils** : deux étudiant-es déclarent que s'être habitué-es aux symboles phonétique leur ont permis de mieux s'adapter à la consigne.

Afin d'être en mesure de prendre en compte tout biais, lié aux différentes approches du trapèze vocalique avec les trois groupes, nous avons cherché à savoir s'ils avaient pu prendre conscience de ces différences pédagogiques et expérimentales. Même si une minorité (8/27) a parlé avec les membres des autres groupes du cours de prononciation pour savoir ce qu'ils ont fait, aucun apprenant ne pense qu'il avait une différence dans le contenu du cours entre les groupes : (3/27) ont répondu qu'il n'y avait pas de différence selon eux et 23/27 ne savaient pas s'il y en avait une.

Pour terminer, sur cette analyse des données, nous nous sommes intéressés plus particulièrement au groupe 3 et les représentations des apprenants sur le lien proposé entre le trapèze vocalique et les notes de musique. A travers la Figure 82, nous remarquons de prime abord que ce lien fait apparaître majoritairement des réponses dans les extrêmes. Il semblerait que l'apport perçu est soit très positif soit très négatif. Lors du questionnaire 2, un nombre important d'apprenants déclarent être tout à fait d'accord sur le fait qu'associer le piano aux voyelles les a aidés à mieux percevoir leurs différences (6/9) et mieux les prononcer (5/9).



**Figure 82:** Graphiques représentant les réponses de 1 à 5 aux affirmations portant sur les liens entre notes de musique et trapèze vocalique lors du questionnaire 2 (à gauche) et 3 (à droite)  $n=9$ /affirmation).

Lors du questionnaire 3, nous remarquons la même tendance vers les extrêmes avec une grande partie de personnes qui déclarent être « tout à fait d'accord » avec le fait que repenser aux notes de piano les aide à prononcer les voyelles du français (6/9) et qu'ils repensent aux notes de piano quand ils doivent les prononcer (4/9). Cependant, un certain nombre déclare n'être « pas du tout d'accord » avec ces affirmations (3/9).

Au regard de l'analyse de ces premières données issues de questionnaires, il est possible de constater qu'une progression positive est mesurée que ce soit de manière indirecte, à travers l'évaluation des représentations sur la capacité à produire les unités orales du français et ce à plusieurs moments, ou de manière directe en demandant aux apprenants d'évaluer leur progression perçue en prononciation. Cette progression positive à petite échelle se retrouve sur l'ensemble des points mesurés et est persistante. Néanmoins, il faut attendre le questionnaire 3 pour que ce sentiment de progression soit majoritairement statistiquement significatif sur quatre des cinq points évalués (prononciation voyelles, consonnes et mélodie). La progression semble plus importante en ce qui concerne les voyelles en lien avec le fait qu'elles aient été le centre de ce travail de médiation phonologique. Il a fallu attendre le troisième questionnaire pour que l'ensemble des apprenants déclarent avoir perçu des changements dans leur prononciation depuis le début du cours de prononciation. Ce sentiment de progression peut-être dû à la fois au travail de médiation mis en place mais également au fait qu'ils soient en contexte homoglotte. Nous avons pu également montrer que, outre l'aspect linguistique, les apprenants déclarent se sentir plus à l'aise avec les outils utilisés que ce soit l'alphabet phonétique international ou les enregistrements ce qui a pu impacter sur les productions. En général, les apprenants pensent que le travail en groupe leur a été plus bénéfique que l'approche théorique. Toutefois, il ne semble pas y avoir de différences particulières entre les trois groupes du dispositif didactique. Néanmoins, le groupe 3 montre des tendances bimodales à savoir si l'association piano/trapèze vocalique les a aidés. En outre, lors du questionnaire 3, l'association proposée semble remporter, soit une adhésion soit un rejet fort. Afin d'aller au-delà de représentations, nous allons nous diriger vers une analyse acoustique et perceptive des productions des apprenants.

### **6.2.2 Analyses préliminaires**

Dans cette partie, nous nous intéresserons aux données obtenues lors des enregistrements 1 et 2. Ces analyses acoustiques et perceptives nous permettront de mesurer l'effet d'imitation pour les groupes 2 et 3 qui ont été exposés à un input et ce en comparaison au groupe 1 qui a eu une tâche de production. Nous serons, par conséquent, en mesure de déterminer l'impact qu'ont eu les inputs dans les productions des apprenants.

### **6.2.2.1 Mise en place**

Nous rappelons que les données ont été enregistrées par l'auteur à l'aide d'un ordinateur portable, d'une carte son externe et d'un micro-casque dans une salle de classe calme. Pour l'ensemble des 27 apprenants, nous avons au total 54 fichiers audio (.wav). Afin de faciliter le traitement automatique des données acoustiques, nous avons nettoyé chaque fichier audio en ne gardant que les 38 phonèmes vocaliques attendus (2 répétitions de 11 voyelles à l'isolée et 2 répétition de 2 continua vocaliques de 4 phonèmes) et en supprimant tout autre production parasite (explications, hésitations, consigne). Nous avons donc obtenu au total 2052 phonèmes vocaliques. Les fichiers audio ont été segmentés et annotés par l'auteur sur Praat en alphabet SAMPA (« *Speech Assessment Methods Phonetic Alphabet* »). En raison de problème de saturation lors du deuxième enregistrement, nous avons dû faire un sous-groupe pour le traitement formantique afin d'avoir des données traitables sans perturber l'analyse statistique (6 apprenants sur 27). Les différentes données acoustiques ont été obtenues automatiquement à l'aide d'un script Praat adapté par Charlotte Koukolia (d'après un script de Cédric Gendrot) qui permet de récupérer, de manière automatique et pour un grand nombre de fichiers, dans un tableau Excel différentes mesures acoustiques à partir d'un .wav et un .txtgrid telles que la durée, les formants F1, F2 et F3 (début/milieu/fin et moyenne) ainsi que la F0 (début/milieu/fin, min/max/étendue, moyenne au cœur et pente). Le traitement statistique a été réalisé sur le logiciel SPSS. Nous nous intéresserons plus particulièrement à l'impact de la modalité de production (isolée ou continuum) et du groupe sur la durée des phonèmes et la F0. Nous ne nous intéresserons pas à des mesures formantiques étant donné la variabilité importante des apprenants (L1, sexe, etc.).

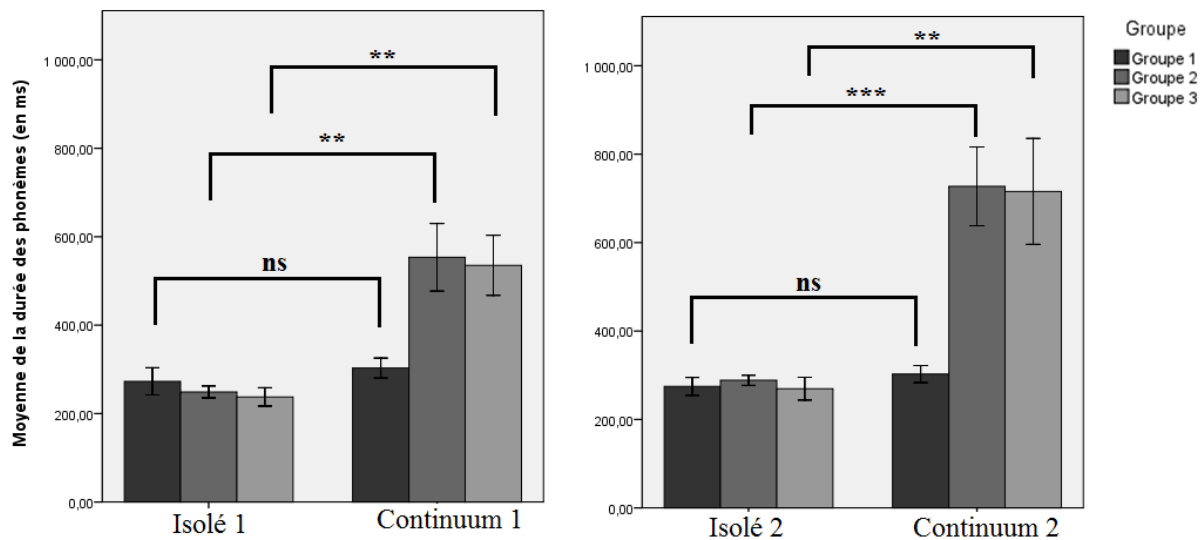
### **6.2.2.2 Durée**

Afin de mesurer l'impact des stimuli auditifs sur la production des phonèmes, nous nous sommes intéressés à la durée des phonèmes vocaliques en fonction de la modalité de production et du groupe. Nous rappelons que seuls les groupes 2 et 3 ont eu un stimulus auditif avant la production des phonèmes vocaliques en continuum. Nous avons concentré notre analyse sur les phonèmes des continua qui nous intéressent « antérieur étiré » et « postérieur arrondi ». Nous avons pris en compte les deux répétitions afin de rendre l'analyse statistique plus représentative (tableau récapitulatif par apprenant en annexe 63).

Statistiques pour échantillons appariés								
		Moyenne	Ecart-type	n	t	ddl	sig	<0,05=*
Groupe 1	Isolé1	272,84	92,43	9	-2,000	8	0,081	ns
	Cont1	302,87	67,93	9				
	Isolé2	274,67	60,70	9	-1,900	8	0,094	ns
	Cont2	302,70	57,98	9				
Groupe 2	Isolé1	248,89	40,54	9	-4,197	8	0,003	**
	Cont1	553,44	229,22	9				
	Isolé2	288,53	34,15	9	-5,121	8	0,001	***
	Cont2	727,28	267,31	9				
Stimuli		1260,30	84,98					
Groupe 3	Isolé1	237,56	62,12	9	-4,713	8	0,002	**
	Cont1	535,278	203,89	9				
	Isolé2	269,48	77,32	9	-3,867	8	0,005	**
	Cont2	715,67	359,35	9				
Stimuli		1416,9	68,39					

**Tableau 89:** Durées et écarts-type moyens (en ms) des phonèmes des deux continua analysés en fonction des tests 1 et 2 et des modalités de production. Les durées moyennes des inputs sont également données.

Nous avons comparé les durées moyennes des phonèmes produits en fonction de la modalité de production (isolée vs continuum) à l'aide d'une analyse statistique (Test t à échantillons appariés), dont les résultats sont indiqués dans la Figure 83 et le Tableau 89. Nous remarquons, dès lors, une tendance similaire entre les enregistrements 1 et 2 à savoir une différence significative entre les deux modalités pour les groupes 2 et 3 seulement. Lors du premier enregistrement, nous constatons que les phonèmes produits en continuum sont en général plus longs que ceux produits à l'isolée de 304,6 ms pour le groupe 2 et 297,7 ms pour le groupe 3. La différence est donc significative pour le groupe 2 ( $t = -4,197$ ,  $ddl = 8$ ,  $p = 0,003$ ) et le groupe 3 ( $t = -4,713$ ,  $ddl = 8$ ,  $p = 0,002$ ). Cette différence n'est pas significative pour le groupe 1 (Test  $T = -2,000$ ,  $ddl = 8$ ,  $p = 0,081$ ) avec une différence moyenne de 30 ms.



**Figure 83:** Diagramme des durées moyennes (en ms) des phonèmes des deux continua analysés en fonction des tests 1 et 2 (T1 à gauche et T2 à droite) et des modalités de production (isolé/continuum).

Le même phénomène est observé, lors de l'enregistrement 2, avec une augmentation significative pour le groupe 2 de 438,8 ms (Test  $t = -5,121$ , ddl = 8,  $p = 0,001$ ) et pour le groupe 3 de 446,2 ms (Test  $t = -3,867$ , ddl = 8,  $p = 0,005$ ). Une analyse de variance, nous indique que la différence entre les trois groupes n'est pas significative à l'isolée que ce soit pour l'enregistrement 1 ( $F = 0,624$ , ddl = 2,  $p = 0,544$ ) ou l'enregistrement 2 ( $F = 0,242$ , ddl = 2,  $p = 0,787$ ). Cependant la différence entre les groupes est significative pour les productions en continuum pour l'enregistrement 1 ( $F = 5,339$ , ddl = 2,  $p = 0,012$ ) et l'enregistrement 2 ( $F = 7,743$ , ddl = 2,  $p = 0,003$ ). Une analyse Post Hoc (Bonferroni) nous indique que dans les deux cas la différence n'est significative qu'entre les groupes 1 vs 2 (enregistrement 1 :  $p = 0,022$  ; enregistrement 2 :  $p = 0,006$ ) et 1 vs 3 (enregistrement 1 :  $p = 0,036$  ; enregistrement 2 :  $p = 0,008$ ).

Il semblerait donc que l'exposition à un stimulus auditif long (moyenne groupe 2 = 1,2 s ; groupe 3 = 1,4 s) ait influencé les productions des apprenants des groupes 2 et 3 pour les productions en continuum. Sans input sonore, il n'y a pas de différence entre les phonèmes produits à l'isolée et ceux en continuum au regard des résultats obtenus pour le groupe 1. Il en va de même entre les trois groupes lorsqu'ils produisent les phonèmes dans les mêmes conditions soit à l'isolée (sans input).



### 6.2.2.3 Fréquence fondamentale

Etant donnée la variabilité des sujets de notre corpus, il est très difficile de mener une analyse statistique comparative en termes de F0 puisque celle-ci varie chez les individus en fonction de leurs caractéristiques physiologiques, entre autres. Nous avons donc décidé de comparer la seconde répétition de chaque continuum pour chaque sujet (tableau récapitulatif par apprenant en annexe 64). Afin de contourner les différences individuelles liées à la f0, nous avons fait la moyenne de l'écart de F0 de chaque paire de phonème pour chaque continuum comme ci-suit :

- Continuum Antérieur (ANT) = moyenne ((e-i) ; (ε-e) ; (a-ε) ; (ε-i) ; (a-i))
- Continuum Postérieur (POST) = moyenne ((o-u) ; (ɔ-o) ; (ɑ-ɔ) ; (ɔ-u) ; (ɑ-u))

Un résultat largement positif indiquera donc une augmentation moyenne de la F0 de /u/ à /A/ et de /i/ à /a/. A l'aide du Tableau 90 (en haut), nous pouvons constater que le groupe 3 montre une tendance positive dans la moyenne des écarts que ce soit pour le continuum antérieur ou postérieur. En détail (en bas), seuls quatre apprenants du groupe 3 montrent une moyenne positive forte (SUN, RAD, NUG et KOZ). Nous remarquons également que KOZ exprime une asymétrie entre les deux continua.

	ANT	POST
Groupe 1	-12,80	-17,82
Groupe 2	-13,09	-13,49
Groupe 3	0,82	3,73

GR	Prénom	ANT	POST
3	SUN	<b>45,2</b>	<b>47,4</b>
3	RAD	<b>41,4</b>	<b>40,4</b>
3	NUG	<b>19,6</b>	<b>19</b>
3	KOZ	-0,4	<b>38,2</b>
2	SHU	-0,6	1
3	PER	-1,2	-22,4
2	YAN	-2,4	-1,8
1	NII	-4,2	1,6
1	HUA	-4,2	-13,6
1	MEH	-6,6	-15,6
1	LYQ	-7,4	-16
1	DEL	-7,6	-23,8
2	KIM	-8,4	-15,6
1	KIT	-8,8	-8,8

GR	Prénom	ANT	POST
3	FAN	-9	-1
3	DUM	-10,6	-9,2
2	IAK	-11,8	-11,2
2	QUI	-14	2
1	MAL	-16	-23,4
2	RIN	-17,2	-12,4
2	GHE	-20,2	-22
2	GOG	-20,4	-22,4
2	WEN	-22,8	-39
3	PAR	-23,6	-22,4
1	DOG	-29,8	-22,2
1	LOP	-30,6	-38,6
3	LEO	-54	-56,4
Moyenne		-8,35	-9,19

**Tableau 90** : Moyenne de la différence de la F0 (en Hz) des différentes paires vocaliques en fonction des continua antérieur (ANT) et postérieur (POST) lors de l'enregistrement 1. En haut, les moyennes en fonction de chaque groupe d'apprenant et en bas (à gauche et à droite), pour chaque apprenant. En gris et gras, les résultats supérieurs à 0. Un nombre positif indique une augmentation positive moyenne entre chaque phonème du continua. La F0 a été calculée par Praat sur la moyenne du phonème.

Afin de préciser cette tendance parmi les apprenants du groupe 3, nous avons pris en compte si les apprenants avaient produit les continua selon le schéma prosodique induit par l'input (une intonation montante). Pour cela, nous avons fait une analyse perceptive puisque la qualité des enregistrements post-test ne nous permettait pas de faire une analyse acoustique.

Sujet	E1		E2		TT E1	TT E2	TT	E2-E1
	ANT	POST	ANT	POST				
DUM			X	X	0	2	2	2
FAN			X	X	0	2	2	2
KOZ		X			1	0	1	-1
LEO					0	0	2	0
NUG	X	X	X	X	2	2	4	0
PAR					0	0	0	0
PER			X	X	0	2	2	2
RAD	X	X			2	0	2	-2
SUN	X	X	X	X	2	2	4	0
TOTAL	3	4	5	5	7	10	17	3

**Tableau 91:** Analyse perceptive de la réalisation de la F0 attendue d'après l'input musical lors des enregistrements 1 et 2. Une croix indique que l'apprenant a réalisé les quatre phonèmes du continua avec une F0 montante proche de celle de l'input auditif. En gris, les apprenants ayant réalisé au moins un continua avec la forme attendue lors de l'enregistrement 1.

En comparant les deux tableaux ci-dessus (Tableau 90 et Tableau 91), nous remarquons que les analyses acoustique et perceptive pour le pré-test sont corrélées puisque nous retrouvons les apprenants KOZ, NUG, RAD et SUN parmi ceux qui ont produit les continus avec la F0 induite par les inputs. Nous constatons également que l'asymétrie entre les continua antérieur et postérieur pour KOZ s'explique par le fait qu'il n'est produit que le deuxième avec la F0 montante. En comparant ces données avec ce qui a été produit lors de l'enregistrement post-test, il est possible de voir que plusieurs profils se détachent : ceux qui ont été constants et qui ont soit tout le temps produit la F0 attenue (2/9) soit jamais (2/9), ceux qui ne l'ont fait qu'au premier test (2/9) et ceux qui ne l'ont fait qu'au deuxième (3/9). Sur les deux enregistrements, seuls deux apprenants n'ont pas produit au moins un continuum avec la F0 attendue d'après l'input écouté.

D'après ces résultats, nous avons souhaité corréler ces tendances avec les réponses des apprenants du groupe 3 aux deux questions du questionnaire : « quand je dois prononcer les voyelles du français, cela m'aide de repenser aux notes de piano » et « quand je dois prononcer les voyelles du français, je repense aux notes de piano? ». En effet, ces deux questions montraient deux tendances extrêmes soit un fort ou faible accord.

Sujet	TT	E2-E1	Q2		Q3	
			Aide à mieux percevoir leurs différences	Aide à mieux les prononcer	Cela m'aide de repenser aux notes de piano	Je repense aux notes de piano
DUM	2	2	1	1	1	1
FAN	2	2	2	2	1	1
KOZ	1	-1	5	5	5	5
LEO	2	0	1	1	5	5
NUG	4	0	2	3	1	1
PAR	0	0	1	1	1	1
PER	2	2	1	1	1	2
RAD	2	-2	1	3	5	5
SUN	4	0	1	1	1	2

**Tableau 92:** Réponses aux questionnaires 2 et 3 des apprenants du groupe 3 concernant l'intérêt de l'association note de piano / phonème corrélée au nombre de continua produits avec la F0 attendue. « 1 » correspond à « tout à fait d'accord » et « 5 » à « pas du tout d'accord ».

Le Tableau 92 nous permet de corréler la production de la forme prosodique attendue avec leurs réponses à savoir si associer les notes de piano aux phonèmes a été une aide et cela lors des questionnaires 2 et 3. Tout d'abord, nous constatons que la personne qui pense le moins que les notes de musique ont été une aide est également une de celles qui a produit le moins de continua avec la F0 attendue (une fois) et n'a pas réitéré cela lors du deuxième enregistrement. Les deux apprenants qui ont considéré lors du troisième questionnaire que repenser aux notes de piano ne les aident pas sont ceux qui ont soit produit moins de continua avec la F0 attendue soit qui ont stagné. Un apprenant a indiqué lors des deux questionnaires que cette association l'avait aidé mais n'a produit aucune occurrence avec la F0 proposée. Les autres ont indiqué que cela les avait aidé et ont produit au moins 2 occurrences de la F0 attendue et ont stagné ou progressé entre les enregistrements 1 et 2. Pour développer cette corrélation cognitive, nous allons tenter d'analyser ces différences sous un aspect psychoaffectif.

Le Tableau 93 nous indique que les deux apprenants qui ont prononcé tous les continua avec la F0 attendue (SUN et NUG) déclarent tous les deux aimer chanter mais l'un d'eux ne se sent pas à l'aise quand il doit chanter. L'apprenant qui n'a jamais prononcé les phonèmes avec la F0 attendue (PER) a déclaré ne pas aimer chanter et ne pas se sentir à l'aise lorsqu'il chante. Les trois apprenants qui ont prononcé plus d'occurrences lors du deuxième enregistrement montrent qu'ils se sentaient lors de ce dernier plus confiants, plus à l'aise et plus libres mais pas plus stressés. Les deux apprenants qui ont régressé entre les deux enregistrements (KOZ et RAD) indiquent qu'ils aiment chanter mais qu'ils ne se sentent pas à l'aise en chantant. RAD indique même qu'il ne se sentait pas plus stressé, confiance, à l'aise et libre. Trois apprenants (DUM,

FAN, PER) n'ont produit des continua avec la F0 attendue qu'au deuxième enregistrement. Il y a une hétérogénéité des réponses quant à savoir s'ils aiment chanter / se sentent à l'aise en chantant mais tous montrent une même tendance à se sentir plus confiant, plus à l'aise et plus libre mais pas plus stressé. Cette tendance se retrouve chez LEO qui a produit deux occurrences lors de chaque enregistrement.

Sujet	TT	E2-E1	Q2						
			Plus stressé-e	Plus confiant-e	Plus à l'aise	Plus libre	Pas de différence	J'aime chanter	A l'aise si je dois chanter
<b>DUM</b>	<b>2</b>	<b>2</b>	--	+	+	+	--	<b>Oui</b>	<b>Oui</b>
<b>FAN</b>	<b>2</b>	<b>2</b>	-	++	++	++	-	<b>Oui</b>	<b>Non</b>
KOZ	1	-1	-	+	+	+	+	Oui	Non
LEO	2	0	--	++	++	+	--	Oui	Non
NUG	4	0	--	+	+	++	--	Oui	Non
PAR	0	0	-	+	+	+	-	Non	Non
<b>PER</b>	<b>2</b>	<b>2</b>	-	++	+	+	--	<b>Non</b>	<b>Non</b>
RAD	2	-2	--	--	--	--	--	Oui	Non
SUN	4	0	-	+	+	+	-	Oui	Oui

**Tableau 93:** Réponses aux questionnaires 2 des apprenants du groupe 3 concernant les différences psycho-affectives entre les deux enregistrements et leur rapport au chant corrélée au nombre de continua produits avec la F0 attendue.

Cette analyse permet de considérer les productions des apprenants sous le prisme de la F0 et notamment concernant le groupe 3 qui a eu un *input* avec une F0 montante. Sur le plan acoustique, nous pouvons notamment voir que seuls les apprenants du groupe 3 (4/9) montrent une augmentation moyenne de la F0 entre les phonèmes de chaque continuum. Ceci indiquerait qu'ils aient suivi la F0 induite par l'*input* sonore. Ces résultats sont confirmés par l'analyse perceptive qui montre que quatre apprenants ont produit la F0 attendue lors de l'enregistrement 1 et cinq lors du deuxième. Ces résultats ne peuvent être que partiellement expliqués par des facteurs acquisitionnels ou psychoaffectifs.

Nous avons donc vu que les productions des apprenants ont pu être influencées par les *inputs* proposés et ce notamment en terme de durée pour les phonèmes produits en continuum. Concernant la F0, le groupe 2 ne semble pas avoir été influencé par l'*input* et le groupe 3 propose des résultats mitigés. Reproduire la F0 attendue dans ce cas impliquait un ensemble de caractéristiques cognitives et psychoaffectives qui ont fait émerger des profils différents.

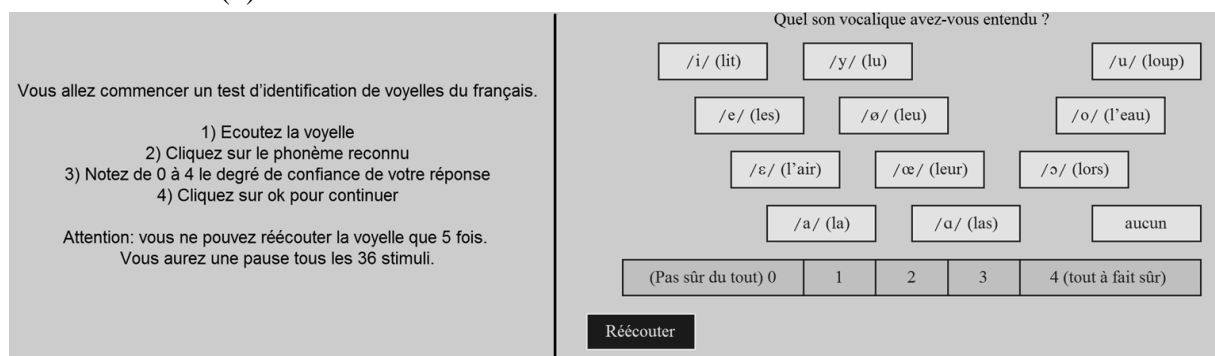
### 6.2.3 Analyse perceptive

Afin de compléter l'analyse des stimuli des apprenants, nous avons mis en place un test d'identification dans le but de repérer de potentielles variations par rapport aux normes internes phonético-phonologiques de francophones natifs.

#### 6.2.3.1 Mise en place

Le test d'identification a été développé sur Praat d'après un script utilisé dans la thèse de Kamiyama (2010). Il se compose de deux échelles (Figure 84) :

- une tâche d'identification : douze réponses possibles parmi les onze phonèmes vocaliques oraux du système maximal du français plus un choix « aucun ».
- une échelle de confiance : cinq degrés de confiance de « pas du tout » (0) à « tout à fait sûr » (4)



**Figure 84:** Test d'identification Praat. Consignes (à gauche) et écran de réponse (à droite).

Nous avons fait le choix de proposer le système maximal du français et non les phonèmes du français contemporain seulement car nous souhaitons obtenir les catégories phonologiques les plus précises ce qui nous permettrait plus tard dans notre analyse d'adapter la grille d'évaluation du système maximal vers un système réduit du français contemporain. De plus, nous avons fait passer le test d'identification à dix locuteurs natifs du français dont le domaine de compétence est la phonétique-phonologie uniquement. Ils étaient donc à même de connaître et reconnaître ces différentes unités phonémiques.

Le test d'identification se compose de quatre parties dissociables pour une durée totale de passation d'environ 35 minutes :

- trois tests (G1, G2, G3) qui comportent respectivement 144 stimuli soit 9 apprenant-es par groupe et pour chaque apprenant-e les occurrences de la deuxième répétition des continua vocaliques « antérieur étiré » (/i e ε a/) et « postérieur arrondi » (/u o ɔ ɑ/) en contexte isolé et en continuum pour le pré-test. Cela représente pour chaque test : 8 phonèmes x 2 contextes x 9 apprenant-e-s.

- un test (E2) : qui comportent respectivement 96 stimuli soit 6 apprenant-es au total (2 par groupe) et pour chaque apprenant-e les occurrences de la deuxième répétition des continua vocaliques « antérieur étiré » (/i e ε a/) et « postérieur arrondi » (/u o ɔ α/) en contexte isolé et en continuum pour le post-test. Cela représente pour chaque test : 8 phonèmes x 2 contextes x 6 apprenant-e-s.

Le test a été passé par 10 francophones natifs (7 femmes et 3 hommes) sans déficience auditive âgés de 20 et 55 ans et qui sont tous dans le domaine de la phonétique et/ou phonologie et/ou didactique de la prononciation. Les tests ont été envoyés par courriel et par conséquent, chaque sujet a passé les tests sur son ordinateur. Il leur était conseillé d'utiliser des écouteurs dans un document d'aide.

### 6.2.3.2 Degré de confiance

Avant d'analyser plus en détail les résultats du test d'identification, nous allons nous intéresser aux résultats du degré de confiance dans la réponse donnée par les auditeurs natifs du français. Ce degré de confiance a été donné après chaque identification sur une échelle de Likert de 0 à 4 (soit 5 niveaux) correspondant respectivement à « pas sûr du tout » à « tout à fait sûr ». Sur le plan individuel et par groupe, nous constatons qu'en moyenne le degré de confiance donné par les auditeurs est élevé et homogène parmi les 27 apprenants testés (min : 4,1 ; max : 4,7 ; Tableau 94).

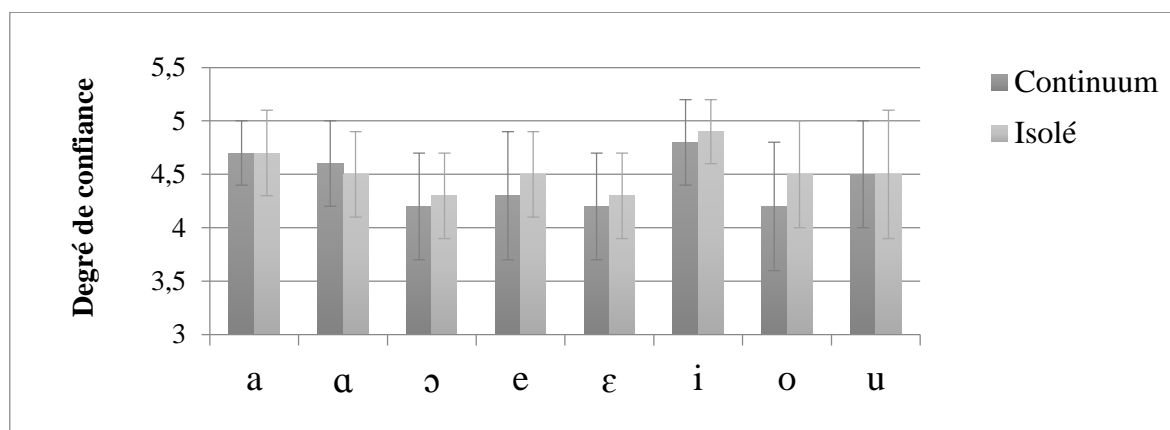
Groupe 1										
Apprenant	DOG	LYQ	LOP	HUA	MEH	DEL	KITT	MAL	NII	TOTAL
Moyenne	4,7	4,5	4,6	4,5	4,1	4,4	4,4	4,6	4,6	4,5
Ecart type	0,4	0,5	0,3	0,4	0,4	0,4	0,4	0,4	0,5	0,4
Groupe 2										
Apprenant	YAN	IAK	SHU	RING	GHE	GOG	QUI	WEN	KIM	TOTAL
Moyenne	4,6	4,6	4,4	4,4	4,5	4,5	4,4	4,5	4,7	4,5
Ecart type	0,5	0,4	0,5	0,7	0,5	0,4	0,4	0,4	0,4	0,4
Groupe 3										
Apprenant	PER	LEO	RAD	DUM	SUN	PAR	NUG	KOZ	FAN	TOTAL
Moyenne	4,6	4,1	4,5	4,4	4,6	4,4	4,4	4,4	4,6	4,4
Ecart type	0,4	0,7	0,5	0,5	0,6	0,5	0,6	0,5	0,4	0,5

**Tableau 94** : Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français en fonction de l'apprenant et du groupe pour l'enregistrement 1.

Nous remarquons également que le degré de confiance moyen est semblable entre les trois groupes 4,5 (0,4) pour les groupes 1 et 2 et 4,4 (0,5) pour le groupe 3. Afin de compléter cette réflexion, nous nous sommes intéressés aux phonèmes (Tableau 95 et Figure 85).

		a	ɑ	ɔ	e	ɛ	i	o	u
Continuum	Moyenne	4,7	4,6	4,2	4,3	4,2	4,8	4,2	4,5
	Ecart type	0,3	0,4	0,5	0,6	0,5	0,4	0,6	0,5
Isolé	Moyenne	4,7	4,5	4,3	4,5	4,3	4,9	4,5	4,5
	Ecart type	0,4	0,4	0,4	0,4	0,4	0,3	0,5	0,6

**Tableau 95 :** Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français en fonction du phonème et de la modalité de production.



**Figure 85 :** Graphique des degrés de confiance au test d'identification par les 10 locuteurs natifs du français en fonction du phonème et de la modalité de production. Les écarts-types sont indiqués.

Nous voyons qu'au regard du degré de confiance et des phonèmes, l'ensemble est homogène que ce soit en production en continuum (min : 4,2 ; max : 4,8) et à l'isolée (min : 4,3 ; max : 4,9). Le /i/ est le phonème pour lequel le degré de confiance est le plus haut quelle que soit la modalité de production. Les voyelles « à double timbre » obtiennent des valeurs plus basses. En général, le degré de confiance est légèrement plus haut à l'isolé qu'en continuum à l'exception du /a/. Ces différences restent toutefois minimes. Pour terminer cette réflexion, nous avons porté notre regard sur les enregistrements 1 et 2 des six apprenants concernés (Tableau 96).

		G1		G2		G3	
Stimuli		DEL	NII	IAK	SHU	PER	LEO
E1	Moyenne	4,4	4,6	4,6	4,4	4,6	4,1
	Ecart type	0,4	0,5	0,4	0,5	0,4	0,7
E2	Moyenne	4,6	4,6	4,7	4,5	4,6	4,2
	Ecart type	0,5	0,4	0,4	0,4	0,4	0,5

**Tableau 96 :** Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français pour les six apprenants retenus lors de l'enregistrement 2.

Une fois de plus, les écarts sont minimes entre les apprenants et les deux enregistrements. Cependant, il est possible de constater qu'aucun des six apprenants ne montre une diminution de la moyenne du degré de confiance entre les enregistrements 1 et 2. Il en va de même concernant les écarts types à l'exception d'une augmentation de 0,1 pour DEL. Pour affiner ces résultats, nous avons analysé le degré de confiance moyen de chaque phonème en fonction de la modalité de production et des enregistrements (Tableau 97).

Stimuli	Continuum			Isole		
	E1	E2	E2-E1	E1	E2	E2-E1
a	4,7	4,7	=	4,7	4,7	=
ɑ	4,8	4,4	-	4,3	4,6	+
ɔ	4,5	4,4	-	4,4	3,9	-
e	4,4	4,6	+	4,5	4,4	-
ɛ	4,1	4,5	+	4,4	4,5	+
i	5,0	5,0	=	4,9	5,0	+
o	4,0	4,3	+	4,4	4,5	+
u	4,4	4,7	+	4,1	4,6	+

**Tableau 97** : Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français pour les six apprenants retenus lors de l'enregistrement 2 en fonction du phonème et de l'enregistrement. L'évolution entre les deux enregistrements est indiquée par un symbole mathématique.

A nouveau, il est possible de remarquer une homogénéité parmi les degrés de confiance moyens entre 4 et 5. Nous remarquons également que le panel de 6 apprenants montre une même tendance que le groupe total (27 apprenants) en ce qui concerne le degré de confiance le plus important pour le phonème /i/. Le phonème /ɔ/ est celui qui en moyenne a été moins bien noté en termes de confiance lors du deuxième enregistrement et ce pour les deux modalités testées. Les phonèmes /ɑ/ et /e/ montrent cette même tendance respectivement pour les modalités en continuum et à l'isolée.

De manière générale, l'analyse du degré de confiance dans la réponse au test d'identification par les dix locuteurs natifs du français ne montre pas de tendance particulière à l'exception d'un fort sentiment moyen de confiance proche de la note maximale (5) et ce dernier est légèrement meilleur pour les voyelles cardinales et à l'isolée.

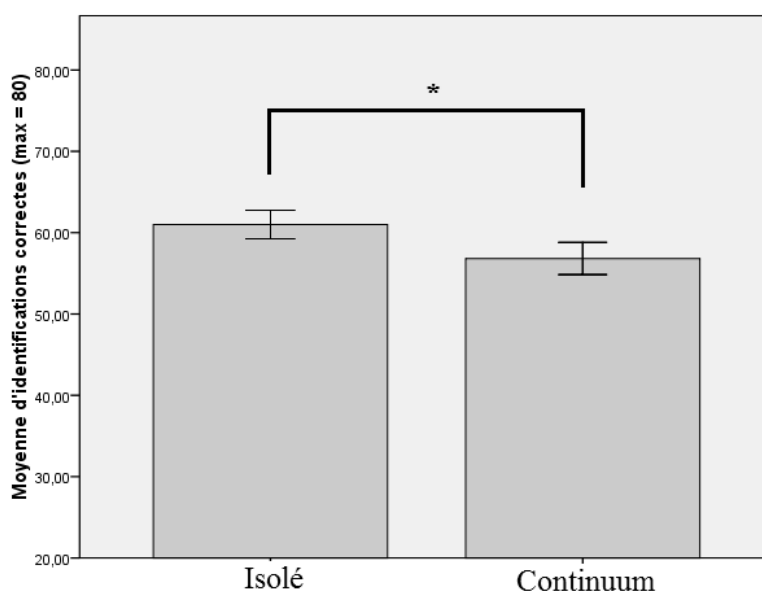
### 6.2.3.3 Analyse perceptive du pré-test

Nous commencerons par analyser les résultats des 9 apprenants des trois groupes expérimentaux. Cette analyse permettra de prendre en considération les différences de production entre les contextes à l'isolée et en continuum mais également de l'impact de la présentation des phonèmes en continuum entre les trois groupes. Nous commencerons par



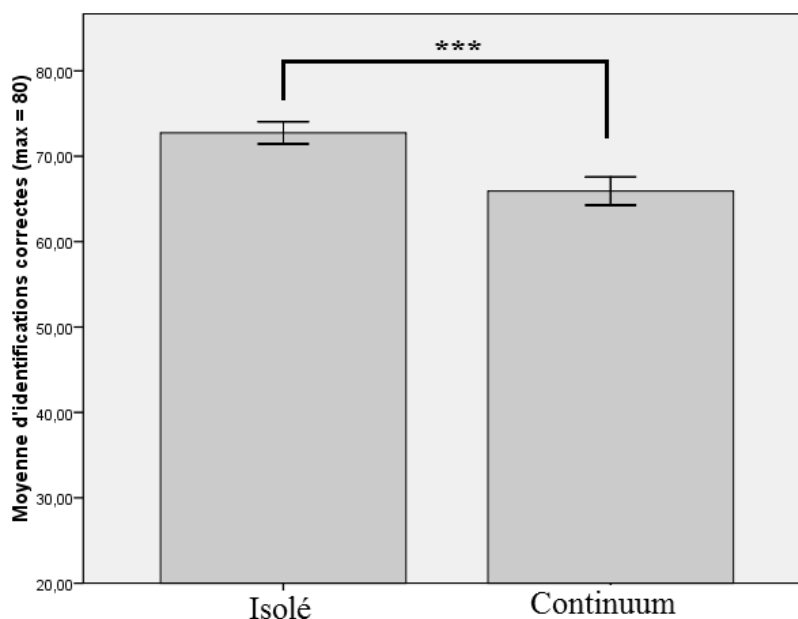
analyser les résultats du test d'identification par les dix locuteurs natifs du français en prenant en compte le système phonologique le plus large soit dix phonèmes vocaliques (sauf nasals et neutralisation de l'opposition /a-ɑ/). Dans un deuxième temps, nous ferons la même analyse en neutralisant tous les archiphonèmes du français contemporain. Nous rappelons que notre test d'identification n'intègre que les deuxièmes répétitions de chaque phonème et continuum.

Pour ce faire, nous avons effectué une analyse de variance (ANOVA à mesures répétées) avec comme facteur intra-sujet les productions à l'isolée et en continuum et comme facteurs inter-sujets les trois groupes expérimentaux. Les résultats statistiques montrent un effet significatif de la modalité de production sur le nombre moyen de réponses correctes en identification soit en isolé (moyenne = 61,0 soit 76,25%, écart type = 9,1) et en continuum (moyenne = 56,8 soit 71,00%, écart type = 10,3) ( $F = 4,749$ ,  $ddl = 1$ ,  $p = 0,039$ ) (Figure 86). Toutefois, l'interaction entre la modalité de production et le groupe n'est statistiquement pas significative ( $F = 0,592$ ,  $ddl = 2$ ,  $p = 0,561$ ).



**Figure 86** : Moyenne d'identification correcte des phonèmes en fonction de la modalité de production selon la grille extrême du français standard. Les erreurs standards sont indiquées.

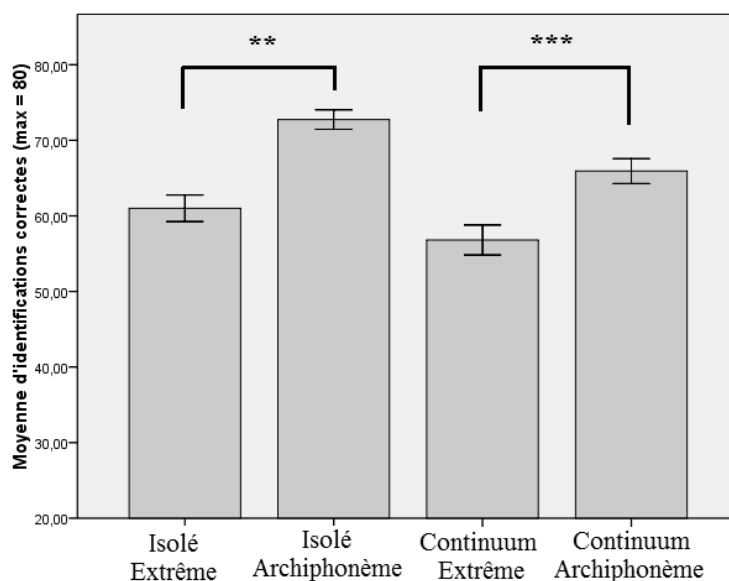
Nous avons mené la même réflexion en neutralisant cette fois les archiphonèmes du français contemporain. Nous obtenons statistiquement les mêmes tendances soit un effet significatif de la modalité de production sur le nombre moyen de réponses correctes en identification soit en isolé (moyenne = 72,7 soit 90,87%, écart type = 6,7) et en continuum (moyenne = 65,9 soit 82,37%, écart type = 8,8) ( $F = 17,880$ ,  $ddl = 1$ ,  $p < 0,0001$ ) (Figure 87). Toutefois, l'interaction entre la modalité de production et le groupe n'est statistiquement pas significative ( $F = 0,264$ ,  $ddl = 2$ ,  $p = 0,771$ ).



**Figure 87** : Moyenne d'identification correcte des phonèmes en fonction de la modalité de production selon la grille avec neutralisation des archiphonèmes du français contemporain. Les erreurs standards sont indiquées.

Les productions à l'isolée semblent donc favoriser une meilleure identification des voyelles des apprenants, tous groupes confondus.

Afin de corrélérer le nombre d'identifications correctes en fonction de la grille de notation (« extrême » ou « archiphonème »), nous avons fait une analyse de variance (ANOVA à mesures répétées) avec comme facteur intra-sujet les productions à l'isolée d'une part et en continuum d'autre part et comme facteurs inter-sujets les trois groupes expérimentaux. Il apparaît une différence significative entre le nombre d'identifications correctes et la grille d'évaluation pour les phonèmes produits à l'isolée avec la grille « extrême » (moyenne = 61 soit 76,25%, écart type = 9,14) et « archiphonèmes » (moyenne = 72,74 soit 90,92%, écart type = 6,71) ( $F = 11,038$ ,  $ddl = 1$ ,  $p = 0,003$ ) et pour ceux produits en continuum avec la grille « extrême » (moyenne = 56,81 soit 71,01%, écart type = 10,30) et « archiphonèmes » (moyenne = 65,92 soit 82,4%, écart type = 8,57) ( $F = 92,919$ ,  $ddl = 1$ ,  $p < 0,0001$ ) (Figure 88). Toutefois, l'interaction entre la modalité de production et le groupe n'est statistiquement pas significative pour les phonèmes produits à l'isolé ( $F = 0,311$ ,  $ddl = 2$ ,  $p = 0,735$ ) et ceux produits en continuum ( $F = 1,137$ ,  $ddl = 2$ ,  $p = 0,338$ ). Les productions montrent donc de meilleurs taux d'identification si elles sont analysées avec la grille « archiphonème » qu'avec la grille « extrême ».



**Figure 88:** Moyenne d'identification correcte des phonèmes en fonction de la grille d'évaluation pour chaque modalité de production. Les erreurs standards sont indiquées.

Pour terminer, nous avons souhaité regarder les principales confusions révélées, tous groupes confondus, dans le but d'analyser plus en détail la nature des confusions. Dans une visée didactique, nous commencerons en neutralisant les archiphonèmes du français contemporain afin de mettre en évidence des confusions phonologiquement pertinentes.

		réponses								Effectif
Stimuli		u	O	A	E	i	y	OE	Aucun	n
Continuum	u	<b>57,0</b>	4,4		0,4		34,1	3,7	0,4	270
	O	4,4	<b>76,9</b>	12,2				3,9	2,6	540
	A		5,6	<b>92,0</b>				0,7	1,7	540
	E				<b>84,3</b>	0,2	3,0	11,7	0,9	540
	i	1,1			3,0	<b>95,9</b>				270
Isolé	u	<b>70,4</b>	4,4			0,4	<b>18,9</b>	5,6	0,4	270
	O	2,6	<b>91,5</b>	3,7	0,2			1,1	0,9	540
	A		<b>1,9</b>	<b>90,4</b>				3,5	4,3	540
	E				<b>98,1</b>	0,7		0,7	0,4	540
	i	1,1			1,1	<b>97,0</b>	0,7			270

		réponses								Effectif
Stimuli		u	O	A	E	i	y	OE	Aucun	n
total	u	<b>63,7</b>	4,4		0,2	0,2	26,5	4,6	0,4	540
	O	3,5	<b>84,2</b>	8,0	0,1			2,5	1,8	1080
	A		3,7	<b>91,2</b>				2,1	3,0	1080
	E				<b>91,2</b>	0,5	1,5	6,2	0,6	1080
	i	1,1			2,0	<b>96,5</b>	0,4			540

**Tableau 98:** Matrices de confusion des 27 apprenant-es en fonction de la modalité de production (en haut) et global (en bas) et de la grille archiphonème. Les résultats sont donnés en pourcentages. Seules les données supérieures à 0 sont indiquées, celles supérieures à 10% sont surlignées en gris et les identifications correctes signalées en gras.

Cette analyse complémentaire permet d'avoir une vision plus générale des confusions phonologiquement pertinentes au système vocalique du français contemporain. Nous voyons, tout d'abord, de voir qu'il y a une symétrie des confusions quelle que soit la modalité même si les phonèmes semblent être moins bien reconnus lorsqu'ils ont été prononcés en continuum qu'à l'isolée (Tableau 98, en haut). De manière globale, le /i/ est très bien reconnu (96,5%), suivi par le /A/ et le /E/ (respectivement 91,2%) puis le /O/ (84,2%) et le /u/ (63,7%). Ce dernier est fortement confondu avec le /y/ (26,5%). En comparant les modalités de production, nous remarquons que lorsque les phonèmes ont été prononcés en continuum le /u/ a plus été confondu avec le /y/ que lorsqu'ils ont été prononcés à l'isolée (+15,2%), le /E/ avec le /OE/ (+11%) et le /O/ avec le /A/ (+8,5%). D'après les deux modalités, le continuum antérieur/étiré a été mieux identifié (92.9% en moyenne) que celui postérieur/arrondi (79.7% en moyenne). De ce fait, nous ne considérerons pas comme un biais le fait que l'association entre notes de musique et l'aperture du continuum antérieur/étiré ait été contre-intuitive avec celle pour le continuum postérieur/arrondi. Si tel était le cas, nous aurions obtenu des résultats un taux d'identification plus faible pour le continuum antérieur/étiré que pour celui postérieur/arrondi. En tout cas, il semblerait que le phénomène de nativisation/assimilation soit plus important que l'aspect contre-intuitif sus-mentionné.

Afin de préciser ces tendances, nous allons prendre la grille extrême dans le but de nous permettre d'affiner notre analyse.

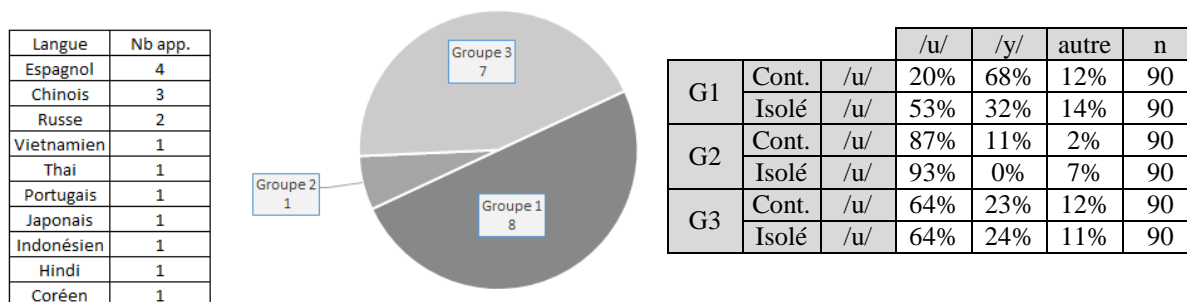
		réponses											
Stimuli		u	o	ɔ	ɑ	a	ɛ	e	i	y	ø	œ	auc.
continuum	u	<b>57,0</b>	4,4					0,4		34,1	3,7		0,4
	o	5,2	<b>72,6</b>	18,5	0,4						0,7		2,6
	ɔ	3,7	21,1	<b>41,5</b>	13,3	10,7					2,6	4,4	2,6
	ɑ		0,4	5,6	<b>11,5</b>	79,6						1,5	1,5
	a			5,2	12,2	<b>80,7</b>							1,9
	ɛ						<b>57,8</b>	27,4	0,4	0,4	8,5	4,1	1,5
	e						24,1	<b>59,3</b>		5,6	3,7	7,0	0,4
i	1,1						<b>3,0</b>	<b>95,9</b>					
isolé	u	<b>70,4</b>	4,4						0,4	18,9	5,6		0,4
	o	4,1	<b>80,4</b>	13,7				0,4				0,7	0,7
	ɔ	1,1	45,2	<b>43,7</b>	3,7	3,7					1,1	0,4	1,1
	ɑ			2,6	<b>11,9</b>	76,3						4,1	5,2
	a			1,1	10,7	<b>81,9</b>						3,0	3,3
	ɛ						<b>60,7</b>	37,0			0,4	1,1	0,7
	e						24,8	<b>73,7</b>	1,5				
i	1,1						1,1	<b>97,0</b>	0,7				

**Tableau 99:** Matrices de confusion des 27 apprenant-es en fonction de la modalité de production et de la grille extrême. Les résultats sont donnés en pourcentages (n=540/ligne). Seules les données supérieures à 0 sont indiquées, celles supérieures à 10% sont surlignées en gris et les identifications correctes signalées en gras.

		réponses											
Stimuli		u	o	ɔ	ɑ	a	ɛ	e	i	y	ø	œ	auc.
Total	u	<b>63,7</b>	4,4					0,2	0,2	26,5	4,6		0,4
	o	4,6	<b>76,5</b>	16,1	0,2			0,2			0,4	0,4	1,7
	ɔ	2,4	33,1	<b>42,6</b>	8,5	7,2					1,9	2,4	1,9
	ɑ		0,2	4,1	<b>11,7</b>	78,0						2,8	3,3
	a			3,1	11,5	<b>81,3</b>						1,5	2,6
	ɛ						<b>59,3</b>	32,2	0,2	0,2	4,4	2,6	1,1
	e						24,4	<b>66,5</b>	0,7	2,8	1,9	3,5	0,2
	i	1,1						2,0	<b>96,5</b>	0,4			

**Tableau 100:** Matrices de confusion globale des 27 apprenant-es globale et de la grille extrême. Les résultats sont donnés en pourcentages (n=1080/ligne). Seules les données supérieures à 0 sont indiquées, celles supérieures à 10% sont surlignées en gris et les identifications correctes signalées en gras.

Au regard du Tableau 99 et du Tableau 100, nous constatons les mêmes tendances que celles observées dans le Tableau 98. Nous constatons toujours que le /u/ est reconnu à 63,7% mais il a été perçu également comme un /y/ à 26,5%. Cette confusion a été identifiée par au moins un des francophones natifs pour 16 apprenant-es dont une minorité du groupe 2 (1 apprenant-e contre 7 pour le groupe 3 et 8 pour le groupe 1) et des origines linguistiques variées mais dont 4 hispanophones, 3 sinophones et 2 russophones. Cette confusion est en moyenne plus importante lorsque le /u/ est produit en continuum pour les groupes 1 et 2. Mais les trois groupes semblent réagir différemment puisqu'à l'isolée nous constatons que la confusion /u-y/ représente 32% pour le groupe 1, 0% pour le groupe 2 et 24% pour le groupe 3 (Figure 89).



**Figure 89:** Apprenant-es concerné-es par au moins une confusion /u-y/ en fonction de la langue première et du groupe expérimental (à gauche) et tableau du pourcentage de confusion en fonction du groupe (à droite).

Nous noterons, cependant, qu'au regard de l'hétérogénéité des langues premières des apprenants d'un dispositif d'apprentissage en France, cela a pu jouer comme un biais dans les tendances mentionnées par groupe au regard de la confusion /u-/y/.

Gr	L1	/u/	/y/	Gr	L1	/u/	/y/	Gr	L1	/u/	/y/
1	portugais	x		2	russe	x		3	hindi	x	
	espagnol	x			vietnamien	x			roumain	x	
	chinois (taiwanais)	x	x		espagnol	x			persan	x	
	coréen	x			chinois	x	x		roumain	x	
	chinois	x	x		russe	x			chinois	x	x
	suédois	x			thaï	x			espagnol	x	
	géorgien	x			espagnol	x			russe	x	
	indonésien	x			chinois (taiwanais)	x	x		chinois	x	x
japonais	x		russe	x		coréen	x				

**Tableau 101** : Langues premières des apprenants en fonction des trois groupes (Gr.) et présence phonologique ou non des sons /u/ et /y/ dans les différentes langues d'après UCLA Phonological Segment Inventory Database (UPSID) (Maddieson, 1984). Le /y/ apparaît également pour le coréen dans UPSID mais il semblerait que cette existence soit remise en question surtout concernant le coréen contemporain (Schuh, 2013).

Cependant, nous pouvons constater, à l'aide du Tableau 101, que malgré cette diversité, l'intégralité des langues premières des apprenants possèdent le /u/ et que seul le chinois semble posséder le /y/. Deux locuteurs dans chaque groupe sont sinophones ce qui ne semble pas permettre d'expliquer les différences intergroupes susmentionnées pour l'opposition /u/-/y/.

		Réponses							
Stimuli		u	O	A	E	i	y	OE	Aucun
G1	u	-33	1		1	-1	<b>36</b>	-3	
	O	4	-8	4	-1			-2	2
	A		9	2				-8	-3
	E				-18	-1	5	<b>13</b>	1
	i				2		-2		
G2	u	-7	-4				<b>11</b>	-1	1
	O	-1	-14	<b>11</b>				1	3
	A		1	-1					1
	E				-6	-1		6	
	i				4	-4			
G3	u		3				-1	-1	-1
	O	2	-22	<b>10</b>				9	1
	A		2	4					-6
	E				-18		4	<b>13</b>	1
	i				-1	1			

**Tableau 102**: Comparaison des matrices de confusions entre les modalités « continuum » et « isolé » en fonction du groupe et selon la grille « archiphonème ». Les résultats sont donnés en pourcentage. Un nombre négatif indique qu'en moyenne cette confusion est plus apparue à l'isolée et un nombre positif indique que cette confusion est plus apparue en continuum. En gris foncé, les données supérieures à 10% et en gris clair, inférieures à -10%.

Les voyelles moyennes /e-ɛ/ et /o-ɔ/ sont principalement confondues entre paires mi-ouverte/fermée même si la voyelle mi-fermée est toujours mieux reconnue par les francophones

natifs que sa correspondante mi-ouverte (/e/ + 33,9% = /ɛ/ et /o/ + 7,2% = /ɔ/). Une différence notable entre les modalités de production est une plus grande variété de confusion en continuum qu'à l'isolée où notamment le /ɔ/ est perçu à 10,7% comme un /a/ et le /ɛ/ à 8,5% comme un /ø/ et le /e/ à 7,0% comme un /œ/.

Afin de comparer les matrices de confusion des trois groupes, nous nous sommes intéressés aux différences entre les confusions des productions en continuum et celle à l'isolée (Tableau 102). Nous remarquons que seul le groupe 1 a d'une manière importante plus fait la confusion entre /u-y/ lors de la production en continuum (36%) même si le groupe 2 montre la même tendance mais de manière plus réduite (11%). De manière générale, nous remarquons que les voyelles moyennes sont moins bien identifiées lorsqu'elles sont produites en continuum qu'à l'isolée. Les groupes 1 et 2 montrent une même tendance pour une confusion plus importante en continuum entre /E-OE/ (13% chacun). Pour finir, les groupe 2 et 3 affichent une moins bonne reconnaissance de /E/ en continuum au profit d'une confusion /O-A/ (respectivement 11% et 10%). Grâce au Tableau 103, nous allons voir plus en détail les différences entre les groupes et les modalités de production en nous concentrant sur les oppositions /e-ɛ/ et /o-ɔ/.

		réponses											
Stimuli		e	ɛ	o	ɔ	a	ɑ	i	u	y	ø	œ	aucun
G1	e	-22	-11					-3		10	6	20	1
	ɛ	4	-7					1			1		
	o	-1		-10	-1				9		1	-2	4
	ɔ			-4		1	8				-2	-1	-1
G2	e	-9	10					-1					
	ɛ	-16	3								9	3	
	o			-10	10		1		-2		1		
	ɔ			-24	-3	1	20				0	1	6
G3	e	-12	-1							7	6	1	
	ɛ	-18	-6							1	14	6	2
	o			-3	6				-3				1
	ɔ			-43	-3	19	1		8		7	12	

**Tableau 103:** Comparaison des matrices de confusion pour les phonèmes /e-ɛ/ et /o-ɔ/ entre les modalités « continuum » et « isolé » en fonction du groupe et selon la grille « extrême ». Les résultats sont donnés en pourcentage. Un nombre négatif indique qu'en moyenne cette confusion est plus apparue à l'isolée et un nombre positif indique que cette confusion est plus apparue en continuum. En gris foncé, les données supérieures à 10% et en gris clair, inférieures à -10%.

Nous confirmons que de manière générale, les voyelles moyennes sont moins bien identifiées lorsqu'elles sont produites en continuum qu'à l'isolée. De plus, ces confusions sont rarement

au profil de leur homologue (mi-fermée ou mi-ouverte). Pour le groupe 1, le /e/ est en moyenne majoritairement plus confondu avec le /y/ (+10%) ou le /œ/ (+20%) et pour le groupe 2, le /e/ est certes plus confondu avec le /ɛ/ (+10%) mais ce dernier est plus confondu avec le /ø/ (+9%) tout comme le groupe 3 (+14%). Concernant les phonèmes /o/ et /ɔ/, nous remarquons pour le groupe 3 que produit en continuum, le /ɔ/ est largement moins perçu comme un /o/ (-43%) et ce au profit d'une confusion avec /a/ (+19%) et /œ/ (+12%) et pour le groupe 2 avec /ɑ/ (+20%) comme pour le groupe 1 (+8%). Pour les groupes 2 et 3, le /o/ a tendance à être plus confondu avec le /ɔ/ (respectivement +10% et +6%) et le groupe 1 avec le /u/ (+9%).

Nous avons que, pour l'enregistrement 1 et de manière globale, les phonèmes sont bien reconnus par les dix locuteurs natifs francophones d'après la grille « archiphonème » à l'exception du /u/. Cependant, la grille « extrême » nous indique un nombre important de confusions entre les homologues des paires de voyelles moyennes. Nous constatons aussi un nombre important de confusion entre /ɔ/ et /ɑ/ ou /a/ en contexte de continuum principalement pour les groupes 2 et 3. Nous avons pu également montrer des différences entre les groupes au regard des résultats en production à l'isolée et en continuum.

### 6.2.3.4 Analyse perceptive du post-test

Dans cette analyse, nous comparerons les résultats d'identification de deux sujets par groupe soit six au total. Comme mentionné précédemment, nous avons dû limiter cette analyse à six individus pour des raisons de qualité insuffisantes des enregistrements (saturation) lors de la deuxième session d'enregistrement. De ce fait, cette analyse se présentera sous la forme d'une analyse de cas et ne représentera pas des phénomènes généralisables. De plus, les six sujets ne sont pas représentatifs du groupe puisque leurs enregistrements n'ont pas saturé par le fait d'une voix de plutôt faible intensité que nous pourrions corrélérer avec un degré de confiance en soi ou d'investissement qui peuvent affecter la précision de la production vocale. Nous noterons, dès maintenant, que les six sujets ont des L1 différentes pour le groupe 1 (japonais et espagnol), le groupe 2 (russe et chinois) et le groupe 3 (espagnol et portugais). Nous ne prendrons pas en compte l'effet inter-sujet du groupe pour des raisons statistiques. Nous avons donc comparé les données du pré-test et post-test avec notre grille « extrême » puis « archiphonème » à l'aide d'un test t pour échantillons appariés. Nous constatons dès lors que quelle que soit la modalité de production la différence entre le pré- et post-test n'est pas significative. Il n'y a pas non plus de différences significatives lorsque nous prenons la grille « archiphonème » (Tableau 104).



Grille	Indicateurs	Groupes	n=	Moyenne	Test t	Valeur de p	<0,05 = *
Extrême	Isolé	T1	6	63,00	-1,728	0,145	ns
		T2	6	67,83			
	Continuum	T1	6	59,33	-1,195	0,286	ns
		T2	6	64,17			
Archiphonèmes	Isolé	T1	6	71,00	-1,439	0,210	ns
		T2	6	74,83			
	Continuum	T1	6	67,50	-1,070	0,334	ns
		T2	6	70,50			

**Tableau 104:** Résultats au Test t pour échantillons appariés entre les pré- et post-test pour les deux modalités de production avec la grille « extrême » et « archiphonème ».

Même s'il n'est pas possible de mesurer de différences significatives entre les pré- et post-tests, nous pouvons cependant constater qu'en moyenne sur les six apprenant-es et quelle que soit la modalité ou la grille, les phonèmes sont mieux reconnus par les francophones natifs de 6% pour la grille extrême et 4,27% pour la grille archiphonème.

Grille	Archiphonème				Extrême			
	Isolé		Continuum		Isolée		Continuum	
	Eff.	Moy.	Eff.	Moy.	Eff.	Moy.	Eff.	Moy.
Progression	3	9	3	8,3	3	10	3	12,7
Stagnation	2	/	0	/	2	/	1	/
Régression	1	-4	3	-2,3	1	-1	2	-4,5

**Tableau 105 :** Nombre d'apprenants montrant une progression, stagnation ou régression entre les deux enregistrements en fonction de la grille et de la modalité. Les moyennes (Moy.) correspondent au nombre moyen d'identifications (n=6).

Plus en détail, nous constatons, d'après le Tableau 105, qu'il y a un nombre constant d'apprenants qui ont progressé, soit au moins trois sujets quelle que soit la grille ou la modalité. Seule la modalité en continuum montre que 3 apprenants ont eu un moins bon score d'identification lors de l'enregistrement 2 par rapport au 1<sup>er</sup>. Néanmoins, nous pouvons constater qu'en moyenne, les apprenants qui ont progressé montrent une tendance de progression plus forte (en moyenne +10) que la tendance des apprenants qui affichent une régression (en moyenne -3).

Nous allons maintenant regarder les principales différences entre les confusions révélées tous groupes confondus dans le but d'analyser les différences entre le pré-test et le post-test. De manière qualitative, nous constatons à l'aide du Tableau 106 (en haut) que les différents phonèmes ont été mieux identifiés lors du post-test que lors du pré-test. Le /u/ a été plus largement identifié comme un /u/ au post-test quelle que soit la modalité (+20% en continuum et +30% à l'isolée), le /E/ a été mieux identifié à +11,7%. Nous notons que le /A/ a été mieux

identifié comme un /A/ (+4,2%) mais il a également été plus confondu avec /O/ (+2,5%) en production à l'isolée. En continuum, le /A/ a été moins bien identifié comme un /A/ (-5,8) mais plus comme un /O/ (+5). De la même manière, le /O/ du post-test a été plus perçu comme un /A/ (+5,8%).

		Stimuli	réponses							Aucun	
			u	O	A	E	i	y	OE		
T2-T1	Continuum	u	<b>20</b>	-10					-1,7	-8,3	
		O	-4,2	-0,8	5,8						-0,8
		A		5	-5,8						0,8
		E				<b>11,7</b>	-0,8	-8,3	-2,5		
		i									
	Isolé	u	<b>30</b>	1,7					<b>-23,3</b>	-8,3	
		O	-3,3	-1,7	0,8					3,3	0,8
		A		2,5	4,2					-0,8	-5,8
		E									
		i	-1,7				-1,7	3,3			

**Tableau 106:** Différence entre les matrices de confusion du post-test et du pré-test des 6 apprenants en fonction du type de grille archiphonème (en haut) et extrême (en bas). Les résultats sont donnés en pourcentages. Seules les données différentes de 0 sont indiquées. Un pourcentage positif indique que l'opposition a plus été identifiée au post-test qu'au pré-test (en gris clair) et un pourcentage négatif indique que l'opposition a moins été identifiée au post-test qu'au pré-test (en gris foncé).

		Stimuli	réponses										Auc.	
			u	o	ɔ	ɑ	a	ε	e	i	y	ø		œ
T2-T1	continuum	u	20,0	-10,0							-1,7	-8,3		
		o	-8,3	11,7	-1,7									-1,7
		ɔ		-21,7	10,0	8,3	3,3							
		ɑ			11,7	1,7	-15,0							1,7
		a			-1,7	-6,7	8,3							
		ε						23,3	-20,0	-1,7	-1,7	-3,3	3,3	
		e						3,3	16,7		-15,0	-5,0		
	isolé	u	30,0	1,7							-23,3	-8,3		
		o	-6,7	18,3	-11,7									
		ɔ		-23,3	13,3	-3,3	5,0					3,3	3,3	1,7
		ɑ			5,0	-6,7	11,7						-1,7	-8,3
		a				-6,7	10,0							-3,3
		ε						13,3	-13,3					
		e						6,7	-6,7					
i	-1,7						-1,7	3,3						

Afin d'analyser l'évolution plus en détail, nous allons corrélérer les résultats du Tableau 106 (haut et bas). De manière globale, nous constatons que l'ensemble des phonèmes a été mieux identifié lors du post-test quelle que soit la modalité à l'exception du /ɑ/ (-6,7%) et du /e/ (-6,7%) à l'isolée et du /i/ en continuum dont les résultats ont stagné. Nous pouvons constater que dans les deux cas le /ɑ/ est celui qui montre une tendance positive soit plus faible soit négative. En continuum, il est légèrement plus perçu comme un /ɑ/ (+1,7%) mais il est moins perçu comme

son homologue phonémique /a/ (-15%) et plus comme un /ɔ/ (+11,7%). A l'isolée, il est moins identifié comme un /ɑ/ (+6,7%) mais plus comme un /ɔ/ (+5%) ou comme /a/ (11,7%). Il est également possible de noter qu'à l'isolée le /ɔ/ a tendance lors du post-test à être plus identifié comme les phonèmes vocaliques arrondis mi-fermé/ouvert antérieurs (+3,3%) ou non-identifié (1,7%). Remarquons que le /e/ à l'isolée a tendance à être moins identifié comme un /e/ (-6,7%) mais comme un /ɛ/ (+6,7%).

### **6.3 Analyses et discussions**

A travers cette étude, nous souhaitons questionner deux éléments expérimentaux à savoir le caractère potentiellement développemental du dispositif d'apprentissage proposé mais également celui d'un outil didactique reliant phonèmes et notes de musique. Nous discuterons ces deux éléments de manière thématique.

#### **6.3.1 Représentations et progression générale**

Comme nous l'avons vu, le dispositif proposé avait pour but d'amener les apprenants à se sentir plus à l'aise dans la prononciation des sons du français à travers une compréhension du système phonético-phonologique de ce code langagier ainsi que des exercices visant les automatismes articulatoires / perceptifs. L'ensemble de ce travail abordait une déconstruction des représentations et habitudes langagières des apprenants afin de favoriser l'apparition de ruptures dans le développement langagier du code français. Nous avons tenté à l'aide de notre protocole de recherche de repérer puis quantifier ces ruptures. De manière générale, nous avons pu constater que soit lors du troisième questionnaire, soit à la fin de la formation, l'ensemble de 27 apprenants déclarait avoir perçu des changements dans leur prononciation depuis le début du semestre. Nous noterons qu'ils étaient 23 lors du deuxième questionnaire à répondre positivement à cette même affirmation. De manière plus spécifique, nous avons questionné les apprenants à trois moments du dispositif d'apprentissage concernant une auto-évaluation de cinq points : (1) leur français à l'oral, (2) leur prononciation générale du français, leur prononciation (3) des voyelles, (4) des consonnes et de (5) la mélodie du français. Lors du questionnaire 2, seule l'évolution de la mélodie du français était significative avec une augmentation de 0,6 points ( $p = 0,001$ ), même si l'ensemble des points mesurés montrait une augmentation comprise entre 0,1 et 0,4 points. Toutefois, nous avons remarqué que lors du questionnaire 3 et ce par rapport au premier, quatre points affichaient une progression significative à l'exception du français oral avec une progression de 0,3 points. Pour conforter

cette évolution positive, nous avons observé que parmi l'ensemble des 27 apprenants, la note maximale (5) était plus sélectionnée lors du questionnaire 3 (13 occurrences) que lors du questionnaire 2 (6 occurrences) ou 1 (aucune occurrence) tandis que la note minimale (1) stagnait à deux occurrences pour les trois questionnaires. Les notes 2 et 3 ont moins été sélectionnées (de 35 occurrences au total) alors que les notes 4 augmentaient (22 occurrences). Au regard de la progression mesurée, soit la comparaison de l'évaluation donnée pour chacun des cinq items susmentionnés entre les questionnaires 1 et 2 ainsi que 1 et 3, nous remarquons une diminution des régressions (de 12% à 7%) et une augmentation des stagnations (47% à 51%) et des évolutions positives (de 41% à 44%). Il est évident que cette tendance positive dans l'auto-évaluation des cinq points mentionnés ne peut pas être que le reflet du travail mené en présentiel puisque dans notre perspective la cours n'est pas un univers clos. Cela se reflète notamment lors 23 apprenants déclarent lors du questionnaire 3 avoir pratiqué le français à l'oral « très souvent » ou « souvent » et 10 affirment avoir beaucoup travaillé sur leur prononciation en dehors du cours de prononciation. Néanmoins, un nombre important d'apprenants déclare que les parties théoriques et en groupe leur ont permis de mieux comprendre la prononciation du français : respectivement 16/27 et 19/27 ont évalué cette aide de « 1 » ou « 2 » sur une échelle à 5 degrés (« 1 » étant le plus positif). Les réponses sont cependant asymétriques en ce qui concerne l'aide que cela a apporté dans l'amélioration de leur prononciation du français étant donné que 20/27 l'ont attribué au travail en groupe et/ou 11/27 à celui de la partie théorique. Nous avons également noté qu'aucun des apprenants n'a donné la note minimale à l'apport de la partie en groupe tandis que nous en dénombrons 7 pour les deux questions relatives à la partie théorique. Il semblerait donc de manière claire que d'après les représentations des apprenants, la partie pratique en groupe a été évaluée de manière plus positive sur le plan potentiellement développemental que la partie théorique. Il semblerait que cette dernière ait eu, selon eux, une influence sur la compréhension du système phonologique mais que son aide pour améliorer la prononciation n'est pas si tranchée. Cependant, il semblerait que l'utilisation des symboles phonétiques ait été perçue comme facilitante pour mieux percevoir et prononcer les voyelles du français, respectivement, avec une moyenne de 2,3 (écart-type = 1,2) et 2,4 (écart-type = 1,1) sur une échelle allant de « 1 » (tout à fait d'accord) à « 5 » (pas du tout d'accord). Dans l'ensemble et au regard des représentations des apprenants, il est clair que le dispositif d'apprentissage mis en place montre, de manière constante, un impact positif sur les représentations des apprenants concernant leur développement langagier. Cet impact semble d'autant plus important au fur et à mesure que le dispositif avance puisqu'il faut attendre le troisième questionnaire pour que l'ensemble des apprenants déclare avoir perçu

des changements dans leur prononciation depuis le début du cours. Ces représentations ont un rôle majeur dans le sentiment de progression nécessaire à un apprenant pour se sentir confiant dans le processus de développement langagier et donc de favoriser l'émergence de ruptures au niveau linguistique.

### **6.3.2 Impact du groupe et de la modalité de production**

Notre but était de mesurer la potentialité acquisitionnelle d'une association des phonèmes en continua avec des notes de musique. Pour cela, les groupes 2 et 3 avaient un input auditif. Afin de prendre en compte l'impact de cet input auditif sur les productions, nous nous sommes intéressés à la durée des phonèmes produits ainsi qu'à la fréquence fondamentale. Nous avons pu voir, de manière statistique, que les groupes 2 et 3 affichaient de manière significative des phonèmes plus longs lorsqu'ils étaient produits en continua, soit après écoute des stimuli auditifs. Ce phénomène se retrouve lors des deux enregistrements avec en moyenne une augmentation de 211,41 ms. Nous notons que les phonèmes dans les stimuli avaient une durée moyenne de 1338,8 ms. Ce phénomène ne se retrouve pas en ce qui concerne le groupe 1 entre les phonèmes produits à l'isolée et en continuum et entre les trois groupes à l'isolée puisque produits sans input. Il semblerait donc clairement que l'input ait été pris en compte par les apprenants des groupes 2 et 3 et cela émerge dans la durée des phonèmes produits. Toutefois, ce phénomène est plus limité au regard de la F0 puisque sur le plan perceptif, il n'a pas été possible de dissocier les groupes 1 et 2. Cependant, nous avons pu mettre en avant différents profils au sein du groupe 3 : deux apprenants ont produits à chaque fois la F0 attendue, quatre ont produit les deux continua soit lors du 1<sup>er</sup>, soit du 2<sup>ème</sup> enregistrement, et un apprenant ne l'a réalisée que lors d'un continua au premier enregistrement. Il semblerait que ces comportements ne puissent pas directement être corrélés avec des facteurs psychoaffectifs car ils sous-tendraient des processus complexes psychologiques. Nous remarquons, cependant, que les deux apprenants du groupe 3 qui ont moins produit les formes prosodiques attendues sont ceux qui ont soit trouvé que cette association ne les avait pas aidé à percevoir ou produire les phonèmes vocaliques, soit qui ne se sont pas sentis plus confiants, plus à l'aise et plus libres lors du 2<sup>ème</sup> enregistrement. Dans l'ensemble, les apprenants, quel que soit le groupe, lors du deuxième enregistrement, se sont sentis moins stressés, plus à l'aise, confiants et libres, ce qui pourra être un facteur expliquant une progression entre les deux enregistrements.

Concernant l'association proposée au groupe 3, il apparaît que son utilité soit sujette à discussion. Lors du questionnaire 2, une majorité des apprenants indiquait qu'associer le piano

aux voyelles les avait aidé à mieux percevoir leurs différences (8/9) ou à mieux les prononcer (6/9). Toutefois, lors du questionnaire 3, aux questions « quand je dois prononcer les voyelles du français, cela m'aide de repenser aux notes de piano » ou « je repense aux notes de piano », deux profils extrêmes s'expriment : ceux qui sont tout à fait d'accord (respectivement 6/9 et 4/9) et ceux qui sont pas du tout d'accord (3/9 pour les deux). Au regard de ces représentations, une première idée semblerait se dégager concernant l'utilisation de cette association qui serait de nature bimodale pour ceux qui y seraient sensibles et ceux qui ne le seraient pas. Cependant, seul un apprenant sur neuf s'est montré fortement en désaccord concernant le fait qu'associer le piano aux voyelles l'a aidé à mieux les percevoir ou prononcer.

Néanmoins, nous devons noter que sur le plan statistique, il n'a pas été possible de montrer de différence significative entre les trois groupes, au sujet des cinq points mesurés (oral, prononciation générale, voyelles, consonnes et mélodie), sur le plan de la progression aux auto-évaluations entre les questionnaires 1 et 2 ainsi que 1 et 3. Ils montrent tous trois le même type de progression, soit une légère augmentation. Cette similitude entre les trois groupes, se retrouve dans la moyenne du degré de confiance donné par les auditeurs natifs du français lors des tests d'identification (environ 4,5/5). Nous n'avons pas pu montrer d'effet d'interaction significatif entre la modalité de production et le groupe, et ce quelle que soit la grille d'évaluation (« extrême » ou « archiphonème »). Il semblerait dès lors, d'après notre protocole expérimental et notre dispositif d'apprentissage, qu'il y ait une potentialité acquisitionnelle cognitive de l'association proposée entre phonèmes et notes de musique ; néanmoins cette dernière montre une tendance bimodale et peut ne pas convenir à tous les profils d'apprenants. De plus, d'après notre protocole, il n'a pas été possible de montrer que cette association avait plus de potentialité développementale que la simple présentation du trapèze vocalique. Cette association apparaîtrait, dès lors, plus comme un outil supplémentaire à l'enseignant dans la prise en compte de la pluralité des profils d'apprentissage. Ce serait un outil cognitif à l'image du « bonhomme sonore » qui favoriserait l'abstraction des différences phonémiques mais ce ne saurait être une révolution acquisitionnelle et pédagogique. Cependant, contrairement à ce que nous pouvions penser, il semblerait que les phonèmes produits à l'isolée soient significativement mieux reconnus par les locuteurs natifs du français que ceux produits en continuum, et ce quel que soit le groupe. La production en continuum pourrait même renforcer les confusions entre /ɔ/ et /a-a/. Il apparaîtrait donc que notre hypothèse basée sur la logique articulatoire des continua n'est pas un élément pertinent dans la réalisation prototypique des phonèmes du français. La production à l'isolée serait celle qui favoriserait le mieux cet exercice. Nous pourrions penser que le poids cognitif de production de quatre phonèmes à la suite serait

trop important et ne permettrait pas à l'apprenant d'atteindre une cible précise articulatoire ou perceptive.

### **6.3.3 Confusions phonémiques et dénativisation/accommodation phonologique**

De manière générale, nous avons souhaité prendre en considération une vision à la fois réductionniste et sociale de la didactique de la prononciation. Nous avons, pour cela, fait le choix d'utiliser deux grilles d'évaluation soit « extrême » (11 phonèmes + aucun), soit « archiphonème » (7 phonèmes + aucun). Tout d'abord, nous avons pu montrer, de manière significative, que la grille choisie influence le nombre moyen d'identifications correctes et ce quelle que soit la modalité de production. Le taux d'identification correcte pour la grille « archiphonème » est en moyenne supérieur d'environ 13% à celui de la grille « extrême ». Il est donc clair que la grille choisie pour évaluer les productions des apprenants est déterminante dans le score obtenu par ces derniers à une évaluation. De manière générale et d'après la grille « archiphonème », nous remarquons que la plupart des phonèmes sont reconnus au-delà de 80% à l'exception du /u/ qui est reconnu à 63,7% et confondu à 26,5% avec le /y/. Cette confusion peut être expliquée par deux éléments dont un phonologique et un expérimental. Tout d'abord, le /y/ est peu présent dans les langues du monde, soit environ 26 langues sur les 451 répertoriées dans UPSID (*UCLA Phonetic Segmental Inventory Database* ; Maddieson, 1984) contre 369 pour le /u/ (*Ibidem*). Cet aspect de phonologie comparée permet dans notre cadre théorique d'identifier le /y/ comme un phonème dont la nativisation/assimilation phonologique pourra être rencontrée par un grand nombre d'apprenants du français. L'autre source de confusion peut porter sur le fait qu'à chaque fois le trapèze vocalique était fourni - et par conséquent les phonèmes - l'était en API. Or, il y a confusion graphique entre le /u/ qui se prononce dans l'alphabet [y] mais dont la graphie est généralement « ou », et le /y/ qui généralement en français se prononce [i] dans l'alphabet mais dont la graphie est « u ». Nous avons vu, à ce titre, que cette confusion était plus importante pour les groupes 1 et 2, lorsque le /u/ était prononcé en continuum. Le groupe 3 est celui pour lequel cette confusion est la plus stable (24% à l'isolée et 23% en continuum) et le groupe 1 où la différence est la plus importante (32% à l'isolée et 68% en continuum). Le groupe 2 montre une augmentation de cette confusion entre isolé et continuum de 11%. Nous pourrions supposer ici que l'input fourni a permis de limiter la confusion, c'est-à-dire que les apprenants des groupes 2 et 3 auraient basé leur production plus sur ce qu'ils entendaient plutôt que ce qu'ils voyaient. Au regard de la grille « extrême » et au

regard des confusions pour les voyelles moyennes, nous avons pu constater comme attendu qu'au total, les phonèmes /e-ɛ/, /o-ɔ/ et /a-ɑ/ ont principalement été confondus ensemble et qu'en général, les phonèmes mi-fermés ont été mieux reconnus que leurs homologues mi-ouverts. Le /ɑ/ a majoritairement été confondu par un /a/ à 78% ce qui semble pertinent au regard de la disparition de cette opposition phonologique en français contemporain. Nous tenons à rappeler ici que cette opposition phonologique a été maintenue, bien que réalisée par très peu de locuteurs francophones, afin de conserver la logique articulatoire du continuum arrondi/postérieur. Il en va de même pour les oppositions /e-ɛ/, /o-ɔ/. Ces disparitions ont pu jouer sur la production des apprenants mais également sur l'évaluation des locuteurs natifs du français qui, bien qu'experts en phonétique/phonologie/prononciation, peuvent identifier de manière moins précise le /ɑ/ ou les oppositions entre voyelles moyennes en fonction de facteurs environnementaux (niveau générationnel, social, etc.). En ce qui concerne les différences entre les trois groupes et les confusions, nous avons vu qu'en moyenne, pour le groupe 1, le /u/ avait moins été identifié comme un /u/ lorsque produit en continuum (33%) et plus perçu comme un /y/ (36%). Cette tendance se retrouve pour le groupe 2 mais de manière plus limitée (+11% de confusion /u-y/) mais pas du tout pour le groupe 3 dans lequel les tendances sont similaires entre isolé et continuum. Il semblerait dès lors que la production en continuum avec pour seul appui : le trapèze vocalique (input visuel), ait été un facteur augmentant les confusions entre /u-y/. L'input audio proposé au groupe 3 ne semble pas avoir amélioré les productions des apprenants mais semble avoir limité une tendance vers une plus grande confusion /u-y/ tant en continuum qu'à l'isolée. D'autres différences entre les trois groupes ont pu être mises en évidence concernant les voyelles moyennes. Ces dernières sont, en général, moins identifiées en continuum qu'à l'isolée mais également que ce processus n'est pas au profit de leur homologue (mi-ouvert/mi-fermé). Le /e-ɛ/ est plus reconnu comme un /y/, /œ/ ou /ø/. Un autre phénomène est celui concernant le /ɔ/ qui en continuum est moins confondu avec le /o/ mais plus avec un /a/ pour le groupe 3 et /ɑ/ pour le groupe 2. Le /o/ a tendance à avoir été plus perçu en continuum comme son homologue mi-ouvert pour les groupes 2 et 3 et un /u/ pour le groupe 1. Il est clair, d'après cette analyse, que la production en continuum n'a pas été une aide articulatoire ou cognitive dans la réalisation canonique des phonèmes vocaliques du français standard et que par conséquent, cette réalisation pourrait induire des confusions supplémentaires que celles qui pourraient apparaître à l'isolée. Les groupes 2 et 3 montrent néanmoins un avantage dans la réalisation du /u/ dont la confusion avec /y/ est importante lorsque l'on demande aux apprenants de produire le continuum /u-o-ɔ-ɑ/ avec pour seule aide le trapèze vocalique. Cet avantage porterait sur l'écoute d'un input auditif qui limiterait le biais



de la représentation graphique. Nous pouvons remarquer, sans pouvoir toutefois généraliser, que l'on observe une stagnation de la confusion /u-y/ pour le groupe 3 avec un input où les phonèmes sont associés aux notes de piano. Le /i/ ne semble pas être affecté par l'effet du groupe et de la modalité et cela pourrait être expliqué par le fait que ce phonème est présent dans beaucoup de langues (393 langues sur 451, UPSID ; Maddieson, 1984) et donc peu sujet à la nativisation/assimilation phonologique.

Nous terminerons par la comparaison des résultats au test de perception entre les enregistrements 1 et 2, soit 8 heures de travail en groupe. Nous rappelons que cette analyse a été réalisée dans un sous-panel de 6 apprenants (2 apprenants x 3 groupes) pour des raisons de qualité d'enregistrement lors du post-test +1. Ces données ne sont donc pas généralisables. Même s'il n'est pas possible de montrer statistiquement une progression entre les deux tests, nous avons pu constater qu'en moyenne, les phonèmes ont été mieux identifiés lors du deuxième enregistrement à l'isolée (+7,1%) et en continuum (+7,5%) pour la grille « extrême », et respectivement +5% et 4% pour la grille « archiphonème ». Ceci est d'autant plus intéressant que les apprenants qui montrent une tendance de progression (+10 identifications correctes en moyenne) est d'autant plus forte que la tendance de ceux qui affichent une régression (-2 identifications correctes en moyenne). Il est donc clair qu'un développement langagier phonologique est déjà visible au bout de 8 heures de travail en groupe même si ce n'est qu'une légère tendance et que celle-ci n'est pas uniforme parmi les six apprenants. De manière plus spécifique, nous avons pu montrer une progression importante entre les deux tests dans l'identification du /u/ (25% en moyenne pour les deux modalités), pour le /E/ à l'isolée (+11,7%) pour le /A/ (+4,2%) et le /i/ (+3,3%) pour la grille « archiphonème ». Grâce à la grille « extrême », nous avons pu préciser ces tendances en montrant une progression positive pour l'ensemble des phonèmes et quelle que soit la modalité de production avec une progression comprise en +30% et +1,7%. Nous remarquons, cependant deux régressions à l'isolée de -6,7% pour le /a/ (plus confondu avec le /a/ et le /ɔ/) et le /e/ (plus confondu avec le /ɛ/).

Il est clair que la grille choisie « extrême » ou « archiphonème » a une influence sur le score d'identification obtenu par les apprenants et que par conséquent, c'est un élément à prendre en compte sur le plan de l'évaluation sommative et certificative mais aussi formative. Les résultats globaux sont en accord avec les principes de la phonologie comparée et les phonèmes les plus présents dans les langues (ex : /i/ ou /a/) ont été identifiés de manière importante. Nous avons pu constater que /u/ pouvait poser un problème lorsqu'il est associé à sa forme API qui crée des confusions avec sa forme alphabétique et graphique. Concernant la confusion /u-y/, il

semblerait que l'input auditif du groupe 3 soit celui qui ait limité le plus cette dernière, même s'il est difficile de généraliser cette analyse. Cette tendance est par ailleurs la seule qui ressorte de notre analyse car il ne semblerait pas que notre proposition didactique montre un apport en termes de potentialité développementale. Toutefois et pour conclure, la comparaison des résultats au test d'identification pour les enregistrements 1 et 2 montrent une progression certes légère mais consistante sur la majorité des phonèmes. De plus, les progressions sont plus importantes que les régressions. Nous noterons la difficulté de comparer ces résultats avec la littérature existante puisqu'en général les analyses empiriques de la prononciation du français en contexte FLES portent sur une des groupes avec des L1 homogènes (ex : des apprenants japonophones ; Kamiyama, 2009) même si elles ne représentent qu'une partie du contexte d'enseignement du français. En phonétique-phonologie, les études portent généralement sur des contrastes consonantiques (/r/-l/ : Strange & Dittmann, 1984 ; Protopapas & Calhoun, 2000 ; /t/-d/ : Flege, 1995b ; /t/-d/ et /p/-b/ : Aliaga-Garcia & Mora, 2008 ; entre autres). Nous nous sommes focalisés pour cette étude sur les résultats d'un test d'identification passé par dix auditeurs natifs du français en raison de la qualité notamment des enregistrements. Cependant, il serait intéressant dans une prochaine étude d'obtenir des enregistrements permettant de comparer les résultats d'une analyse acoustique avec ceux d'un test d'identification par un panel d'auditeur natifs du français.

### 6.4 Conclusion générale de l'étude 3

A travers cette étude, nous souhaitons questionner le caractère potentiellement développemental de notre dispositif d'apprentissage, dans une perspective de didactique interventionniste mais également quasi-expérimentale, en étudiant l'impact d'une association entre continua vocaliques et notes de musique. Cette association reposait sur une association arbitraire mais néanmoins motivée par une réflexion théorique et pédagogique et notamment sur la notion d'abstraction conceptuelle. Nous souhaitons voir si cette association pouvait être une aide pour les apprenants dans l'appropriation de la logique articulatoire et perceptive des continua vocaliques du français et notamment ceux « antérieur-étiré » et « postérieur-arrondi ». Nous avons principalement traité ce sujet à partir de représentations mesurées par trois questionnaires à trois moments-clé (pré-test, post-test t+1 et t+2) et par la passation d'un test d'identification par dix locuteurs francophones natifs sur les phonèmes vocaliques produits par les apprenants lors des deux enregistrements et en prenant en compte deux modalités de production : à l'isolé ou en continuum. Afin de prendre en compte une approche réductionniste et sociale de la prononciation, nous avons décidé d'analyser les confusions sous le prisme de

deux grilles : une extrême qui prend en compte le système vocalique extrême du français et une « archiphonème » qui ne retient que les oppositions phonologiques pertinentes en français contemporain. Nous avons pu voir, tous groupes confondus, que les auto-évaluations des apprenants révèlent une progression significative entre les questionnaires 1 et 3 à l'exception du niveau d'oral général. Nous avons également vu que les apprenants étaient plus sensibles à l'utilité d'une approche en groupe et basée sur la pratique plutôt que sur une approche raisonnée théorique. Cependant, un grand nombre ont affirmé que les symboles phonétiques les avaient aidé à mieux percevoir et produire les voyelles du français et ce grâce au fait qu'ils se sentent plus à l'aise avec ces derniers au bout de sept semaines. Il semblerait que cet aspect ne soit pas confirmé par les matrices de confusion et surtout pour le /u/ dont la forme API et graphique ont pu ajouter une difficulté à de possibles confusions phonologiques. Le sous-groupe de six apprenants n'a pas permis de mesurer une progression significative même si nous avons pu constater qu'en moyenne, les phonèmes vocaliques avaient été mieux identifiés lors de l'enregistrement 2 par rapport au premier.

Au regard de notre proposition pédagogique, il n'a pas été possible de montrer de forte tendance en faveur d'une potentialité plus développementale pour le groupe 3 que la simple exposition à un trapèze vocalique dans le cas du groupe 1, par exemple. Cependant, nous avons vu qu'au plan représentationnel, l'utilisation du piano avait été clivante puisque soit les apprenants montrent, dans le questionnaire, une forte adhésion, soit un fort rejet. Cela se reflète dans les productions des apprenants où en moyenne 4 apprenants sur 9 ont produit la forme prosodique attendue. Cependant, il semblerait que la présentation de l'input avec notes de musique ait limité l'augmentation de confusion /u-y/ par rapport à l'isolée et par rapport aux groupes 1 et 2. Les notes de musique auraient pu jouer un rôle dans une attention plus forte au son qu'à la forme graphique proposée dans le trapèze vocalique. Cependant, il semblerait que la présentation d'un input sonore ait favorisé des confusions /ɔ-a/ ou /ɔ-ɑ/ pour les groupes 2 et 3 et contrairement au groupe 1. Il semblerait donc que l'utilisation des notes de piano dans l'acquisition de nouvelles formes phonologiques puisse avoir un impact cognitif chez certains apprenants mais il semblerait également que la production de phonèmes en continuum soit une source d'erreur comparée à l'isolé.

**Synthèse de positionnement :**

Cette étude didactique s'est voulue résolument pluri-inter-transdisciplinaire puisqu'elle tente de faire coïncider les buts développementaux et les pressions d'un dispositif d'apprentissage en didactique des langues avec les outils expérimentaux et d'analyse de la phonétique-phonologie. Nous avons voulu également y ajouter une volonté créative en proposant un outil pédagogique nouveau en lien à la musicologie. Intégrer l'ensemble de ces composantes dans une même étude, au nom de la complexité, a imposé de faire des concessions fortes ce qui a pu me pousser dans une approche plus réductionniste que je ne l'imaginais. Il est clair qu'une telle étude mériterait d'être approfondie dans une recherche ultérieure et notamment afin de montrer si l'acquisition des phonèmes à l'isolée (ou en continuum) implique (in)directement des ruptures dans la production en discours et, plus particulièrement, face aux phénomènes de coarticulation. Il faudra également penser à l'intégration de ce type d'outil dans une tâche sociale plus large permettant de stimuler l'intérêt, l'envie d'apprendre et d'intégrer la prononciation dans un processus plus complexe d'interaction. Cependant, je pense que cette étude a permis de montrer la possibilité de questionner une approche créative à l'aide d'outils scientifiques et expérimentaux et donc de faire coïncider à la fois création et déconstruction. Elle a permis notamment de mettre en évidence des points de rupture (progression dans les représentations et dans les productions) mais également des points de résistance, notamment sur les facteurs psychoaffectifs liés à la production de parole chantée.

Cette étude m'a permis également, en tant qu'individu-chercheur, de me rendre compte de la difficulté de mettre en place une recherche à la fois interventionniste et quasi-expérimentale pour faire émerger des tendances dans un dispositif à durée limitée et avec des publics hétérogènes. Il me semble que la notion de potentialité développementale permette de solutionner ce problème et de limiter le nombre de groupes-témoin. Multiplier le nombre de groupe semble mettre en avant les limites des approches plutôt que leur(s) potentialité(s) développementale(s) qui, de manière comparée, montre(nt) peu de fortes différences en raison de l'évolution naturelle des organismes vivants quelle que soit l'approche choisie (cf. théorie du chaos ; Larsen-Freeman, 1997).

## **DISCUSSION GENERALE**

Dans cette section, nous mettrons tout d'abord en corrélation les points importants qui ressortent des trois études que nous avons présentées et analysées. Cela nous permettra de les situer par rapport aux questionnements initiaux. Dans un deuxième temps, nous présenterons les limites du travail ainsi que les perspectives et implications didactiques.

### **Bilan global**

Cette thèse avait pour but de questionner les liens entre musique et parole dans une perspective de didactique de la prononciation en contexte FLE/S. Pour ce faire, nous avons mené une réflexion à la fois théorique et empirique qui nous a permis de venir confirmer ou remettre en question certaines de nos hypothèses. Plus qu'un recueil et une analyse de données, ce travail doctoral nous a plongé dans des problématiques disciplinaires anciennes et complexes. La linguistique, la linguistique appliquée et la didactique des langues (ou du moins les chercheurs de ces domaines) n'ont pas encore totalement réussi à se situer en France. Ces obstacles ont été à l'origine d'un besoin de définir les caractéristiques de la pluri-inter-transdisciplinarité qui impose de respecter les exigences de chaque domaine tout en en créant un nouveau en même temps. Nous avons dû également prendre le temps de nous former à un certain nombre d'outils pour lesquels la didactique des langues n'avait pas apporté de réponse dans notre cursus universitaire pré-doctoral. En effet, la didactique ne possède pas ou peu d'outils propres à l'analyse des données et il importe donc que le jeune chercheur de trouver les moyens de se former aux outils nécessaires au traitement de ces données. Nous avons tenté également, à travers cette réflexion, de déconstruire un certain nombre de représentations. Cette déconstruction a nécessité de déterminer un éclairage intégré des théories émergentiste et socioconstructiviste. Elle nous a aussi permis de définir les processus langagiers comme l'émergence de propriétés de l'objet complexe qui est plus que la simple somme de ces parties. Nous avons ainsi vu de ce fait que les liens entre musique et parole sont complexes même si un certain nombre d'éléments convergent vers un « ancêtre » commun (un protolangage) caractérisé par des propriétés holistique, multimodale, intentionnelle et musicale dont les traces pourraient être mesurées aujourd'hui dans l'hypothèse de ressources d'intégration syntaxique partagées même si les représentations à long terme, quant à elles, seraient gérées de manière distinctes. Toutefois, il est clair, d'après un certain nombre de travaux en neurocognition, que l'entraînement musical a des effets sur les capacités auditives d'autres objets et notamment en ce qui nous concerne la parole. C'est à ce titre que nous avons pu recenser un certain nombre

de pratiques pédagogiques qui lient les objets verbaux et musicaux, même si aucune d'entre elles n'a été testée de manière empirique. Notre travail doctoral a cherché à contribuer à la mise en place de fondations dans cette thématique. Dans notre introduction, nous mentionnions que cette thèse avait pour but d'amorcer une réflexion sur les niveaux auxquels les différences entre musique et parole se situaient. Le niveau évolutif a été traité dans le cadre théorique. Les autres niveaux ont été abordés à travers chacune des trois études qui visaient à répondre à trois questionnements :

- 1) Le rapport que les individus portent sur l'objet verbal ou musical modifie-t-il la manière dont ils les conçoivent ? Que peut-on apprendre de l'un pour enrichir l'autre ?
- 2) Existe-t-il des différences entre musiciens et non-musiciens chez des apprenants du français en contexte FLE/S ? Si oui, à quel niveau et peuvent-elles être mesurées empiriquement ?
- 3) Une approche pédagogique qui lie musique et parole peut-elle avoir un apport potentiellement développemental ? Si oui, est-il supérieur à d'autres approches ?

Cependant, plutôt que de répondre de manière linéaire à ces questionnements, nous les regrouperont de manière thématique afin de mettre en avant les apports de cette thèse au(x) domaine(s). Deux thématiques générales sont envisagées :

### - **L'impact de la pratique musicale**

Etude 1 : Les musiciens semblent être fortement sensibles à la différence entre ce qu'ils produisent et ce que les autres perçoivent. Les locuteurs plurilingues veulent principalement valider l'interaction et la pratique régulière d'au moins une langue étrangère semblent limiter cette perception.

Etude 2 : Les musiciens répondent, en moyenne, plus rapidement que les non-musiciens à un test de discrimination AXB qui focalise sur le traitement acoustique. Cependant, il n'y a pas de différence significative en ce qui concerne le taux de discrimination.

Les études 1 et 2 nous ont permis d'apporter de nouvelles données en ce qui concerne les particularités des musiciens. Les musiciens qui ont fait du solfège montrent une forte tendance en faveur du fait qu'ils entendent la différence entre ce qu'ils jouent et ce qu'entendent les spectateurs (65% sont tout à fait (30%) ou plutôt d'accord (35%)). Cette tendance est plus faible pour les locuteurs plurilingues qui pratiquent au moins une langue dans un contexte naturel en ce qui concerne leur sentiment lorsqu'ils parlent (36% sont tout à fait (9%) ou plutôt d'accord

(27%)). Nous avons également pu montrer de manière statistique que les musiciens répondent en moyenne plus rapidement que les non-musiciens de 148,12 ms lorsqu'ils ont une tâche qui met en avant le traitement acoustique des phonèmes vocaliques du français. Cependant, cette différence n'est pas significative au regard du taux de discrimination. L'ensemble de ces points nous permet de mieux appréhender l'impact de l'entraînement musical sur le traitement des unités de la parole. Les musiciens et notamment ceux ayant fait du solfège ont une meilleure sensibilité entre ce qu'ils produisent et ce qui est perçu. Cet entraînement aurait des effets sur le plan acoustique en leur permettant de traiter plus rapidement les informations sonores et donc de prendre une décision de manière plus rapide. Ces données confirment donc le fait que l'entraînement musical aurait des effets sur d'autres objets que seulement le musical. Dans notre cas, cela fait référence aux unités vocaliques de la parole. Cela confirme donc les récents travaux en neurocognition. Cependant, nous devons prendre en considération le fait que ces différences entre musiciens et non-musiciens n'ont été mesurées que le plan acoustique et non au niveau phonologique ce qui serait en accord avec le fait que la musique et la parole possèdent des corrélats partagés de bas niveau mais que leur stockage en représentations à long terme est dissocié. De plus, il semblerait que plus un individu pratique une langue plus il favorise la validation du sens par l'interaction moins il prend conscience de la matérialité de la parole en son.

### - **La musique instrumentale et la didactique des langues**

Etude 2 : Une exposition limitée à de la musique instrumentale n'a pas eu d'effet significatif sur la répétition de la mesure du taux de discrimination et du temps de réaction entre les tests 1 et 2.

Etude 3 : Associer les notes de musique et l'aperture n'a pas d'apport potentiellement développemental particulier face à d'autres approches. Cette association montre une adhésion bimodale.

Les études 2 et 3 nous ont permises de porter un regard critique sur les utilisations possibles de la musique instrumentale dans la didactique des langues. Nous avons vu qu'une exposition en temps limité (3 minutes) n'as pas d'effet significatif sur le traitement d'unités vocaliques pour des apprenants du français et ce en comparaison à l'exposition à un document visuel ou de parole en français. De la même manière, l'association proposée entre notes de musique et aperture n'a pas montré d'apport potentiellement développemental supérieur à d'autres modes de présentation (F0 plat ou non-contrôlée). Au regard de ces deux études, il nous semble clair

que les effets de la musique ne peuvent avoir d'impact à court terme. De plus, l'entraînement musical, pour qu'il puisse être mesuré, nécessite une pratique quantitativement et qualitativement importante afin de développer ce que l'on appelle une expertise (Bigand, 2004) dans le respect du temps que nécessite la plasticité cérébrale. De plus, une approche qui associe musique instrumentale et parole ne semble pas avoir d'effet négatif sur le développement langagier même si linguistiquement parlant, cela n'a pas d'effet positif non plus en comparaison à d'autres approches. Il semblerait donc qu'une telle approche puisse permettre à certains apprenants de prendre conscience des éléments phonémiques sur lesquels on mettrait l'accent mais que cela n'est pas supérieur à d'autres approches et que cela fait émerger une adhésion bimodale entre ceux qui trouvent cela très utile ou pas du tout. Une approche sensible peut faire émerger des résistances psycho-émotionnelles chez certains apprenants. L'utilisation d'éléments à fort potentiel émotionnel nécessite de prendre en considération et d'anticiper des obstacles d'ordre psychoaffectif supplémentaires. L'importance de la prise en compte de ces obstacles réside dans le fait qu'ils peuvent jouer sur l'adhésion des apprenants à une approche pédagogique et par conséquent jouer négativement sur leur progression.

#### - **Autres résultats intéressants**

Etude 2 : Il n'y a pas de tendances significatives concernant le taux de discrimination et le temps de réponse des phonèmes vocaliques entre apprenants et locuteurs natifs du français.

Etude 3 : La production en continuum augmente le nombre de confusion des voyelles par rapport à une production à l'isolée.

Ces deux points ne touchent pas directement les questionnements de ce travail de thèse mais semblent intéressants dans une perspective plus large de la didactique de la prononciation. Nous avons montré, en outre, qu'il n'y avait pas de différence significative entre apprenants et locuteurs natifs du français que ce soit pour le taux de discrimination ou le temps de réaction à un test de discrimination visant le traitement acoustique des phonèmes vocaliques. Cela renforce l'idée qu'il existe un double traitement de la parole : d'une part acoustique et d'autre part phonologique. Sur le plan didactique, cela remet en cause la notion de « surdité » (symbolisme) au profit de la nativisation/assimilation phonologique (cognitivisme). Ce processus peut, dès lors, être vu comme un traitement descendant (des automatismes cognitifs langagiers) qui vient perturber le traitement ascendant (analyse acoustique de l'*input* sonore). Ce double processus provoque des interférences entre le signal acoustique réel et ce qui est perçu dans sa dimension psychocognitive. L'objet central de la médiation de la prononciation



se situe donc, dans notre approche, dans le développement de nouveaux automatismes phonologiques permettant à la fois de limiter le biais phonologique de la/les L1(s) tout en favorisant un traitement écologique des unités acoustiques de la « langue » visée. Nous rappelons que le traitement phonologique tend à faciliter le décodage des caractéristiques acoustiques propres à un code langagier.

Le deuxième point a remis en question une représentation forte chez nous qui consistait à penser que la production des phonèmes en continuum facilitait cette dernière à travers la logique articulatoire de chaque continuum antérieur arrondi/étiré et postérieur arrondi. Cependant, il semblerait d'après notre troisième étude (didactique interventionniste et quasi-expérimentale) que les phonèmes produits par les apprenants montrent plus de confusion, dans un test d'identification passé par des auditeurs natifs du français, lorsqu'ils sont produit en continuum qu'à l'isolée. Cette tendance se reproduit que les apprenants aient un input auditif ou non. Il semblerait donc, d'après ces résultats, que le caractère facilitant de la logique articulatoire ne soit pas suffisant pour contrebalancer le poids cognitif demandé par l'articulation de plusieurs phonèmes les uns à la suite des autres. Cela renforce également l'idée que la performance est conditionnée par le type de tâche imposée. Il est donc important, en didactique des langues, de prendre en considération le fait que la logique articulatoire ou perceptive, entre les unités sonores langagières, ne se présentera pas forcément comme facilitatrice dans la production elle-même. Cela montre également que chaque modalité de production (isolée, continuum, logatome, mot, énoncé) requiert des capacités liées mais dont les liens ne sont pas directs et exclusifs.

### **Limites et implications didactiques**

A partir des perspectives globales mises en évidence par notre cadre théorique et nos trois études, nous allons nous pencher sur les limites de cette approche ainsi que sur les implications didactiques possibles. Nous avons tout d'abord compris la difficulté de mettre en place une recherche en didactique des langues à l'aide d'outils de la phonétique-phonologie qui imposent généralement un contrôle des données. Cela nous a amené vers une approche plus réductionniste que nous ne le pensions étant donné que les liens entre musique et parole semblent être plus forts sur les processus de bas niveau et étant donné que la méthodologie empirique en phonétique-phonologie favorise, autant que faire se peut, des données contrôlables en laboratoire. Nous avons dû à ce titre modifier les objectifs de l'étude 3 en raison de données audio saturées liées à des problèmes techniques et l'impossibilité de refaire des enregistrements à cause du grand nombre d'apprenants du dispositif d'apprentissage ainsi que

l'obligation de faire ces enregistrements pendant le temps du cours. Avoir une approche quasi-expérimentale en didactique des langues est extrêmement chronophage notamment avec des grands groupes (35 étudiants). Nous avons aussi pu percevoir que la mise en place de cette étude interventionniste a pu perturber un certain nombre d'apprenants et certainement que notre volonté d'intégrer le dispositif expérimental au sein du dispositif d'apprentissage n'a pas été totalement réussi. En outre, nous avons pu voir que la production en continuum pouvait demander une charge cognitive importante ce qui a pu augmenter les erreurs d'identification par les auditeurs natifs du français. Les limites de notre première étude portent sur le fait que nous avons voulu comparer les rapports du musicien sur la musique et des locuteurs plurilingues sur la parole uniquement. Cela nous a permis de limiter les effets du fait que des locuteurs plurilingues soient des musiciens mais cela ne nous a également pas permis confronter le rapport des musiciens à la musique et à la parole.

Malgré les limites mentionnées, de nombreuses implications didactiques et perspectives de recherche émergent de ce travail doctoral. De manière générale, il semblerait que des associations entre unités phonologiques et musicales pourraient faire partie des outils pédagogiques existants pour aider certains apprenants mais ce ne serait qu'un des outils à disposition des apprenants/enseignants. Nous pourrions intégrer cela dans une approche sensible de la didactique de la musique afin de permettre d'amener les apprenants à prendre conscience de leur prononciation dans une langue étrangère. En effet, la didactique de la musique pourraient inspirer la didactique de la prononciation pour une prise de conscience plus importante entre ce qui est produit et ce qui est perçu afin de sensibiliser les apprenants à leur prononciation. Les applications concrètes que nous pouvons envisager à travers nos résultats ne sont ni novatrices, ni innovantes (comme c'est à la mode d'en qualifier toute proposition pédagogique), mais viennent confirmer voire conforter des approches déjà existantes telles que la suggestopédie, la méthode verbo-tonale, etc. (voir Lauret, 2007 pour une synthèse de ces approches). Cependant, nous avons pu montrer que plus les locuteurs plurilingues pratiquent une langue, moins ils sont sensibles aux caractéristiques des sons qu'ils produisent. Cela irait à l'encontre d'une approche sociale de la didactique de la prononciation et irait en faveur d'une déconstruction des unités de la parole afin de faciliter la dénativisation/accommodation. Cette déconstruction pourrait s'inspirer de la didactique de la musique en proposant des tâches d'entraînement visant à aller de l'unité minimale (segmental) à l'unité maximale (discursif). Un travail proprioceptif serait également pertinent afin de se réapproprier des mécanismes vocaux internalisés depuis la très jeune enfance dans la/les langue(s) première(s).

Sur le plan de la recherche, ce travail doctoral ouvre plus de questions qu'il n'a apporté de

réponses, même si celles-ci ont été nombreuses. Ces questions ne relèvent, par ailleurs, pas forcément de notre thématique principale, mais elles incitent à développer les connaissances sur les processus de perception et de production des sons des langues étrangères. Nous avons notamment une volonté forte de créer du lien entre les chercheurs de nos deux disciplines de référence: la phonétique-phonologie et la didactique des langues. Les perspectives expérimentales, que nous envisageons, sont les suivantes :

- Etude 1 et 3: ces deux études nous amènent à vouloir tester une approche sensible tirée de la didactique de la musique sur la perception de sa propre prononciation. Nous aimerions reproduire le dispositif d'apprentissage (distanciation, décontextualisation puis contextualisation des sons de la parole) de l'étude 3 tout en mettant au centre du dispositif une prise de conscience entre ce qui est produit et ce qui est perçu. Cela pourrait se faire à travers un enregistrement systématique de l'apprenant et une série d'auto-confrontation individuelle ou en groupe. Une tâche orale esthétique finale pourrait être envisagée afin de valoriser l'apprenant dans son interprétation du français. Ce pourrait être des lectures de textes (poèmes, écrits personnels, etc.) en musique devant un public interne ou externe au dispositif ; des vidéos en ligne représentant la diversité des individus francophones.
- Etude 2 et 3: ces deux études nous conduisent à nous questionner sur l'impact de ruptures du système phonologique à l'isolée dans l'émergence (probable) de ruptures dans le discours. Nous avons focalisé notre attention pour ce travail sur les phonèmes vocaliques du français. Il serait intéressant de mesurer si un travail spécifique au niveau segmental (des voyelles, par exemple) provoquerait systématiquement une réorganisation du système phonologique au niveau du discours ou s'il y aurait une asymétrie. Cela permettrait de mieux comprendre les mécanismes qui sous-tendent le développement de nouvelles catégories phonologiques.
- Etude 2: en complément au point précédent, nous aimerions corrélérer le taux de discrimination AXB avec ISI long (1 seconde) et ISI court (200 ms) chez des apprenants en contexte FLE/S. Cela nous permettrait de mettre en évidence différents profils entre traitements acoustique et phonologique. Nous pourrions également mesurer l'effet de variables interindividuelles comme le niveau évalué de langue ou le statut musical. Le but serait de mieux comprendre et appréhender les processus qui sous-tendent le traitement de la parole à plusieurs niveaux et ainsi contribuer aux connaissances en phonétique-phonologie.

- Étude 2: sur le plan expérimental, nous souhaiterions reproduire la même expérience de l'étude 2, en psychoacoustique, en allongeant la durée de l'input musical et en supprimant le questionnaire afin limiter les interférences probables qui ont pu jouer en faveur d'un effet non significatif. De plus, nous pourrions venir confirmer ou infirmer les résultats obtenus lors de notre première étude et ainsi mesurer la robustesse des effets et interactions significatifs.

A travers cette discussion générale, nous avons souhaité montrer que les trois études expérimentales, au-delà de leur situation sur des terrains expérimentaux différents, avaient pour but de soutenir une même réflexion quant aux liens entre musique et parole dans une perspective de didactique de la prononciation en contexte FLE/S. Nous avons également mis en évidence le fait que peu de travaux dans cette thématique précèdent le nôtre ce qui en fait un travail pionnier et que par conséquent, les perspectives sont nombreuses tant sur le plan de la recherche fondamentale qu'appliquée.

## CONCLUSION GENERALE

Ce travail de thèse a dépassé le cadre de la réflexion doctorale académique : il a été le catalyseur d'un positionnement en tant que jeune chercheur. Il a permis de questionner les liens entre musique et parole dans le but de compléter les outils de la didactique de la prononciation en contexte FLE/S. Etant que chercheur, un tel cadre pluriel nécessitait de se questionner sur la notion de discipline elle-même. Cette nécessité apparaît au regard de la tendance naturelle de l'humain à catégoriser le monde qui l'entoure afin de le comprendre et l'appréhender (Claverie, 2010). Nous notons également, sur le plan épistémologique, un besoin actuel dans la recherche de donner du lien à des travaux issus du structuralisme qui ont permis de mieux comprendre les processus langagiers à travers leur déconstruction mais qui ont fragmenté ces processus dans leur complexité. Les travaux inter- et transdisciplinaires contemporains répondent partiellement à ce besoin. En parallèle, nous avons tenté de définir un nouveau cadre théorique émergentiste et socioconstructiviste qui permette de prendre en considération à la fois les processus ascendants et descendants constitutifs du langage. Même si cet éclairage nouveau nécessitera du temps et la collaboration de chercheurs de différents domaines pour trouver sa place dans le paysage des sciences du langage, nous voyons dès à présent qu'il constitue un tournant dans la prise en compte du langage dans sa dimension à la fois neurocognitive mais également psychoaffective et sociale. Cette approche permet de réintégrer le rôle des émotions dans les processus cognitifs et sociaux puisque l'humain semble être prédisposé à chercher la proximité et l'attention d'autrui (cf. théorie de l'attachement : Ainsworth & Bowlby, 1991) et dont on a pu retrouver des corrélats neurologiques (cf. rôle de l'amygdale et du cortex septal ; Joseph, 1990 : 123). Même si la question des émotions n'a pas directement été traitée dans ce travail de thèse, nous savons qu'elle fera partie intégrante de nos futurs travaux puisque, dans notre perspective, les émotions sont à la fois le résultat de l'interaction avec le monde qui nous entoure tout en permettant également de déclencher ou d'inhiber des processus cognitifs. Chez l'adulte, il nous semble déterminant de prendre en compte les facteurs psychoaffectifs dans l'activation qualitative et quantitative des codes de la/les L1(s) sur le plan phonologique en ce qui nous concerne. Nous pourrions ainsi mesurer l'impact de la peur, par exemple, dans l'émergence d'un accent chez des apprenants du français. Dans cette perspective, la musique peut avoir un double emploi : contribuer à induire des états relaxant face à des émotions comme la peur qui peuvent se traduire en anxiété, angoisse, etc. mais également à apprendre à gérer ces états qui inhibent la production langagière.

Le sujet des émotions nous ramène également à la place du chercheur dans la recherche, sa recherche et la production de données scientifiques. Ce travail de thèse, qui est né initialement d'une curiosité plus que de besoins sociaux contemporains, répond néanmoins à des problématiques actuelles récurrentes. L'influence du socio-interactionniste, en didactique des langues, a amené les didacticiens à minimiser le rôle de la prononciation dans le rapport à l'autre en dehors de la composante purement linguistique. La parole ne donne pas accès seulement au sens mais également à l'état émotionnel de l'individu. J'ai pu me rendre compte, d'après mon expérience d'enseignement et de discussion avec les apprenants, que la volonté de « perdre ou limiter son accent » est souvent motivée par la peur d'être face à un interlocuteur qui ne vous comprend pas ou qui vous répondra dans une autre langue que celle que vous souhaitez pratiquer. La volonté de contrôle ou le perfectionnisme peuvent également être l'origine de ce désir. Le rôle du médiateur est alors autant d'amener l'individu à déconstruire ses peurs qu'à lui apporter des éléments (linguistiques) qui peuvent lui permettre d'éviter ces situations pouvant lui faire perdre la face. A ce titre, il est important pour tout pédagogue de se situer clairement dans son rapport à sa propre prononciation. Chaque pédagogue peut être amené à gérer l'émergence de pratiques pédagogiques conditionnées par ses propres expériences passées (familiales, éducatives, sociales, etc.). Sans ce recul, chaque pédagogue pourrait, de manière inconsciente, imposer sa vision de ce qu'est une « bonne » prononciation sans prendre en compte les buts, besoins et sensibilités de chaque apprenant.

Traiter la notion d'émotion pour un chercheur n'est pas aussi aisé que nous pourrions le croire. Ce travail de thèse a finalement pris une direction plus réductionniste que nous ne le souhaitions face à notre cadre théorique. Cette tendance a été influencée d'une part par les outils de la phonétique-phonologie mais également la volonté primaire de tout chercheur d'être objectif (ou se persuader de l'être). Je crois que la déconstruction de cet idéal est un processus qui ne sera jamais stabilisé. L'homme est fait pour la stabilité dans un monde où tout est instable (Claverie, 2010 : 23). Utiliser les outils réductionnistes de la phonétique-phonologie était à la fois une manière de (me) prouver que je n'étais pas que didacticien mais à la fois de mettre de la distance avec des facteurs psychoaffectifs personnels trop déstabilisants. Cependant, au terme de ce travail doctoral, j'ai pu prendre conscience de ces mécanismes afin de mesurer l'impact de ces obstacles dans la préhension des mécanismes complexes du vivant. Nous sommes des humains qui étudions des humains qui sont le miroir de ce que nous sommes. Se placer en transdisciplinarité, c'est se mettre dans un espace instable tant pour soi que pour les autres (chercheurs, institution) et qui requiert de trouver des certitudes dans un paradigme qui relève de l'épistémè de l'incertitude (Morin, 2000).

Ce travail a également été l'occasion de venir consolider les connaissances sur les facteurs environnementaux qui peuvent influencer la perception et la production des sons de la parole. Nous avons principalement mis en avant deux tendances intéressantes chez les musiciens qui ont une meilleure perception entre ce qui est produit et ce qui est perçu au niveau représentationnel et qui ont de meilleurs temps de réaction par rapport aux non-musiciens à un test de discrimination AXB. Nous avons pu également déterminer qu'une approche associant notes de musique et phonèmes vocaliques possède un atout potentiellement développemental même si celui-ci n'est pas supérieur à d'autres approches et qu'il n'est, par conséquent, qu'un outil supplémentaire de la mallette de l'enseignant au service de l'apprenant. Ce dernier point nous rappelle que les méthodes pédagogiques alternatives ne sont pas un substitut aux méthodes plus traditionnelles. Il nous faut réfléchir, en tant que didacticiens, aux apports potentiellement développementaux de chaque méthode et comment les articuler entre elles afin de les rendre complémentaires. Cette complémentarité permettra de prendre en considération les profils pluriels des apprenants et leur offrir toutes les possibilités de progresser. La capacité de l'enseignant à croire en ce qui fait et sa capacité de prendre du recul sur ses pratiques sont, d'après nous, déterminantes dans le succès d'une approche pédagogique.

Pour terminer, nous noterons qu'étant donné le peu de recherches dans cette thématique pour le moment dans le domaine, l'intérêt important de la neurocognition pour les liens musique-parole et les implications possibles de ces liens au niveau social, de nombreuses ouvertures en didactique des langues s'offrent à nous. Ce travail doctoral a ouvert plus de possibilités qu'il n'a permis de faire le tour de notre problématique.

## BIBLIOGRAPHIE

- Abgrall, J. M. (2001). *Healing or Stealing? Medical charlatans in the New Age*. New York: Algora Publishing.
- Abry, D. & Chalaron, M.-L. (2011). *Les 500 exercices de phonétique B1/B2*. Paris : Hachette Français Langue Etrangère.
- Aiello, L.C. & Dunbar, R.I.M. (1993). "Neocortex size, group size, and the evolution of language". *Current Anthropology*. 34. pp.184-93.
- Aiello, L.C. & Wheeler, P. (1995). "The expensive tissue hypothesis". *Current Anthropology*. 36. pp. 199-221.
- Aiello, L.C. (1996). "Terrestriality, bipedalism and the origin of language". In Runcinman, W. G., Maynard-Smith, J. & Dunbar, R.I.M. Eds. *Evolution of social behaviour Patterns in Primates and Man*. Oxford: Oxford University Press. pp. 269-90.
- Ainsworth, M. D. S. & Bowlby, J. (1991). "An ethological approach to personality development". *American Psychologist*. 46. 331-41.
- Ait Abdelmalek, A. (2004). « Edgar Morin, sociologue et théoricien de la complexité : des cultures nationales à la civilisation européenne ». *Sociétés*. N° 86. pp. 99-117.
- Alexander, J.A., Bradlow, A.R., Ashley, R.D. & Wong, P.C.M. (2008). "Music Melody Perception in Tone-language and Non-tone-language Speakers". *156th Meeting of the Acoustical Society of America*. pp. 10-14.
- Aliaga-Garcia, C. & Mora, J.C. (2008). « Perception and production of oral stops by Catalan/Spanish learners of English: A phonetic training experiment. In Monroy, R. & Sanchez, A. (Eds). *25 years of applied linguistics in Spain: Milestones and Challenges. Proceedings of the 25<sup>th</sup> AESLA Applied Linguistics Conference*. Murcia: Servicio de Publicaciones de la Universidad de Murcia. pp. 9-15.
- Andersen, R. (1983). *Pidginisation and Creolisation as Language Acquisition*. Rowley: Newbury House.
- Aubin, S. (1996). *La didactique de la musique du français : sa légitimité, son interdisciplinarité*. Villeneuve d'Ascq : ANRT. Thèse de doctorat soutenue à Rouen.
- Aubin, S. (2008). « Maître de langue, professeur de langue et enseignement de la musique du français ». *Synergies Espagne*. N° 1. pp. 14-57.
- Auroux, S. (2001). « Le langage n'est pas dans le cerveau ». In Dortier, J.-F. (coord.). *Le langage*. Auxerre : Éditions Sciences humaines.
- Baddeley, A. & Hitch, G.J. (1974). "Working Memory". In Bower, G. Eds. *Recent Advances in learning and motivation*. Volume 8. New York: Academic Press. pp. 47-90.
- Barth, B.-M. (2004). *L'apprentissage de l'abstraction*. Paris: Retz
- Bei, G. X. (2013). "Effects of immediate repetition in L2 speaking task: a focused study". *English Language Teaching*. Volume 6, N° 1. pp. 11-19.



- Bernardi, L., Porta, C. & Sleight, P. (2006). "Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence". *Heart*. 92. pp. 445-452.
- Berthoz, A. (2009). *La Simplexité*. Paris: Odile Jacob.
- Best, C., McRoberts, G., & Sithole, N. (1988). "The Phonological Basis of Perceptual Loss for Non-Native Contrasts: Maintenance of Discrimination Among Zulu Clicks by English-Speaking Adults and Infants". *J. Exp. Psychol. Hum. Percept. Perform.* 14. pp. 345–360.
- Best, C. (1993). "Emergence of Language-Specific Constraints in Perception of Non-Native Speech: A Window on Early Phonological Development". In de Boysson-Bardies, B., de Schonen, S., Jusczyk, P., MacNeilage, P. & Morton Klewer, J. Eds. *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. The Netherlands: Dordrecht.
- Best, C. (1995). "A Direct Realist View of Cross-Language Speech Perception". In Stange, W. Ed. *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Timonium, MD: York Press. pp. 171–204.
- Best, C. T., Hallé, P., Bohn, O.-S. & Faber, A. (2003). "Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages". *Actes du 15ème congrès de l'ICPhS*. pp. 2889-92.
- Bigand, E. (2004). « L'oreille musicale experte peut-elle se développer par l'écoute passive de la musique ? ». *Revue de Neuropsychologie*. Volume 14, N°1-2. pp. 191-221.
- Billières, M. (2005). « Les pratiques du verbo-tonal. Retour aux sources ». In Berré, M. (2005). *Linguistique de la parole et apprentissage des langues. Questions autour de la méthode verbo-tonale de P. Guberina*. Mons: Centre International de Phonétique Appliquée. pp. 67-87.
- Blanc, J.-M. (2005). *Traitement de la Prosodie par un Réseau Récurrent Temporel : Un Cadre Unifié pour l'identification automatique des langues, des attitudes prosodiques, et des catégories lexicales*. Thèse d'état de sciences cognitives soutenue à l'Université Lyon 2 Lumière. En ligne : [http://theses.univ-lyon2.fr/documents/lyon2/2005/blanc\\_jm/info](http://theses.univ-lyon2.fr/documents/lyon2/2005/blanc_jm/info) . (Consulté le 05/08/13).
- Bourguignon, A. (1997). « De la pluridisciplinarité à la Transdisciplinarité ». Centre International de Recherches et Études Transdisciplinaire. Congrès de Locarno, 30 avril - 2 mai. Annexes au document de synthèse. Paris : CIRET-UNESCO. En ligne : <http://ciret-transdisciplinarity.org/locarno/loca5c1.php>. Consulté le 03/08/13.
- Bransford, J.D., Brown, A.L., & Cocking, R.R. Eds. (2002). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press.
- Brief, J.-C. (1977). « Le constructivisme piagétien et les épistémologies traditionnelles ». *Philosophiques*. Volume 4, N°2. pp. 195-224.
- Broca, P. (1861). « Remarques sur le siège de la faculté du langage articulé; suivies d'une observation d'aphémie (perte de la parole) ». *Bulletins de la Société Anatomique*. Paris. 6. pp. 330–357, 398–407.

- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A. & Rizzolatti, G. (2004). "Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study". *J. Cogn. Neurosci.* Jan-Feb, 16(1). pp. 114-26.
- Bygate, M. (2001). "Effects of task repetition on the structure and control of oral language". In Bygate, M., Skehan, P. & Swain, M. (2001). Eds. *Researching pedagogic tasks second language learning and testing*. Longman.
- Caclin, A. (2004). *Interactions et indépendances entre dimensions du timbre des sons complexes : approche psychophysique et électrophysiologique chez l'Humain*. Thèse de doctorat soutenue à l'université Paris VI en neurosciences.
- Callamand, M. (1981). *Méthodologie de l'enseignement de la prononciation, Organisation de la matière phonique du français et correction phonétique*. Paris : CLE International.
- Callon, M. & Latour, C. (1991). *La science telle qu'elle se fait*. Paris : La découverte.
- Campos, R., Donin, N. & Keck, F. (2006). *Musique et sciences humaines. Rendez-vous manqués ?* Paris : Revue d'histoire des sciences humaines. 2006/1, N°14.
- Caplan, D. (1992). *Language: Structure, Processing, and Disorders*. Cambridge, MA: MIT Press.
- Carignan, C. (2013). *When nasal is more than nasal: The oral articulation of nasal vowels in two dialects of French*. Ph.D. thesis, University of Illinois at Champaign-Urbana.
- Carruthers, P. (2002). "The cognitive functions of language". *Brain and Behavioral Sciences*. 25. pp. 657-674.
- Chan, A.Y. W. (2013). « The discrimination of English vowels by Cantonese ESL learners in Hong Kong: a test of the perceptual assimilation model ». *Open journal of modern linguistics*. Volume 3, N°3. pp. 182-9.
- Charlemagne, P. (2010). *Psychoacoustique: mesure subjective des paramètres d'intensité, de fréquence et de temps*. Mémoire soutenu dans le cadre du Diplôme d'Etat d'Audioprothésiste à Nancy-Université, Université Henri Poincaré.
- Chistovich, L. A. & Lublinskaja V. (1979). "The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli". *Hear. Res.* 1, 185–195.10.1016/0378-5955(79)90012-1
- Chobert, J. & Besson, M. (2012). « Influence de l'apprentissage de la musique sur la perception des syllabes chez les enfants normolecteurs et dyslexiques ». In Bigand, E. et al. *Musique et Cerveau : nouveaux concepts, nouvelles applications*. Montpellier: Sauramps Medical. pp. 61-75.
- Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.
- Chouard, C.-H. (2001). *L'oreille musicienne. Les chemins de la musique de l'oreille au cerveau*. Paris: Gallimard.
- Clark, A. & Misyak, J.-B. (2009). "Language, Innateness, and Universals". In Christiansen, M.H., Collins, C. & Edelman, S. Eds. *Language Universals*. Oxford: Oxford University Press. pp. 253-261.

- Claverie, B. (2010). « Pluri-, inter-, transdisciplinarité : ou le réel décomposé en réseaux de savoir ». *Projectics/Proyéctica/Projectique*. N°4. pp. 1-14.
- Collet, G., Leybaert, J., Serniclaes, W., Deltenre, P., Markessis, E., Hoonhorst, I. & Colin, C. (2014). « Les entraînements auditifs : des modifications comportementales aux modifications neurophysiologiques ». *L'Année psychologique*. Volume 114, Issue 02. pp. 389-418.
- Conein, B. (1990). « Formalité et contingence interactionnelle : échange des noms et des salutations au téléphone ». *Réseaux. Communication – Technologie – Société*. Volume 8, N° 1. pp. 145 – 160.
- Corbel, E. (2005). « Considération sur l'interprétation chez Vladimir Jankélévitch et Igor Stavinsky ». *Horizons philosophiques*. Volume 16, N°1. pp. 74-85.
- Cornaz, S., Henrich, N. & Vallée, N. (2010). « L'apport d'exercices en voix chantée pour la correction phonétique en langue étrangère : le cas du français langue étrangère appliqué à des apprenants italiens d'âge adulte ». *Cahiers de l'APLIUT*. Volume XXI, N° 2. pp. 103-119.
- Cornaz, S. & Caussade, D. (2014). « Musique, voix chantée et apprentissage : une revue de la littérature et quelques propositions d'exploitation en didactique de la phonétique des langues ». *Editions du CRINI*. N°6. 34 pages.
- Cornaz, S. & Fonio, F. (2014). « Présentation et premiers résultats de la conception d'un référentiel de compétences en pratiques artistiques et apprentissage des langues ». *Editions du CRINI*. N°6. 32 pages.
- Corral Fulla, A. (2008). « Méthodologie musicale et enseignement-apprentissage du FLE ». *Synergies Espagne*. N° 1. pp. 141-152.
- Crepault, C. (1997). *La sexoanalyse*. Paris : Payot.
- Cross, I. (2003). "Music, cognition, culture, and evolution". In Wallin, N.L., Merker, B. & Brown, S. Eds. *The Origins of Music*. Cambridge, MA: MIT Press. pp 42-56.
- Damasio, A. (2010). *L'autre moi-même : les nouvelles cartes du cerveau, de la conscience et des émotions*. Paris : Odile Jacob.
- Darsel, S. (2010). *De la musique aux émotions*. Rennes : Presses universitaires de Rennes.
- Darwin, C. (1871). *The descent of Man, and Selection in Relation to Sex*. London: John Murray.
- Daumas, C. (2005). « Homme ou femme, le cerveau ne voit pas la différence ». *Libération en ligne*. 04/06/05. <http://www.liberation.fr/week-end/0101531664-hommes-ou-femmes-le-cerveau-ne-voit-pas-la-difference> (Consulté le 08/08/12).
- Davies, P. (1988). *The cosmic Blueprint*. New-York: Simon and Schuster.
- De Bot, K., Lowie, W. & Verspoor, M. (2007). "A dynamic systems theory approach to second language acquisition". *Bilingualism: Language and Cognition*. 10 (1). pp. 7-21.
- Deacon, T.W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York : W.W. Norton.

- Dehaene, S. (2007). *Les neurones de la lecture*. Paris : Odile Jacob.
- Dehaene, S. (2012). « Introduction au raisonnement Bayésien et à ses applications ». Cours donné dans le cadre de la chair de psychologie cognitive du Collège de France le 10 Janvier.
- Della Chiesa, B. (2007). *Comprendre le cerveau : naissance d'une science de l'apprentissage*. Ouvrage collaboratif : collectifs CERI et OCDE. Paris : édition de l'OCDE.
- Demaizière, F. (2007). « Didactique des langues et TIC : les aides à l'apprentissage ». *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*. Volume 10, N° 1. pp. 5-21.
- Deonna, J.A. & Teroni, F. (2009). « L'intentionnalité des émotions : du corps aux valeurs ». *Revue européenne des sciences sociales*. Tome XLVII, N°144. pp. 25-41.
- Détrie, C., Siblot P. & Vérine B. (2001). *Termes et concepts pour l'analyse du discours : Une approche pragmatique*. Collection « Bibliothèque elzévirienne. Paris : Honore Champion.
- Deutsch, D. & Dooley, K. (2009). “Absolute pitch among students in an American music conservatory: Association with tone language fluency”. *J. Acoust. Soc. Am.* 125 (4). pp. 2398-403.
- Di Cristo, A. (2002). « Prosodie et cognition ». Conférence. Université Toulouse le Mirail. Non publiée.
- Diehl, R.L., Lotto, A.J. & Holt, L.L. (2004). “Speech Perception”. *Annu. Rev. Psychol.* 55. pp. 149-79.
- Doudin, P.-A. & Martin, D. (1992). *De l'intérêt de l'approche métacognitive en pédagogie*. Lausanne : Centre Vaudois de Recherches Pédagogiques.
- Dubé, S. (2001). « La musique, un extraordinaire support à l'apprentissage ». Atelier présenté dans le cadre du *colloque Cerveau, et apprentissage, Intelligences multiples* 8, 9, 10 novembre 2001, Rimouski.
- Dunbar, R.I.M. (2004). *The Human Story*. London: Faber and Faber.
- Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B. & Taub, E. (1995). “Increased cortical representation of the fingers of the left hand in string players”. *Science*. 270. pp. 305–307.
- Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P. & Paabo, S. (2002). “Molecular evolution of FOXP2, a gene involved in speech and language”. *Nature*. 418. pp. 869-72.
- Ernst, M.O. & Bühlhoff, H. (2004). “Merging the senses into a robust percept”. *Trends in cognitive Sciences*. 8 (4). pp 162-9.
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002). “Speech listening specifically modulates the excitability of tongue muscles: a TMS study”. *European Journal of Neuroscience*. 17. pp. 1703-14.
- Fant, G. (1960). *Acoustic theory of speech production*. Mouton: The Hague.
- Fitch, W. T. (2006). “The biology and evolution of music: a comparative perspective”. *Cognition*. 100. pp. 173-215.

- Flege, J.E. & Hillenbrand, J. (1984). "Limits on phonetic accuracy in foreign language speech production". *Journal of the Acoustical Society of America*. 76(3). pp. 708-721.
- Flege, J.E. (1995a). "Second language speech learning: Theory, findings, and problems". In Strange, W. Ed. *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. pp. 233–277.
- Flege, J.E. (1995b). "Two procedures for training a novel second language phonetic contrast". *Applied Psycholinguistics*. 16. pp. 425-42.
- Flege, J.E., Yeni-Komshian, G.H. & Liu, S. (1999). « Age constraints on second-language acquisition ». *J. Mem. Lang.* 41. pp. 78–104.
- Fodor, J.A. (1975). *The Language of Thought*. New York: Crowell.
- Fonagy, I. (1983). *La vive voix : Essais de psycho-phonétique*. Paris: Payot.
- Fowler, C.A. (1996). "Listeners do hear sounds, not tongues". *Journal of the Acoustical Society of America*. 99. pp. 1730–41.
- Frith, C. (2010). *Comment le cerveau crée notre univers mental*. Paris: Odile Jacob.
- Fujisaki, H. & Kawashima, T. (1971). "A model of the mechanism for speech perception: quantitative analysis of categorical effects in discrimination". *Annual Report of the Engineering Research Institute*. Faculty of Engineering, University of Tokyo. Volume 30. pp. 59–68.
- Gardner, H. (1997). *Les formes d'intelligence*. Paris: Edition Odile Jacob sciences.
- Gardner, C. (2009). "The ears have it". *BBC Music Magazine*. pp. 37-8.
- Gazzaniga, M. S., Ivry, R. B. & Mangun, G. R. (2001). *Neurosciences cognitives : la biologique de l'esprit*. Paris: DeBoeck Université.
- Gelman, A. (2013). "P Values and Statistical Practice". *Epidemiology*. Volume 24, N°1. January. pp. 69-72.
- Gendrot, C. & Adda-Decker, M. (2005). "Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German". *Processings of Interspeech*. pp. 2453-2456.
- Gendrot, C., Adda-Decker, M. & Vaissière, J. (2008). « Les voyelles /i/ et /y/ du français : focalisation et variations formantiques ». *Actes des XXVIIèmes Journées d'Étude sur la Parole*. Avignon. pp. 205-8.
- Geoffroy, R., Lo Cascio, D. & Rivas, M. (1991). *La Clé des champs I*. Paris : CLE international.
- Georgeton, L., Paillereau, N., Landron, S., Gao, J. & Kamiyama, T. (2012). « Analyse formantique des voyelles orales du français en contexte isolé : à la recherche d'une référence pour les apprenants de FLE ». *Actes des XXIXièmes Journées d'Étude sur la Parole*. Grenoble. Juin. pp. 145–152.
- Gerrits, E. (2001). *The categorisation of speech sounds by adults and children*. The Netherlands: Netherlands Graduate School of Linguistics.

- Gerrits, E. & Schouten, M.E.H. (2004). "Categorical perception depends on the discrimination task". *Perception & Psychophysics*. 66(3). pp. 363-76.
- Geschwind, N. & Levitsky, W. (1968). "Human brain: left-right asymmetries in temporal speech region". *Science*. Volume 161. pp. 186-7.
- Gladwell, M. (2008). *Outliers, the story of success*. London: Penguin.
- Gleick, J. (1987). *Chaos making a New Science*. New-York: Penguin Book.
- Goody, J. (1994). *Entre l'oralité et l'écriture*. Paris: PUF.
- Gordon, J. & Grey, J. M. (1978). "Perceptual Effects of Spectral Modifications on Orchestral Instrument Tones". *Computer Music Journal*. 2(1). pp. 24-31.
- Graham, C. (2000). *Jazz Chants: Old and New*. Oxford: Oxford University Press.
- Grahn, J.A. (2011). "Advances in neuroimaging techniques: implications for the shared syntactic integration resource hypothesis". In Rebuschat, P., Rohrmeier, M., Hawkins, J. A. & Cross, I. *Language and Music as Cognitive Systems*. Oxford: Oxford University Press. pp. 235-42.
- Green, K.P. (1996). "Studies of the McGurk effect: implications for theories of speech perception". *Spoken Language, ICSLP, Proceedings*. pp. 1652-1655.
- Guimbretière, E & Kaneman-Pougatch, M. (1991). *Plaisir des sons: Enseignement des sons du français*. Paris: Les Éditions Didier.
- Guimbretière, E. (1997). *Phonétique et enseignement de l'oral*. Paris: Didier/Hatier.
- Guion, S. G., Flege, J. E., Akahane-Yamada, R. & Pruitt, J. C. (2000). "An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants". *Journal of the Acoustical Society of America*. 107 (5). pp. 2711-24.
- Hallé, P., Chéreau, C. & Segui, J. (2000). « Where is the /b/ in « absurde » [apsyrd]? It is in French Listeners' Minds ». *Journal of Memory and Language*. 43. pp. 618-639.
- Hallé, P., Best, C., & Bachrach, A. (2003). "Perception of /dl/ and /tl/ clusters: A cross-linguistic study with French and Israeli listeners". *International Congress of Phonetic Sciences*. Barcelona, Spain. pp. 2893-96.
- Hallé, P., Chang, Y.-C. & Best, C. (2004). "Categorical perception of Taiwan Madarin Chinese tones by Chinese versus French native speakers". *Journal of Phonetics*. 32. pp. 395-421.
- Hallé, P. & Best. C. (2007). "Dental-to-velar perceptual assimilation: a cross-linguistic study of the perception of dental stop+/l/ clusters". *Journal of the Acoustical Society of America*. 121 (5). pp. 2899-2914.
- Hardison, D.M. (1996). "Bimodal speech perception by native and nonnative speakers of English: factors influencing the McGurk effect". *Lang Learn*. 46. pp. 3-73.
- Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). "Revisiting the dissociation between singing and speaking in expressive aphasia". *Brain*. 126. pp.1838-1850.

- Hebert, S. & Cuddy, L.L. (2006). "Music-reading deficiencies and the brain". *Advances in Cognitive Psychology*. Volume 2, N°2-3. pp. 199-206.
- Hesling, I. (2002). « L'hémisphère cérébral droit : un atout en anglais de spécialité ». *ASp*. pp. 37-38.
- Hickok, G., Bellugi, U. & Klima, E. S. (1998). "The neural organization of language: evidence from sign language aphasia". *Trends in Cognitive Sciences*. 2(4). pp. 129-36.
- Hoch, L., Tillmann, B & Poulin-Charronat, B. (2008). « Musique, syntaxe et sémantique: des ressources d'intégration structurale et temporelle partagée? ». *Revue de Neuropsychologie*. 18(1-2). pp. 33-59.
- Hombert, J.-M., Ohala J. J. & Ewan W. G. (1979). "Phonetic Explanation for the Development of Tones". *Language*. 55(1). pp 37-58.
- Honikman, B. (1964). "Articulatory Settings". In Abercrombie, D., Fry, D.B., MacCarthy, P.A.D., Scott, N. C. & Trim, J.L.M. Eds. In *Honour of Daniel Jones: Papers Contributed on the Occasion of His Eightieth Birthday*. 12 September 1961. pp. 73-84.
- Huhtanen, K. (2008). "Adult as learners in the field of music: a pedagogical challenge". *Proceedings of the 17<sup>th</sup> International Seminar of the Commission for the Education of the Professional Musician (CEPROM)*, International Society for Music Education (ISME). pp. 57-61.
- Hunt, K.D. (1994). "The evolution of human bipedality: ecology and functional morphology". *Journal of Human Evolution*. 26. pp. 183-202.
- Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C. & Schlaug, G. (2009). "Musical training shapes structural brain development". *J. Neurosci*. 29. pp. 3019–3025.
- Hyman, L. (1975). *Phonology: Theory and Practice*. New York: Holt.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J. C. & Rizzolatti, G. (1999). "Cortical mechanisms of human imitation". *Sciences*. pp. 2526-2528.
- Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C. & Rizzolatti, G. (2001). "Reagengerent copies of imitated actions in the right superior temporal cortex". *Proceedings of National Academy of Sciences of USA*. 98, 24. pp. 13995-13999.
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Iverson, P & Kuhl, P. (1995). "Mapping the perceptual magnet effect for speech using signal theory and multidimensional scaling". *Journal of the Acoustical Society of America*. 94 (1). pp. 553-62.
- Jalilzadeh, K. (2011). "Emergentism in First Language Acquisition". *European Journal of Scientific Research*. 57(1). pp. 152-155.
- Jodelet, D. (2003). *Les représentations sociales*. Paris : PUF.
- Jones, D. (1931). "On phonemes". *Travaux du Cercle Linguistique de Prague*. N°4. pp. 74-79.

- Jones, D. (1956). *Cardinal Vowels*. 8<sup>th</sup> ed. Cambridge: W. Heffer & Sons.
- Joseph, R. (1990). *Neuropsychology, Neuropsychiatry, and Behavioral Neurology*. New York: Plenum Press.
- Kamiyama, T. (2009). *Apprentissage phonétique des voyelles du français langue étrangère chez des apprenants japonophones*. Thèse de doctorat de phonétique soutenue à l'université Paris 3 Sorbonne Nouvelle.
- Kim, H.-G., Moreau, N., & Sikora, T. (2005). *MPEG-7 Audio and Beyond*. Wiley.
- Kirby, S. (2001) "Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity". *IEEE Transactions on Evolutionary Computation*. 5(2). pp. 102-110.
- Kirby, S. & Christiansen, M. H. (2003). "From language learning to language evolution". In Christiansen, M. and Kirby, S. Eds. *Language Evolution*. Oxford: Oxford University Press. pp. 272–294.
- Kivy, P. (2006). *The performance of reading: an essay in the philosophy of literature*. Blackwell Publishing.
- Krashen, S.D. (1988). *Second Language Acquisition and Second Language Learning*. London: Prentice Hall.
- Kraus, N. & Chandrasekaran, B. (2010). "Music training for the development of auditory skills". *Nature Reviews: Neuroscience*. Volume 11. pp. 599-605.
- Kris, E. (1952). *Psychoanalytic Explorations in Art*. New York: International University Press.
- Kuhl, P. K. (1993). "Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception". *Journal of Phonetics*. 21. pp. 125-39.
- Kuhl, P.K. & Iverson, P. (1995). "Linguistic experience and the "perceptual magnet effect"". In Strange, W. Ed. *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press. pp. 21-154.
- Kuhl, P. (2004). "Early language acquisition: cracking the speech code". *Nature Reviews Neurosciences*. 5. pp. 831-843.
- Kuhl, P.K. (2010). "Brain mechanisms in Early Language". *Neuron*. 67. pp. 713-87.
- Kutas, M., & Hillyard, S.A. (1983). "Event-Related brain Potentialsto grammatical errors and semantic anomalies". *Memory & Cognition*. 11. pp. 539–550
- Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. & Monaco, A.P. (2001). "A forkhead-domain gene is mutated in a sever speech and langage disorder". *Nature*. 413. pp. 519-23.
- Larsen-Freeman, D. (1997). "Chaos/Complexity Science and Second Language Acquisition". *Applied Linguistics*. 18 (2). pp. 141-165.
- Lashley, K. S. (1951). "The problem of serial order in behaviour". In Jeffress, L. A. (Ed.). *Cerebral mechanisms in behavior: The Hixon symposium*. New York: Wiley. pp. 112-136.



- Lauret, B. (1998). *Aspect de Phonétique Expérimentale Contrastive : l'accent anglo-américain en français*. Thèse de doctorat de phonétique soutenue à l'université Paris 3 Sorbonne Nouvelle.
- Lauret, B. (2007). *Enseigner la prononciation du français : questions et outils*. Paris : Hachette.
- Legallois, D. (2006). « Mémento sur quelques rapports entre mémoire et linguistique ». *Questions de style*. N°6. pp. 1-21.
- Legrain, M. (2007). *Théorie et pratique des enquêtes par questionnaire*. Liège: Edipro.
- Leipp, E. (1989). *Acoustique et musique*. Paris : Masson.
- Léon, P. R. (2000). *Phonétisme et prononciations du français*. (4ème édition). Paris: Nathan.
- Levitin, D. & Zatorre, R. J. (2003). "On the nature of early music training and absolute pitch: A reply to Brown, Sachs, Commuso and Foldstein". *Music perception*. 21. pp. 105-10.
- Levitin, D. (2010). *De la note au cerveau : l'influence de la musique sur le comportement*. Paris: Editions Héloïse d'Ormesson.
- Lhote, E. (1995). *Enseigner l'oral en interaction : percevoir, écouter, comprendre*. Paris: Hachette, Collection Autoformation.
- Li, C. & Hombert, J.-M. (2002). "On the evolutionary origin of language". In Stamenov M. I. & Gallese, V. (2002). *Mirror Neurons and the Evolution of Brain and Language*. Amsterdam: John Benjamins Publishing. pp. 175-205.
- Lieberman, A., Delattre P., Cooper F.S. (1952). "The role of selected stimulus-variables in the perception of the unvoiced stop consonants". *Am. J. Psychol.* 65. pp. 497-516.
- Lieberman, A., Delattre, P.C., Cooper, F.S., & Gerstman, L.J. (1954). "The role of consonant-vowel transitions in the stop and nasal consonants". *Psychol. Monogr.* 68. pp. 1-13.
- Lieberman, A., Cooper, F., Shankweiler, D. & Studdert-Kennedy, M. (1967). "Perception of the speech code". *Psychological Review*. Volume 74. pp. 431-461.
- Lieberman, M. (1984). *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press.
- Lieberman, A. & Mattingly, I. (1985). "The motor theory of speech perception revised". *Cognition*. Volume 21. pp. 1-36.
- Lindbom, B. & Sundberg, J. (1969). "A quantitative theory of cardinal vowels and the teaching of pronunciation". *Speech Transmission Laboratory Quarterly Progress Status Report*. Stockholm, 10. pp. 19-25.
- Llorca, R. (1998). *Les Ritmimots, exercices de groupes avec la voix et le geste sur les rythmes du français parlé*. Munich : Bayerischer Rundfunk.
- Llorca, R. (1999). « Les Ritmimots : le français par l'oreille musicale ». *Schulfunk*. N° 10. pp. 44-45.

- Ludke, K. M. & Weinmann, H. (2012). *European Music Portfolio: a creative way into languages. Comenius Lifelong Learning Project*. En ligne: <http://www.emportfolio.eu/> . (Consulté le 05/08/13).
- Luzzati, D. (2007). « Le dialogue oral spontané: quels objets pour quels corpora ». *Revue d'interaction Homme-Machine*. Volume 8, N°2. pp. 23-36.
- Macaire, C., Narcy-Combes, J.-P. & Portine, H. (2010). « Interrogations épistémologiques en didactique des langues ». *Le français dans le monde, Recherches et applications*. Paris: CLE International. 192 pages.
- Macaire, D. (2010). « Recherche-action et didactique des langues : du positionnement du chercheur à une posture de recherche ». *Les après-midis du LAIRDIL, numéro thématique : Didactique des langues, didactique des sciences* ». N°17, novembre 2010, Toulouse : LAIRDIL. pp. 21-32.
- MacLarnon, A. & Hewitt, G. (1999). "The evolution of human speech: The role of enhanced breathing control". *American Journal of Physical Anthropology*. 109. pp. 341-363.
- MacWhinney, B. (1998). "Models of the emergence of language". *Annual Review of psychology*. N°49. pp. 199-227.
- MacWhinney, B. (2006). "Emergentism- Use often and with Care". *Applied Linguistics*. 27:4. pp. 729-40.
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press
- Malinen, A. (2000). *Towards the essence of adult experimental learning*. Finlande: Sophi Academic Press.
- Mandelbaum, D.G. (1949). *Selected Writings of Edward Sapir in Language, Culture and Personality*. Los Angeles : University of California Press.
- Marler, P. (2000). "Origins of music and speech: Insights from animals". In Wallin, N.L., Merker, B. & Brown, S. Eds. *The Origins of Music*. Cambridge, MA: MIT Press. pp 31-48.
- Marschark, M., Siple, P., Lillo-Martin, D., Campbell, R. & Everhart, V. S. (1997). *Relations of language and thought: the view from Sign Language and Deaf Children*. New York: Oxford University Press.
- Martinet, A. (1970). *Éléments de linguistique générale*. 4ième édition. Paris : Armand Colin.
- Massaro, D.W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Massaro, D.W. (1998). *Perceiving talking faces. From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- McGurk, H. & MacDonald, J. (1976). "Earing lips and seeing voices". *Nature*. 264. pp. 746-748.
- Memmi, D. (1995). « Le connexionnisme est-il réductionniste ? ». *Intellectica*. N°21. pp. 262-265.

- Meunier, C. (2007). « Chapitre 13 : Phonétique Acoustique ». In Auzou, P. Eds. *Les Dysarthries*. pp. 164-173.
- Milgrom, M. (2002). “Does dark matter really exist?” *Scientific American*. pp. 42-52.
- Miller, G. (2000). “Evolution of human music through sexual selection”. In Wallin, N.L., Merker, B. & Brown, S. Eds. *The Origins of Music*. Cambridge, MA : MIT Press. pp. 29-60.
- Miras, G. (2013). « « Enseigner/apprendre » la prononciation autrement : une approche psychosociale musique-parole ». *Les cahiers de l'ACEDLE*. Volume 10, N°1. pp. 53-87.
- Miras, G. & Narcy-Combes, J.-P. (2014). « Conséquences sur les pratiques d'une prise en compte intégrée des théories socio-constructivistes et émergentiste ». *Travaux et document*. Mars, N°46. pp. 15-26.
- Miras, G. (à paraître). « Facteurs environnementaux et représentations sur la lecture-écriture. Une étude transdisciplinaire : musique, parole et langue des signes ». *Histoire Epistémologie Langage*. 22 pages.
- Mithen, S. (2006). *The Singing Neanderthals*. Cambridge: Harvard University Press.
- Mohanan, K. P. (1992). “Emergence of complexity in phonological development”. In Ferguson, C., Menn, L. & Stoel-Gammon, C. Eds. (1992). *Phonological Development*. Timonium: York Press. pp. 635-62.
- Montagu, J. (2007). *Analyse acoustique et perceptive des voyelles nasales et nasalisées du français parisien*. Thèse de Doctorat en phonétique, Université Paris III.
- Morin, E. (1994). « Sur l'interdisciplinarité ». Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires. N° 2. Juin. En ligne : <http://ciret-transdisciplinarity.org/bulletin/b2c2.php> (consulté le 09/08/13). Une version de ce texte a été publiée dans Kourilsky, F. (1990). Carrfour des sciences. Actes du colloque du Comité national de la Recherche Scientifique. Paris: Edition du CNRS. pp. 1-7.
- Morin, E., & Lemoigne, J.-L. (1999). *Introduction à la pensée complexe*. Paris: L'Harmattan.
- Morin, E. (2000). *Les Septs savoirs nécessaires à l'éducation du futur*. Paris : Le Seuil.
- Morley, I. (2003). *The evolutionary Origins and Archaeology of Music: An investigation into the Prehistory of Human Musical Capacities and Behaviours*. Ph.D. Dissertation, University of Cambridge.
- Mucchielli, R. (1967). *Le questionnaire dans l'enquête psycho-sociale*. 10<sup>ème</sup> édition. Paris: Formation permanente en sciences humaines.
- Münste, T. F., Altenmüller, E. & Jäncke, L. (2002). “The musician's brain as a model of neuroplasticity”. *Nature Reviews, neurosciences*. Volume 3. pp. 473-8.
- Näätänen, R., Gaillard A.W. & Mäntysalo S. (1978). “Early selective-attention effect on evoked potential reinterpreted”. *Acta Psychol.* 42 (4). pp. 313-29.
- Narcy-Combes, J.-P. (2002). « Comment percevoir la modélisation en didactique des langues. ». *ASp*. pp.35-36. En ligne: <http://asp.revues.org/1678> (consulté le 25/09/11).

- Narcy-Combes, J.-P. (2005). *Didactique des langues et TIC : Vers une recherche-action responsable*. Paris : Ophrys.
- Narcy-Combes, J.P. (2006a). « Deux modes de fonctionnement mémoriel en production langagière et tâches d'apprentissage des langues ». *Cahiers de l'APLIUT*. Volume XXV, N°2. pp. 77-87.
- Narcy-Combes, J.-P. (2006b). « La didactique de L2 à la croisée des chemins ». *Les Cahiers de l'Acedle*. N° 2. pp. 301-18
- Narcy-Combes, J.-P. & Bertin, J.-C. (2012a). "Tutoring at a distance, Modelling as a tool to control chaos". *CALL*. Volume 25, N°2. pp. 111-127.
- Narcy-combes, J.-P. (2012b). « Propositions pour intégrer contenu et langue(s): allier l'approche par tâches en langue et une pédagogie disciplinaire de projet ou de résolution de problèmes ». Dans Causa, M., Derivrey-Plard, M., Lutrand Pezant, B. & Narcy-Combes, J.-P. Eds. (2012). *Les langues dans l'enseignement supérieur: quels contenus pour les filières non linguistiques?* Paris: Riveneuve
- Narcy-Combes, J.-P. & Miras, G. (2013). « 40 ans de modélisation en didactique des langues ». *Mélanges CRAPEL*. N° 33. pp. 25-44.
- Newman, AJ., Bavelier, D., Corina, D., Jezzard, P. & Neville, HJ. (2002). "A critical period for right hemisphere recruitment in American Sign Language processing". *Nature Neurosciences*. Jan, 5 (1). pp. 76-80.
- Nguyen, N. (2005). « Perception de la parole ». In Nguyen, N., Wauquier-Gravelines, S. & Durand, J. *Phonologie et Phonétique : Forme et Substance*. Chapitre 16. Paris : Lavoisier. En ligne: [http://aune.lpl.univ-aix.fr/~nguyen/nm\\_hermes.pdf](http://aune.lpl.univ-aix.fr/~nguyen/nm_hermes.pdf) (consulté le 17/04/2013)
- Niemitz, C. (2010). "The evolution of the upright posture and gait – a review and a new synthesis ». *Naturwissenschaften*. 97. pp. 241-263.
- Norrix, L. W. & Green, K.P. (1996). "Auditory-visual context effects on the perception of /r/ and /l/ in a stop cluster". *Journal of the Acoustical Society of America*. 99(4). pp. 2591-2603.
- Noyau, C. (1988). « Recherches sur l'acquisition spontanée d'une langue étrangère par des adultes dans le milieu social ». *Dialogue et cultures*. pp. 208-218.
- O'Grady, W. Lee, M. & Kwak, H.-Y. (1997). "Emergentism and Second Language Acquisition". In Ritchie, W. & Bhatia, T. Eds. *The handbook of Second Language Acquisition*. Second edition. pp. 69-88.
- O'Grady, W. (2010). "Emergentism". Cambridge Encyclopedia of Language Science. En ligne: <http://www.ling.hawaii.edu/faculty/ograde/Emergentism.pdf> (consulté le 22/09/11)
- Ollat, H. (1999). « Schizophrénie et mémoire de travail ». *Neuropsychiatrie : Tendances et Débats*. 4. pp. 43-50.
- Pamula, M. (2008). « Sensibiliser les enfants à une langue étrangère par le biais d'une activité musicale ». *Synergies Espagne*. N° 1. pp. 133-140.

- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E. & Hoke, M. (1998). "Increased auditory cortical representation in musicians". *Nature*. 392. pp. 811–814.
- Pantev, C., Roberts, L.E., Schulz, M., Engelien, A. & Ross, B. (2001). "Timbre-specific enhancement of auditory cortical representations in musicians". *Neuroreport*. 12. pp. 169–174.
- Pastore, R.E., Friedman, C.J. and Buffato, K.J. (1976). "A comparative evaluation of the AX and two ABX procedures". Paper presented at the 92nd meeting of the Acoustical Society of America, San Diego.
- Patel, A. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
- Peretz, I. (1985). « Les différences hémisphériques dans la perception des stimuli musicaux chez le sujet normal. II. Les sons simultanés ». *L'année psychologique*. 85(4). pp.567-576.
- Peretz, I. & Kolinsky, R. (1993). "Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective". *The Quarterly Journal of Experimental Psychology*. 46. pp. 301-325.
- Peretz, I., Belleville, S., & Fontaine, F. (1997). « Dissociations entre musique et langage après atteinte cérébrale: un nouveau cas d'amusie sans aphasie ». *Canadian Journal of Experimental Psychology*. 51. pp. 354-367.
- Peretz, I. & Coltheart, M. (2003). "Modularity of music processing". *Nature Neuroscience*. 6(7). pp. 688-91.
- Peretz, I., Gagnon, L., Hébert, S., & Macoir, J. (2004). "Singing in the brain: Insights from cognitive neuropsychology". *Music Perception*. 21. pp. 373-390.
- Peretz, I. (2006). "The nature of music from a biological perspective". *Cognition*. 100. pp. 1-32.
- Pernon, G. (1998). *Histoire de la musique*. Paris : J-P Gisserot.
- Perrot, X. (2010). « Anatomie et physiologie du système nerveux auditif central ». *Les cahiers de l'Audition*. N°6. pp. 5-16.
- Piaget, J. (1964). *Six études de psychologie*. Paris : Folio.
- Piaget, J. (1968). *Le structuralisme*. Paris : PUF. 12ième édition, n°1311.
- Piaget, J. (1970). *Genetic epistemology*. New York: W.W. Norton & Company.
- Pillot, C. (2004). *Sur l'efficacité vocale dans le chant lyrique. Aspects physiologiques, cognitif, acoustique et perceptif. Thèse de doctorat de phonétique*. Université Paris III Sorbonne Nouvelle.
- Pillot-Loiseau, C., Amelot, A. & Fredet, F. (2010). « Apports de la phonétique expérimentale à la didactique de la prononciation du français langue étrangère ». *Cahiers de l'APLIUT*, Volume XXIX, N° 2. pp. 75-88.
- Pinker, S. & Jackendoff, R. (2005). "The faculty of language: what's special about it?". *Cognition*. 95. pp. 201-36.

- Pisoni, D.B (1973). "Auditory and phonetic memory codes in the discrimination of consonants and vowels". *Perception & Psychophysics*. Volume 13. pp. 253-260.
- Pisoni, D.B. & Tash, J. (1974). "Reaction times to comparisons within and across phonetic categories". *Perception and Psychophysics*. Volume 15, N°2. pp. 285-290.
- Porée, R. (2012). *Evaluation de l'ergonomie d'un logiciel d'apprentissage de la phonétique française: Speedlingua*. Mémoire de Master 2 professionnel de l'Université Stendhal Grenoble 3. En ligne : <http://dumas.ccsd.cnrs.fr/dumas-00743370/> (consulté le 24/04/13)
- Posner, M.I. & Mitchell, R.F. (1967). "Chronometric analysis of classification". *Psychol Rev.* 74(5). pp. 392-409.
- Posner, M.I. (1969). "Abstraction and the process of recognition". In Bower, G.H., Spence, J.T. Eds. *The psychology of learning and motivation*. New York: Academic Press. pp. 44–100.
- Posner, M.I., Boies, S.J., Eichelman, W.H. & Taylor, R.L. (1969). "Retention of visual and name codes of single letters". *Journal of Experimental Psychology Monograph*. 79(1). pp. 1–16.
- Protopapas, A. & Calhoun, B. (2000). "Adaptive phonetic training for second language learners". *Proceedings of InSILL*. 8 pages.
- Randall, M. (2007). *Memory, Psychology and Second Language Learning*. Amsterdam: John Benjamins.
- Regan, V. (2002). « Le contexte d'acquisition: la variation du groupe et de l'individu ». *AILE*. 17. pp. 123-142.
- Reitan, I. E. (2008). "Stress and Well-being in the Aural Training Class – The Psychological Aspect of Training for Enhanced Musician's Skills". Proceedings of the 17th International Seminar of the Commission for the Education of the Professional Musician (CEPROM), International Society for Music Education (ISME). pp. 85-90.
- Renard, R. (1978). *Initiation phonétique à l'usage des professeurs de langues*. Bruxelles : Didier.
- Renard, R. (2002). *Apprentissage d'une langue étrangère-seconde. 2, La phonétique verbo-tonale*. Bruxelles: De Boeck Université.
- Rialland, A. (2005). « Phonological and phonetic aspects of whistled languages ». *Phonology*. Volume 22, N°2. pp. 237-271.
- Rizzolatti, G. & Sinigaglia, C. (2006). *Les neurones miroirs*. Paris : Odile Jacob.
- Rolland, Y. (2011). *Apprendre à prononcer. Quels paradigmes en didactique des langues?* Paris: Belin.
- Rosenthal, V. (2011). « Synesthésie en mode majeur. Une introduction ». *Intellectica*. Volume 55. pp. 7-46.
- Rutherford, W. (1987). *Second Language Grammar: Learning and Teaching*. London: Longman.

- Sapir, E. (1985). *Culture, language and personality: Selected essays by Edward Sapir*. Berkeley: University of California Press.
- Saussure (de), F. (1967). *Cours de linguistique générale*. Originellement publié par Bailly, C. & Sechehaye, A. Édition critique par Mauro (de), T. Paris : Payot.
- Savouret, A. (2002). « Electroacoustique et perspectives phonoculturelles ». *Circuit : musique contemporaines*. 13 (1). pp. 9-20.
- Schaeffer, P. (1966). *Traité des objets musicaux*. Paris : Éditions du Seuil.
- Schiffmann, S. N. (2001). « Le cerveau en constante reconstruction : le concept de plasticité cérébrale ». *Cahiers de psychologie clinique*. 1/2001, N°16. pp. 11-23.
- Schneider, K., Dogil, G. and Möbius, B. (2011). “Reaction time and decision difficulty in the perception of intonation”. *Proc. Interspeech*. pp. 2221-2224
- Schneuwly, B. (2008). « Vygotski, l'école et l'écriture ». *Cahiers de la section des sciences de l'éducation : Pratiques et théorie*. N°118.
- Schön, D., Anton, J.-L. Roth, M. & Besson, M. (2002). “An fMRI study of music sight-reading”. *Brain Imagin, NeuroReport*. 13(17). pp. 2285-9.
- Schrenk, F., Sandrock, O., Kullmer, O. (2004). “An “open source” perspective of earliest hominid origins”. *Coll Anthropol*. 2, Suppl. 2. pp. 65–76.
- Schuh, R. G. (2013). “An impertinent account of Korean vowels”. Manuscript. UCLA. En ligne:[http://www.linguistics.ucla.edu/people/schuh/Papers/ms\\_2013\\_Korean\\_vowels.pdf](http://www.linguistics.ucla.edu/people/schuh/Papers/ms_2013_Korean_vowels.pdf) (consulté le 31/07/14).
- Schwartz, J.-L., Boë, L.-J., Vallée, N. et Abry, C. (1997). “The Dispersion-Focalization Theory of vowel systems”. *Journal of Phonetics*. 25(3). pp. 255-286.
- Scotto Di Carlo, N. (1990). « L'arme secrète des chanteurs d'opéra ». *La Recherche*. XXI. 218. pp. 246-248.
- Seron, X. (1998). « Le programme de la neuropsychologie ». In Seron, X. & Jeannerod, M. (1998). *Neuropsychologie humaine*. Paris : Mardaga. pp. 35-56
- Shaw, C. A. & McEachern, J. (2001). *Toward a Theory of Neuroplasticity*. Philadelphia: Psychology Press.
- Shiller, D. M., Rvachew, S. & Brosseau-Lapré, F. (2010). “Importance of the auditory perceptual target to the achievement of speech production accuracy”. *Canadian Journal of Speech-Language Pathology and Audiology*. 34(3). pp. 181-92.
- Shors, T. J. (2004). “Learning during stressful times”. *Learn. Mem.* 11. pp. 137-144.
- Simondon, G. (1989). *L'individuation psychique et collective*. Paris : Aubier.
- Simmonds, A.J., Wise, R.J.S, Dhanjal, N.S. & Leech, R. (2011). “A comparison of sensory motor activity during speech in first and second languages”. *Journal of neurophysiology*. July. pp. 470-478.



- Singleton, D. (1989). *Language Acquisition: the Age Factor*. London: Multilingual Matters.
- Stephan, A. (1999). "Varieties of emergentism". *Evolution and Cognition*. Volume 5, N°1. pp. 49-59.
- Strange, W. & Dittmann, S. (1984). "Effects of discrimination training on the perception of /r- l/ by Japanese adults learning English". *Perception & Psychophysics*. 36 (2). pp. 131-45.
- Tantibundhit, C.; Onsuwan, C.; Phienphanich, P. & Wutiwiwatchai, C. (2012). "Methodological Issues in Assessing Perceptual Representation of Consonant Sounds in Thai". *Interspeech*, ISCA, 4 pages.
- Terrien, P. (2012). « Le chant : de l'oralité à l'auralité ». In Terrien, P. & Leroy, J.-L. (2012). *La voix et l'éducation musicale. Contribution à la réflexion et à l'action pédagogique* (II). Paris : L'Harmattan.
- Thaut, M.H., McIntosh, K.W., McIntosh, G.C. & Hoemberg, V. (2001). "Auditory rhythmicity enhances movement and speech motor control in patients with Parkinson's disease". *Functional Neurology*. 16. pp. 163-172. Paris : Edition Mardaga.
- Tiberghien, G. (1999). « Chapitre 9 La mémoire Partie 1: Psychologique cognitive de la mémoire humaine ». In Seron, X. & Jeannerod, M. (1999). *Neuropsychologie humaine*.
- Tillmann, B., Madurell, F., Lalitte, P. & Bigand, E. (2005). « Apprendre la musique : perspectives sur l'apprentissage implicite de la musique et ses implications pédagogiques ». *Revue française de pédagogie*. N°152. pp. 63-77.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). "Understanding and sharing intentions: the origin of cultural cognition". *Behavioral and Brain Sciences*. 28. pp. 675-691.
- Tripier-Mondancin, O. (2008). « Structuration des programmes d'enseignement musical, en France, de 1925 à 1997 ». *L'Education musicale*. N°27. <http://www.leducation-musicale.com/> (consulté le 28/08/13).
- Troubetzkoy, N.S. (1939). *Principes de phonologie*. Paris: Klincksieck.
- Tsukada, K. (2012). "Non-native Japanese listeners' perception of vowel length contrast in Japanese and Modern Standard Arabic (MSA). *Second Language Research*. 28(2). pp. 151-168.
- Tubach, J.-P. (1989). *La parole et son traitement automatique*. Calliope. Paris: Masson.
- Twaddell, W. F. (1935/57). "On Defining the Phoneme". *Language Monograph* 16. Baltimore. Reprinted in Joos, 1957. pp. 55-80.
- Tyne, H. (2012). *Acquisition d'une langue seconde en milieu naturel : contextes, contacts, enjeux. Les migrants face aux langues des pays d'accueil : acquisition en milieu naturel et formation*. Lille : Septentrion.
- Vaissière, J. (2006). *La phonétique*. Que sais-je. Paris: PUF.



- Vaissière, J. (2007). "Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of sounds across languages". In Solé M. J., Beddor, P. S. & Ohala M. *Experimental Approaches to Phonology*. Oxford : Oxford University Press. pp. 54-71.
- Vaissière, J. (2008). « Ivan Fonagy ; un pionnier dans l'étude de l'expressivité dans la parole ». Conférence donnée dans le cadre du Festival Agora de l'Ircam. 5-20 Juin.
- Vaissière, J. (2009). "Articulatory modeling and the definition of acoustic-perceptual targets for reference vowels". *The Chinese Phonetics Journal*. 2. pp. 22-33.
- Vaissière, J. (2011). "On the acoustic and perceptual characterization of reference vowels in a cross-language perspective". *Actes du colloque ICPHS XVII à Hong Kong*. pp. 52-8.
- Vallejo-Gomez, N. (2008). « La pensée complexe : Antidote pour les pensées uniques. Entretien avec Edgar Morin ». *Synergies Monde*. N°4. pp. 249-262.
- Van Geert, P. (1994). *Dynamic systems of development. Change between complexity and chaos*. Hertfordshire, England: Harvester Wheatsheaf.
- Van Hessen, A. J., & Schouten, M. E. H. (1999). "Categorical perception as a function of stimulus quality". *Phonetica*. 56. pp. 56-72.
- Verspoor, M., de Bot, K. & Lowie, M. (2011). Eds. *A dynamic Approach to Second Language Development: Methods and Techniques*. John Benjamins Publishing Company.
- Verspoor, M., Schmid, M. S. & Xu, X. (2012). "A dynamic usage based perspective on L2 writing". *Journal of Second Language Writing*. pp. 239-263
- Vidal, C. (2005). *Cerveau, Sexe & Pouvoir*. Paris : Belin.
- Vos, J. & Dryden, G. (1999). *The Learning Revolution*. USA : Jalmar Pr.
- Vygotski, L. (1934). *Pensée et langage*. 3<sup>ème</sup> édition parue en 1997. Paris: La Dispute.
- Wachs, S. (2011). « Tendances actuelles en enseignement de la prononciation du français, langue étrangère (FLE) ». *Revista de Lenguas Modernas*. N°14. pp. 183-196.
- Watson, A. H. D. (2006). "What can studying musicians tell us about motor control of the hand". *J. Anat.* 208. pp. 527-542.
- Werker, J. F., & Tees, R. C. (1984a). "Cross-language speech perception: evidence for perceptual reorganization during the first year of life". *Infant Behavior and Development*. Volume 7. pp. 49-63.
- Werker, J.F. & Tees, R.C. (1984b). "Phonemic and phonetic factors in adult cross-language speech perception". *J. Acoust. Soc. Am.* 75(6). pp. 1866-76.
- Wernicke, C. (1874-1977). "Der aphasische symptomenskomplex: eine psychologische studie auf anatomischer basis" [la complexité du symptôme aphasique: une étude psychologique sur des bases anatomiques]. In Eggert, G.H. Eds. *Wernicke's works on aphasia: a sourcebook and review*. The Hague: Mouton. pp. 91-145.
- Whalen, D. & Levitt, A. G. (1995). "The universality of intrinsic F0 of vowels". *Journal of phonetics*. Volume 23, N° 3. pp. 349-66.

- Wheeler, P. (1988). "Stand tall and stay cool". *New Scientist*. 12. pp. 60-65.
- Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge: MIT Press.
- Wilhelm, S. (2012). « Prosodie et correction phonétique (partie 2) ». La Clé des Langues. Lyon: ENS LYON/DGESCO. Octobre. Consulté le 6 août 2013. pp. 1-22. En ligne: <http://cle.ens-lyon.fr/anglais/prosodie-et-correction-phonetique-partie-2--171334>.
- Wilson, I. L. (2006). *Articulatory Settings of French and English Monolinguals and Bilinguals*. Thèse de doctorant en phonétique soutenue à University of British Columbia.
- Wray, A. (2000). "Holistic utterances in protolanguages: the link from primates to humans". In Knight, C., Studdert-Kennedy, M. & Hurford, J.R. Eds. *The Evolutionary Emergence of Language: Social Function and the Origin of Linguistic Form*. Cambridge: Cambridge University Press. pp. 285-302.
- Wray, A. (2002a). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2002b). *The transition to language*. Oxford: Oxford University Press.
- Zatorre, R.J. (2003). "Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function". *Nature Neuroscience*. Volume 6, N°7, July. pp. 692-5.
- Zwicker, E. & Fastl, H. (2003). "Psychoacoustics Facts & Models, Chapter 7, Just Noticeable Changes". In *Sound*. (Ed.). Springer. pp. 174-185.

## Site internet

- Bbemg.ulg.ac.be (2012). Consulté en Février 2012 :  
<http://www.bbemg.ulg.ac.be/FR/2Notions/freqlg.html>
- Britannica.com (2013). Consulté en Février 2013 :  
<http://media-1.web.britannica.com/eb-media/89/389-004-2259D9A1.jpg>
- Encyclopedia Britannica (1997). Consulté en Mars 2013 :  
[http://theses.univ-lyon2.fr/documents/getpart.php?id=lyon2.2005.gabriel\\_d&part=97746](http://theses.univ-lyon2.fr/documents/getpart.php?id=lyon2.2005.gabriel_d&part=97746)
- Ciral.ulaval.ca (2012). Consulté en Mars 2013 :  
[http://www.ciral.ulaval.ca/phonetique/connaissances/phonetic\\_generale/acoustique/acoustique.htm](http://www.ciral.ulaval.ca/phonetique/connaissances/phonetic_generale/acoustique/acoustique.htm)
- CNRS.fr (2012). Consulté en Mars 2013 :  
<http://www2.cnrs.fr/journal/3451.htm>
- Coclea.org (2013). Consulté en Mars 2013 :  
<http://www.coclea.org/spe/cerveau-auditif-2.html>

Med.univ-montp1.fr (2013). Consulté en Mars 2013 :  
[http://www.med.univ-montp1.fr/enseignement/cycle\\_1/PCEM2/mod-integres/MI3\\_neuro/sensorialite/audition/cours/cours2-1.htm](http://www.med.univ-montp1.fr/enseignement/cycle_1/PCEM2/mod-integres/MI3_neuro/sensorialite/audition/cours/cours2-1.htm)

Larousse.fr (2013). Consulté en Mars 2013 :  
<http://www.larousse.fr/>

### **Logiciels**

Boersma, P. & Weenink D. Praat: doing phonetics by computer (mise à jour: 2013).  
<http://www.praat.org/>

Laprie, Y. WinSnoori (mise à jour : 2010). [www.loria.fr/~laprie/WinSnoori/](http://www.loria.fr/~laprie/WinSnoori/)

Martin, P. WinPitch (mise à jour : 2005). <http://www.winpitch.com>

UCL Division of Psychology and Language Sciences. RTGram (mise à jour : 2010).  
[www.phon.ucl.ac.uk/resource/sfs/rtgram/](http://www.phon.ucl.ac.uk/resource/sfs/rtgram/)

Psychology Software Tools. E-Prime 2.0. <http://www.pstnet.com/eprime.cfm>

Statistical Package for the Social Sciences (SPSS) (mise à jour: 2008).  
<http://www-01.ibm.com/software/fr/analytics/spss/>

**INDEX DES AUTEURS ET DES CONCEPTS**

acoustique 7, 23, 24, 34, 35, 36, 37, 41, 43, 45, 50, 51, 149, 151, 152, 153, 162, 164, 172, 199, 200, 203, 204, 206, 208, 209, 214, 216, 217, 219, 238, 243, 245, 267, 271, 272, 273, 276	Callon..... 48
acquisition 3, 12, 17, 18, 19, 46, 53, 57, 59, 60, 62, 67, 68, 74, 75, 78, 80, 81, 88, 96, 268, 269	Campos ..... 51
apprenant 2, 53, 63, 64, 66, 70, 73, 74, 80, 81, 88, 89, 90, 91, 94, 101, 149, 155, 194, 205, 214, 225, 229, 234, 236, 237, 239, 242, 243, 244, 246, 247, 252, 253, 254, 258, 262, 263, 264, 276, 279, 280	Caplan ..... 20
apprentissage 1, 2, 3, 5, 17, 18, 19, 25, 26, 27, 28, 46, 51, 53, 55, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 72, 73, 76, 77, 81, 86, 88, 89, 90, 92, 94, 95, 96, 99, 101, 102, 106, 110, 111, 112, 130, 132, 133, 136, 138, 140, 144, 145, 152, 158, 159, 161, 162, 205, 206, 207, 210, 211, 212, 213, 217, 221, 222, 226, 230, 232, 254, 260, 263, 267, 269, 274, 276	Carignan..... 219
approche par tâche.....46, 74, 97, 210	Carruthers ..... 16
archiphonème 206, 207, 251, 252, 255, 257, 258, 259, 263, 264, 266, 268	Chan..... 199
Auroux.....8	Charlemagne ..... 215
AXB 5, 149, 151, 153, 154, 162, 163, 164, 165, 166, 167, 168, 169, 170, 172, 173, 174, 183, 194, 199, 203, 271, 276, 280	Chistovich..... 209
Baddeley.....56	Chobert ..... 25, 45, 96, 149, 152, 164
Barth .....67, 80, 205, 210	Chomsky ..... 60, 61
Bei .....151, 200	Chouard..... 35
Bernardi.....92, 149, 174	Clark ..... 17
Berthoz .....46, 54, 55, 62	Claverie..... 46, 47, 48, 49, 102, 278, 279
Best.....67, 76, 164, 165, 199	cognition ..... 47, 51, 56, 60, 61, 98, 102
Bigand .....7, 27, 152, 273	Collet..... 167
Billières .....208	communicatif ..... 50, 67, 72, 73, 90, 207
Blanc.....81	compétence ..... 72, 73, 74, 246
Bourguignon.....48	complexité 11, 13, 18, 27, 46, 49, 52, 54, 60, 62, 64, 97, 152, 269, 278
Bransford.....56	Conein..... 34
Brief.....60	Corbel ..... 93, 139
Broca .....17, 21, 23, 85	Cornaz..... 90
Buccino.....85, 86	Crepault..... 106
Bygate.....151, 200	crible ..... 75, 76, 97, 151
Caclin .....38	Cross ..... 15
Callamand.....39, 208	Damasio ..... 32
	Darsel..... 32
	Darwin ..... 15
	Daumas ..... 106
	Davies ..... 52
	Deacon ..... 17
	Dehaene ..... 55, 101, 144
	Demaizière ..... 64
	dénativisation/accommodation 13, 77, 264, 275
	Deonna..... 32
	Détrie ..... 107
	Deutsch ..... 7, 29
	diachronie ..... 11, 6, 97
	didactique 3, 4, 5, 6, 39, 45, 46, 47, 50, 51, 52, 57, 60, 62, 66, 67, 68, 72, 74, 78, 81, 87, 88, 90, 93, 95, 97, 98, 100, 102, 148, 152, 153, 165, 172, 175, 200, 204, 205, 207, 210, 220, 222, 238, 247, 252, 260, 264, 267, 269, 270, 272, 273, 274, 275, 276, 277, 278, 279, 280

didactique de la prononciation 3, 4, 6, 45, 67, 68, 78, 87, 88, 90, 93, 94, 95, 148, 247, 264, 270, 273, 275, 277, 278	Fastl..... 215
didactique des langues 3, 6, 39, 46, 60, 62, 66, 72, 90, 93, 94, 97, 98, 100, 152, 153, 165, 175, 204, 207, 269, 270, 272, 274, 276, 279, 280	filtre..... 32, 37, 38, 75, 208
Diehl.....83, 84	Fitch ..... 18
discours 8, 16, 17, 54, 70, 71, 74, 96, 100, 137, 138, 139, 140, 218, 269, 276	FLE/S 2, 3, 4, 5, 45, 66, 67, 68, 78, 81, 87, 88, 90, 91, 94, 96, 204, 207, 210, 270, 271, 276, 277, 278
discrimination 5, 149, 151, 153, 154, 158, 162, 163, 164, 165, 173, 174, 183, 187, 188, 194, 195, 199, 202, 203, 204, 221, 271, 273, 280	Flege ..... 67, 76, 144, 164, 267
dispositif 3, 5, 62, 63, 64, 65, 204, 205, 206, 207, 210, 211, 212, 213, 214, 217, 221, 222, 226, 230, 232, 238, 254, 260, 263, 267, 269, 274, 276	Fodor..... 61
Doudin.....107	Fonagy ..... 31
Dubé .....92	Fowler..... 84
Dunbar .....13	français 1, 3, 5, 6, 7, 39, 40, 41, 42, 43, 44, 45, 54, 66, 67, 68, 77, 78, 79, 80, 81, 87, 88, 90, 94, 96, 99, 101, 111, 112, 136, 149, 150, 151, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 167, 168, 169, 170, 171, 172, 173, 174, 175, 183, 184, 185, 186, 187, 189, 194, 195, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 230, 231, 232, 234, 236, 237, 238, 243, 246, 247, 248, 249, 250, 251, 252, 253, 260, 263, 264, 267, 271, 272, 273, 274, 275, 276, 278
écrit 71, 100, 116, 117, 120, 121, 124, 137, 138, 142, 143, 144, 145, 156, 157, 174, 211	Frith..... 32, 55
Elbert.....26	Fujisaki ..... 83, 84
émergentisme 6, 46, 52, 58, 59, 60, 62, 65, 66	Gardner ..... 16, 28
émergentiste 4, 52, 57, 58, 59, 62, 64, 66, 74, 75, 96, 97, 199, 204, 219, 270, 278,	Gazzaniga ..... 21
émotion 7, 14, 21, 22, 30, 31, 32, 92, 98, 99, 102, 118, 119, 120, 121, 122, 123, 135, 136, 139, 144, 152, 173, 177, 178, 183, 218, 278, 279	Gelman..... 183
Enard .....15	Gendrot 7, 42, 43, 44, 169, 170, 171, 205, 209, 216, 217, 239
enregistrement 170, 207, 214, 223, 236, 239, 240, 241, 242, 243, 244, 245, 247, 248, 249, 257, 258, 262, 266, 268, 276,	Geoffroy..... 89
enseignant 3, 88, 93, 95, 161, 207, 220, 222, 230, 263, 280	Georgeton..... 7, 42, 43, 44, 216
enseignement 1, 2, 6, 46, 66, 67, 68, 79, 81, 88, 89, 90, 102, 108, 136, 155, 207, 222, 267, 279	Gerrits ..... 163, 164
épistémologie .....9, 60	Geschwind ..... 21
épistémologique ..1, 4, 52, 62, 68, 158, 278	Gladwell..... 151, 202
Ernst .....55, 56	Gleick..... 53
extrême 43, 113, 206, 207, 250, 251, 253, 254, 256, 257, 258, 259, 263, 264, 266, 268	Goody ..... 70
Fadiga.....86	Gordon ..... 209
Fant.....37, 208	Graham ..... 68, 90
	Grahn ..... 7, 24
	Green..... 83, 84
	Guimbretière 68, 71, 72, 88, 89, 90, 207, 208, 210
	Guion ..... 76
	Hallé..... 164, 165, 199
	Hardison..... 83
	Hébert ..... 20, 287
	Hesling..... 21
	Hickok..... 23

- Hoch .....24, 25  
Hombert.....8, 14, 209  
Honikman .....67, 77, 78, 148  
Huhtanen .....68, 94  
Hunt.....10  
Hyde .....27  
Hyman .....70  
Iacoboni .....86  
identification 162, 164, 171, 172, 205, 206,  
221, 246, 247, 248, 249, 250, 251, 252,  
253, 257, 258, 263, 264, 266, 267, 274,  
275  
incertitude.....9, 3, 57, 279  
interventionniste.....97  
Iverson.....76  
Jalilzadeh.....59  
Jodelet.....101  
Jones .....7, 43, 70, 101  
Joseph.....278  
Kamiyama 43, 162, 163, 164, 214, 246, 267  
Kim.....209  
Kirby.....14  
Kivy.....93  
Krashen.....70, 101  
Kraus .....25, 27, 96, 149, 152, 164, 201  
Kris.....101  
Kuhl.....67, 76, 144  
Kutas.....25  
Lai.....15  
langage 6, 8, 9, 10, 11, 12, 13, 14, 15, 16,  
17, 18, 19, 20, 22, 23, 24, 25, 34, 45, 48,  
50, 52, 53, 54, 55, 59, 61, 66, 72, 92, 98,  
99, 100, 101, 102, 105, 113, 130, 132,  
134, 136, 138, 143, 146, 147, 148, 152,  
207, 218, 278  
langue 1, 6, 8, 29, 34, 37, 39, 43, 46, 53, 54,  
59, 60, 62, 63, 67, 68, 70, 71, 72, 73, 74,  
75, 76, 77, 78, 79, 81, 86, 87, 88, 89, 90,  
91, 98, 99, 100, 101, 102, 103, 111, 112,  
113, 118, 119, 120, 121, 122, 123, 124,  
127, 128, 129, 130, 135, 136, 137, 138,  
139, 140, 141, 142, 143, 144, 145, 146,  
147, 148, 154, 155, 156, 157, 159, 161,  
163, 168, 199, 201, 205, 210, 213, 218,  
226, 254, 271, 274, 275, 276, 279  
langue étrangère 1, 53, 67, 68, 71, 73, 76, 78,  
79, 81, 98, 99, 101, 102, 103, 111, 112, 113,  
118, 119, 120, 121, 123, 124, 130, 135, 136,  
137, 139, 140, 147, 154, 156, 157, 159, 161,  
199, 205, 218, 271, 275  
Larsen-Freeman ..... 46, 52, 53, 54, 94, 269  
Lashley..... 172  
Lauret41, 67, 71, 72, 75, 88, 90, 91, 101,  
208, 275  
Legallois ..... 12  
Legrain..... 103, 104  
Leipp..... 7, 33, 34, 35, 37  
Léon ..... 44  
Levitin..... 15, 26, 28, 130  
Levitsky ..... 21  
Lhote ..... 78  
Li .....8, 14  
Lieberman..... 11, 67, 83  
Lindbom..... 80  
linguistique appliquée..... 97, 153, 270  
Llorca..... 68, 90  
locuteur-plurilingue 12, 14, 98, 99, 100,  
104, 105, 108, 110, 111, 112, 113, 118,  
119, 121, 124, 127, 128, 130, 136, 138,  
139, 140, 141, 142, 143, 144, 145, 146,  
147, 148  
locuteur-signeur 12, 14, 98, 99, 100, 104,  
105, 108, 112, 113, 122, 123, 124, 129,  
130, 143, 144, 145, 146, 147  
Ludke ..... 68, 93  
Luzzati ..... 80  
Macaire ..... 72  
MacLarnon..... 11  
macro-tâche..... 63  
MacWhinney..... 59  
Maddieson..... 255, 264  
Malinen ..... 68, 94  
Mandelbaum ..... 107  
Marler ..... 9  
Marschark ..... 61  
Martinet..... 70  
Massaro..... 83, 164  
McGurk..... 11, 67, 82, 83, 84, 87  
médiateur ..... 9, 64, 210, 279  
médiation 2, 4, 64, 67, 74, 75, 78, 81, 88,  
96, 97, 98, 101, 102, 152, 204, 205, 206,  
210, 218, 238, 273  
Memmi..... 52  
Meunier..... 41, 42, 43  
micro-tâche ..... 63, 210  
Milgrom ..... 48  
Miller ..... 15  
Miras ..... 63, 64, 66, 74, 148  
Mithen 6, 9, 10, 11, 12, 13, 14, 15, 16, 20,  
96

- Mohanan.....54  
 Montagu .....219  
 Morin .....48, 52, 54, 57, 279  
 Morley .....16  
 Mucchielli.....103  
 Münte .....26  
 musicien 14, 51, 99, 114, 115, 116, 117,  
 131, 132, 133, 134, 138, 139, 140, 146,  
 158, 178, 183, 187, 188, 189, 190, 192,  
 194, 195, 198, 203, 213, 275  
 musicologie .....50, 51, 269  
 musique 2, 3, 4, 6, 7, 15, 16, 17, 18, 19, 20,  
 21, 22, 23, 24, 25, 26, 27, 28, 30, 32, 34,  
 37, 40, 45, 46, 47, 50, 51, 66, 67, 68, 69,  
 87, 88, 89, 90, 92, 93, 94, 96, 97, 98, 99,  
 102, 103, 108, 110, 114, 115, 116, 117,  
 124, 125, 126, 130, 131, 132,133, 134,  
 138, 139, 140, 141, 145, 146, 148, 149,  
 150, 151, 152, 153, 154, 155, 157, 158,  
 160, 161, 162, 173, 175, 178, 179, 182,  
 183, 185, 186, 187, 189, 194, 195, 201,  
 203, 204, 205, 206, 207, 209, 213, 215,  
 237, 244, 253, 260, 262, 263, 267, 268,  
 270, 271, 272, 274, 275, 276, 277, 278,  
 280  
 musique instrumentale 3, 4, 5, 6, 67, 89, 90,  
 93, 96, 149, 150, 151, 153, 154, 155, 173,  
 175, 183, 185, 186, 187, 189, 194, 195,  
 201, 203, 204, 205, 272  
 Näätänen.....84  
 Nancy-Combes 2, 8, 46, 62, 63, 64, 66, 72,  
 74  
 nativisation/assimilation 77, 96, 151, 153,  
 199, 203, 204, 220, 253, 264, 273  
 Newman .....144  
 Nguyen .....83, 84  
 Niemitz .....11  
 non-musicien 3, 5, 27, 32, 86, 92, 149, 150,  
 151, 152, 153, 155, 157, 161, 174, 175,  
 177, 178, 183, 187, 188, 189, 192, 193,  
 194, 195, 198, 199, 200, 201, 203, 204,  
 213, 271, 272, 280  
 Norrix .....83  
 note (de musique) 21, 22, 28, 34, 37, 205,  
 207, 209, 229, 232, 244, 249, 261  
 Noyau .....110  
 Ollat.....56  
 oral 18, 69, 70, 77, 80, 93, 94, 99, 100, 101,  
 113, 118, 119, 120, 121, 122, 123, 124,  
 127, 128, 129, 135, 136, 137, 140, 142,  
 143, 145, 146, 147, 156, 157, 211, 218,  
 222, 223, 225, 227, 230, 232, 233, 260,  
 263, 268  
 Pamula .....90  
 panel 98, 99, 103, 104, 106, 107, 108, 109,  
 113, 114, 115, 116, 117, 118, 119, 120,  
 121, 122, 123, 124, 125, 126, 127, 128,  
 129, 138, 143, 144, 146, 162, 213, 249,  
 266, 267  
 Pantev .....26  
 parole 1, 3, 4, 6, 7, 8, 11, 15, 17, 18, 19, 21,  
 23, 24, 25, 26, 28, 30, 32, 35, 37, 38, 40,  
 43, 44, 45, 46, 47, 50, 51, 54, 61, 66, 68,  
 70, 72, 74, 75, 76, 80, 81, 82, 83, 84, 85,  
 86, 88, 96, 97, 98, 100, 101, 107, 143,  
 146, 148, 149, 152, 154, 155, 162, 170,  
 173, 175, 176, 177, 183, 185, 186, 201,  
 203, 209, 217, 218, 219, 269, 270, 271,  
 272, 273, 274, 275, 276, 277, 278, 279,  
 280  
 Pastore.....163  
 Patel 7, 9, 15, 16, 17, 18, 19, 20, 24, 34, 96,  
 perception 2, 4, 18, 20, 22, 25, 27, 30, 33,  
 40, 55, 56, 57, 67, 72, 74, 75, 76, 78, 81,  
 82, 83, 84, 87, 88, 89, 91, 94, 96, 98, 101,  
 102, 120, 121, 124, 140, 142, 145, 148,  
 149, 152, 153, 162, 163, 164, 177, 184,  
 199, 200, 202, 204, 209, 210, 214, 217,  
 221, 266, 271, 276, 280  
 Peretz .. 7, 20, 22, 24, 25, 45, 96, 149, 152,  
 performance 1, 29, 74, 93, 100, 141, 148,  
 235, 274  
 Pernon .....16  
 Perrot.....30  
 phonation .....37, 81, 90  
 phonème 37, 39, 41, 70, 86, 89, 144, 171,  
 172, 216, 217, 220, 242, 244, 248, 249,  
 250, 264  
 phonétique 1, 3, 7, 23, 37, 40, 41, 44, 45,  
 47, 50, 59, 69, 74, 77, 79, 97, 153, 161,  
 162, 163, 164, 165, 172, 175, 199, 201,  
 204, 211, 213, 214, 218, 219, 220, 236,  
 238, 246, 247, 265, 267, 269, 274, 276,  
 279

- phonologie 47, 50, 52, 59, 72, 85, 97, 153, 161, 165, 172, 175, 211, 218, 246, 247, 264, 266, 269, 274, 276, 279
- Piaget.....46, 47, 60, 61, 75
- Pillot-Loiseau .....9, 81, 90
- Pinker .....17
- Pisoni.....84, 149, 163, 164, 165, 189
- pluri-inter-transdisciplinarité 4, 6, 46, 47, 48, 49, 50, 51, 54, 66, 96, 97, 102, 153, 204, 269, 270
- Porée.....92
- Posner.....165
- production 2, 4, 16, 18, 20, 33, 37, 45, 67, 72, 74, 75, 77, 79, 81, 82, 83, 84, 87, 88, 90, 93, 94, 96, 98, 100, 101, 102, 120, 121, 124, 138, 140, 145, 148, 204, 205, 206, 207, 210, 211, 214, 217, 219, 221, 226, 238, 239, 240, 241, 244, 248, 249, 250, 251, 252, 253, 256, 257, 258, 259, 262, 263, 264, 266, 267, 268, 269, 273, 274, 275, 276, 278, 279, 280
- production orale.....4, 90
- prononciation 1, 4, 46, 64, 66, 67, 68, 70, 72, 73, 74, 77, 78, 79, 81, 88, 90, 91, 93, 95, 97, 101, 102, 118, 119, 120, 121, 122, 123, 124, 135, 137, 142, 143, 146, 148, 152, 156, 157, 205, 207, 210, 211, 212, 213, 217, 218, 222, 223, 224, 225, 226, 227, 230, 231, 232, 233, 234, 235, 237, 238, 260, 263, 265, 267, 269, 273, 275, 276, 279
- prosodie 2, 16, 22, 37, 45, 73, 81, 209, 222, Protopapas .....267
- psychoacoustique 3, 12, 5, 98, 149, 203, 277
- psychoaffectif .....244, 273
- psychologie..... 50, 70, 101, 202, 285
- questionnaire 3, 13, 103, 105, 106, 113, 114, 115, 116, 118, 119, 120, 122, 123, 124, 125, 126, 127, 128, 129, 131, 132, 133, 135, 136, 137, 139, 142, 143, 146, 148, 149, 153, 154, 155, 156, 174, 175, 177, 181, 184, 185, 202, 205, 221, 222, 223, 224, 225, 226, 229, 230, 232, 233, 234, 237, 238, 243, 244, 260, 262, 268, 277
- Randall .....70, 101
- Regan.....110
- Reitan .....94
- Renard .....7, 36, 72, 75
- représentation 24, 26, 28, 38, 39, 47, 49, 57, 64, 78, 81, 89, 101, 109, 110, 113, 186, 203, 208, 219, 230, 266, 274
- Rialland..... 210
- Rizzolatti..... 67, 85, 86, 87, 283, 285, 288
- Rolland..... 60, 71, 72, 75, 101, 106
- Rosenthal ..... 28
- Rutherford..... 53
- Sapir..... 61, 107
- Saussure ..... 8, 48, 69
- Savouret ..... 95
- Schaeffer ..... 37
- Schiffmann..... 26
- Schneider ..... 189
- Schneuwly..... 61
- Schön ..... 21
- Schrenk ..... 11
- Schuh ..... 255
- Schwartz ..... 44
- sciences du langage..... 50, 51, 60, 278
- Scotto Di Carlo ..... 75
- Seron ..... 1
- Shaw ..... 47
- Shiller..... 80
- Shors ..... 92
- Simmonds ..... 23, 27, 144
- Simondon..... 8
- simplicité..... 11, 46, 52, 54, 55, 62
- Singleton ..... 101
- socioconstructivisme..... 4, 6, 46, 60, 61
- socioconstructiviste 3, 62, 66, 74, 75, 96, 97, 199, 270, 278
- statistique 50, 82, 150, 184, 192, 194, 205, 224, 233, 239, 240, 242, 262, 263, 272
- Stephan ..... 58
- stimuli 12, 23, 25, 28, 30, 56, 84, 149, 153, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 200, 206, 216, 217, 239, 246, 247, 262
- stimulus 60, 145, 163, 171, 173, 189, 198, 206, 239, 241
- Strange ..... 267
- synchronie..... 11, 6, 8, 16, 96, 97
- tâche sociale..... 63, 64, 269
- Tantibundhit..... 189, 199
- taux de discrimination 5, 150, 151, 153, 183, 187, 188, 194, 195, 198, 199, 200, 201, 202, 203, 204, 271, 272, 273, 276,



temps de réaction 5, 150, 151, 153, 162, 165, 183, 189, 190, 191, 193, 194, 195, 196, 197, 198, 200, 201, 202, 203, 272, 273, 280	verbo-tonal..... 67, 72, 275
Terrien .....68, 94	Verspoor..... 53, 54
Thaut.....14	Vidal..... 106
théorie du chaos.....94	voix 34, 38, 75, 88, 90, 91, 109, 173, 176, 200, 207, 257
Tiberghien.....57	Vos ..... 92
Tillmann .....7, 27	Vygotski ..... 46, 61, 101
Tomasello .....17	Wachs..... 9, 67, 71, 72
Tripier-Mondancin .....108	Watson ..... 27
Troubetzkoy.....69, 75	Werker..... 23, 162, 201
Tsukada.....199	Wernicke ..... 21, 23
Tubach .....42, 43, 170, 216	Whalen ..... 209
tuteur.....63, 64	Wheeler..... 10, 12
Twaddell .....70	Whorf ..... 61, 107
Tyne .....110	Wilhelm ..... 78
Vaissière 7, 23, 31, 33, 35, 38, 39, 42, 43, 44, 50, 69, 79, 169	Wilson..... 77
Vallejo-Gomez.....54	Wray..... 6, 9, 12, 14, 45
	Zatorre..... 28, 130
	Zwicker ..... 215

## LISTE DES FIGURES

- Figure 1:** Classification de Homo sapiens dans l'ordre des primates (Britannica en ligne, 2013) ..... 9
- Figure 2:** Une représentation modulaire de la musique (Peretz & Coltheart, 2003 : 690). Chaque rectangle représente un bloc processuel et chaque flèche un flux d'information. Les blocs spécifiques à la musique apparaissent en noir foncé et les autres en noir/gris clair. .... 24
- Figure 3:** Pourcentage de réponses correctes pour le test AP en fonction de l'âge de premier entraînement musical et le niveau de performance à parler une langue à ton. « Les individus des groupes “tone very fluent” (parle couramment une langue à ton), “tone fairly fluent” (parle plutôt couramment une langue à ton) et “tone non-fluent” (parle une langue à ton mais de manière non courante) étaient tous d'origine d'Asie de l'Est. Les individus du groupe “nontone” (ne parle pas de langue à ton) étaient tous caucasiens et ne parlaient aucune langue à ton. La courbe “chance” (hasard) représente une réponse au hasard à la tâche. » (Deutsch & Dooley, 2009 : 2400)..... 29
- Figure 4:** les deux voies de la perception auditive (med.univ-montp1, 2013)..... 30
- Figure 5:** Intégration comportementale de l'information auditive (Cochlea.org, 2013). .... 31
- Figure 6:** Le double encodage de la communication verbale (Fonagy, 1983 : 14)..... 31
- Figure 7:** Calcul de la fréquence d'une onde (bbemg.ulg.ac, 2012)..... 34
- Figure 8:** Représentation d'une onde périodique simple (à gauche) et d'une onde apériodique (à droite) en fonction de l'amplitude (en ordonnée) et du temps (en abscisse) (Ciral.ulaval.ca en ligne, 2012). .... 35
- Figure 9:** Représentation d'une onde complexe et ses harmoniques (Renard, 1978 : 6). .... 36
- Figure 10:** Représentation de la répartition tonotopique de la cochlée. Les sons sont traités dans la cochlée des fréquences aiguës vers les graves respectivement de la base vers l'apex de la membrane basilaire (Encyclopaedia Britannica en ligne, 1997)..... 36
- Figure 11:** Une représentation du timbre musical (McAdams et al. 1995 dans Caclin, 2004 : diapositive 2). Le timbre musical se compose d'au moins trois dimensions que sont le temps d'attaque (court/long), le centroïde spectral (aigu/grave) et le flux spectral (plus/moins). .... 38
- Figure 12:** Représentation spectrale d'une voyelle. Analyse formantique du son vocalique isolé /ø/ avec le logiciel Praat. Les 4 premiers formants ont été calculés automatiquement par le programme dans la représentation de gauche. Le [ø] du français standard se caractérise par quatre formants (plus ou moins) équidistants..... 39
- Figure 13:** Représentation spectrale de trois voyelles isolés /i y u/ avec le logiciel Praat. Les rectangles représentent les renforcements d'énergie. Le /i/ est caractérisé par un renforcement d'énergie dans les hautes fréquences, le /y/ dans les moyennes fréquences et le /u/ dans les basses fréquences. .... 40
- Figure 14:** Représentation du trapèze général vocalique (à droite). Les phonèmes vocaliques du français sont entourés (Meunier, 2007 : 2)..... 41
- Figure 15:** Réalisations acoustiques des voyelles /i/, /e/, /a/, /o/, /u/ prononcées isolément (à gauche) et dans un texte (à droite). On observe la réduction considérable de la taille globale du système vocalique; les valeurs moyennes de formants observées en lecture de texte sont très éloignées des valeurs prototypiques (Meunier, 2007 : 9). .... 42
- Figure 16:** Comparaison des triangles vocaliques sur le plan F1-F2 de Gendrot & Adda (2005) (en noir), Calliope (Tubach, 1989) (en bleu) et Groupe Didactique (GD) (Georgeton et al., 2012) (en rouge) à gauche. Valeurs des formants sur un plan F1/ F2 et F2/ F3 de GD (à droite) (Georgeton et al., 2012 : 150). .... 42
- Figure 17:** Voyelles cardinales primaires (à gauche) et secondaires (à droite) telles que définies par Jones (Vaissière, 2011 : 53)..... 43

<b>Figure 18:</b> Représentation spectrographique des six voyelles focales du français. Les renforcements d'énergie sont signalés par des crochets et les formants impliqués inscrits à côté. Le /i/ montre une focalisation F3F4, le /y/ F2F3 et /u o o α/ F1F2.....	44
<b>Figure 19:</b> Deux modélisations (A et B) de l'interdisciplinarité, multidisciplinarité et transdisciplinarité (Della Chiesa, 2007 : 148-149). .....	49
<b>Figure 20:</b> La boucle perception-action dans un cadre Bayésien (Ernst et Bühlhoff, 2004 : 164). Dans un cadre bayésien, la perception émerge à travers l'organisation d'un ou plusieurs stimuli, émis par l'environnement, traités à travers des règles bayésiennes nécessitant un traitement sensoriel et une connaissance a priori. A partir de ce bit d'information, une décision sera prise provoquant une action ou non. Le choix pourra être motivé par une fonction gain/perte dont le poids dépendra du but à atteindre. L(a non-)action ainsi produite pourra avoir un impact sur l'environnement lui-même à travers ses interactions avec l'acteur. ....	56
<b>Figure 21:</b> Décomposition structuraliste des systèmes (par l'auteur). .....	57
<b>Figure 22:</b> Recomposition émergentiste des systèmes (par l'auteur). .....	58
<b>Figure 23:</b> Dispositif par tâche (d'après Narcy-Combes & Miras, 2013 : 40). La relation tuteur-apprenant(s) (T/A(s)) est au centre du système d'apprentissage où gravitent des tâches sociales (ou macro-tâche : Ma) et un suivi/évaluation (Suivi Eval). Chaque tâche sociale peut motiver le déclenchement d'une ou plusieurs tâches d'entraînement (ou micro-tâche : Mi). Le suivi/évaluation pourra également intégrer des micro-étapes non modélisées ici. ....	63
<b>Figure 24:</b> Dispositif global sous le prisme de l'émergentisme (par l'auteur). Le dispositif évolue dans le temps et à travers les pressions environnementales et les objectifs. Ces changements sont stimulés et stimulent des niveaux inférieurs qu'ils soient biologiques, cognitifs, sociaux et/ou liés aux mécanismes d'apprentissage. Les changements sont imprévisibles car s'effectuent de manière non-linéaire. ....	65
<b>Figure 25:</b> La double réalité des unités sonores verbales (par l'auteur). .....	70
<b>Figure 26:</b> Représentation des rapports entre les notions de processus, compétence et performance (Miras & Narcy-Combes, 2014 : 19). .....	74
<b>Figure 27:</b> Modélisation de l'effet d'aimant perceptif. Les sons perceptivement proches (A) sont attirés par le son prototypique le plus proche (B) (Kuhl, 1993 dans Kuhl et Iverson, 1995 : 124). .....	76
<b>Figure 28:</b> Représentation spectrographique des quatre voyelles antérieures du français standard. L'opposition /e/-/ɛ/ est réalisée par la double stratégie : une augmentation de F1 et un abaissement de F2 et F3. Calcul formantique automatique par Praat. ....	79
<b>Figure 29:</b> Cochléogramme des onze voyelles orales du français standard. Représentation automatique par Praat. En encadré orange, les renforcements d'énergie les plus saillants pour ce locuteur avec atténuation quasi-totale des autres fréquences et en encadré jaune, les renforcements d'énergie avec une atténuation moindre des autres fréquences. ....	80
<b>Figure 30:</b> Aires corticales activées durant l'observation d'actes communicatifs oraux (d'après Buccino et al., 2004 dans Rizzolatti & Sinigaglia, 2006 : figure 5.10). .....	86
<b>Figure 31:</b> Potentiels évoqués moteurs enregistrés par les muscles de la langue durant l'écoute de matériaux verbaux et de sons bitonaux. (Rizzolatti et Sinigaglia, 2006 : 180). .....	87
<b>Figure 32:</b> Représentation corporelle des sons vocaliques [i] et [u] selon le bonhomme sonore (Guimbretière & Kaneman-Pougatch, 1991 : 16). .....	89
<b>Figure 33:</b> Fréquences dominantes des langues selon la méthode SpeedLingua et d'après le principe Tomatis. ....	91
<b>Figure 34:</b> Répartition quantitative des cinq panels de l'enquête. Les sigles signifient respectivement musiciens sans cours de solfège (MSS), musiciens avec cours de solfège (MAS), locuteurs-plurilingues avec pratique régulière en contexte naturel (LPR), locuteurs-plurilingues sans pratique régulière en contexte naturel (LPNR) et locuteurs-signeurs (LS). ....	105

<b>Figure 35</b> : Continuum de prise de conscience entre ce qui est produit et ce qui est perçu chez les 4 panels étudiés. ....	141
<b>Figure 36</b> : Déroulement des trois phrases principales du test général.....	154
<b>Figure 37</b> : Mesure de la progression entre les tests 1 et 2 afin de mettre en évidence la progression intra-individuelle.....	154
<b>Figure 38</b> : Les différents groupes d'auditeurs retenus pour l'étude. Les tâches sont une exposition à un document visuel (Visuelle) ou de parole continue en français (Parole FR) ou de musique instrumentale (Musique). ....	155
<b>Figure 39</b> : Les trois types d'expérience en perception phonético-phonétique avec mise en avant des différentes cibles, contextes et mesures possibles (Kamiyama, 2009 : 109).....	163
<b>Figure 40</b> : Déroulement du test AXB avec mise en avant des intervalles interstimulus et inter-trial. ....	164
<b>Figure 41</b> : Déroulement des tests AXB 1 et 2 en cinq étapes selon l'affichage E-Prime. ....	165
<b>Figure 42</b> : Déroulement des phases d'entraînement et de test selon l'affichage E-Prime. La phase test comprend sept étapes et celle d'entraînement, une de plus soit le « feedback ». ....	166
<b>Figure 43</b> : Placement des touches de réponse sur le clavier d'ordinateur pour les tests AXB. La pastille bleue est sur la touche « CTRL » et la verte sur « flèche droite ». ....	166
<b>Figure 44</b> : Affichage des feedbacks positif (à gauche) et négatif (à droite) au test AXB.....	167
<b>Figure 45</b> : Triangle vocalique du français standard. Les sept paires de la phase test du test AXB sont entourées. En abscisse, l'antériorité/postériorité et en ordonnée, l'aperture. ....	168
<b>Figure 46</b> : Représentation spectrale sur Praat des dix voyelles orales du français standard enregistrées à l'isolée par LOC1 et utilisées comme stimuli pour les tests AXB. En abscisse, le temps et en ordonnée, la fréquence (en Hz). ....	169
<b>Figure 47</b> : Analyse formantique (formants F1 à F2) des stimuli du test AXB. Les valeurs sont présentées sous forme de tableau (à gauche) et dans un graphique en nuage de points (à droite). Le F4 des quatre voyelles postérieures et le /a/ n'a pas pu être déterminé. Valeurs en Hz. ....	169
<b>Figure 48</b> : Triangle vocalique de LOC1 d'après l'analyse formantique des stimuli du test AXB. Le premier formant (F1) est représenté en ordonnée et le deuxième formant (F2) en abscisse (en Hz). ....	170
<b>Figure 49</b> : Déroulement de la phase tâche en trois étapes.....	174
<b>Figure 50</b> : Document 4 visuel (DV4).....	176
<b>Figure 51</b> : Document 5 visuel (DV5).....	177
<b>Figure 52</b> : Moyennes et écart-types évalués en fonction de chaque groupe pour le tempo perçu après exposition à l'un des trois documents de la tâche. 1 = tempo très lent : 5 = tempo très vif. Les barres d'erreur indiquent l'erreur standard. ....	184
<b>Figure 53</b> : Carte heuristique des premiers mots donnés par chaque sujet en fonction des quatre groupes (visuel, parole en français, musique instrumentale : apprenants et locuteurs natifs du français). Ce type de représentation permet mettre en lumière les liens qui existent entre un concept ou une idée, et les informations qui leur sont associées. ....	186
<b>Figure 54</b> : Diagramme du taux de discrimination moyen (en %) en fonction du groupe, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard. ....	187
<b>Figure 55</b> : Impact du statut musical sur le taux de discrimination moyen (en %). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,24)=4,015$ , $p=0,057$ ; $F(1,6)=64,751$ , $p=0,036$ ). Les barres d'erreur indiquent l'erreur standard. ....	188
<b>Figure 56</b> : Diagramme du temps de réaction moyen (en ms) en fonction du groupe, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard.....	189
<b>Figure 57</b> : Impact du statut musical sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,24)=4,381$ , $p=0,047$ ; $F(1,6)=324,149$ , $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard. ....	190

<b>Figure 58</b> : Impact du Test sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,24)=12,284$ , $p=0,002$ ; $F(1,6)=233,315$ , $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard. ....	191
<b>Figure 59</b> : Impact de l'interaction Statut X Test sur le temps de réaction moyen (en ms). La différence est marginalement significative d'après une analyse de variance à mesures répétées ( $F(1,24)=3,297$ , $p=0,082$ ; $F(1,6)=16,637$ , $p=0,007$ ). Les barres d'erreur indiquent l'erreur standard. ....	191
<b>Figure 60</b> : Impact de l'interaction Groupe X Test sur le temps de réaction moyen (en ms). La différence n'est pas significative pour F1 d'après une analyse de variance à mesures répétées ( $F(1,24)=0,197$ , $p=0,0,823$ ; $F(1,307 ; 7,84)=14,198$ , $p=0,004$ ). Les barres d'erreur indiquent l'erreur standard. ....	193
<b>Figure 61</b> : Diagramme du taux de discrimination moyen (en %) en fonction du statut langagier, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard. ....	194
<b>Figure 62</b> : Diagramme du temps de réaction moyen (en %) en fonction du statut langagier, du test et du statut musical. Les barres d'erreur indiquent l'erreur standard. ....	196
<b>Figure 63</b> : Impact du statut sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures ( $F(1,16)=7,242$ , $p=0,016$ ; $F(1,6)=324,990$ , $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard. ....	196
<b>Figure 64</b> : Impact du Test (à gauche) sur le temps de réaction moyen (en ms). La différence est significative d'après une analyse de variance à mesures répétées ( $F(1,16)=22,732$ , $p<0,0001$ ; $F(1,6)=204,656$ , $p<0,0001$ ). Les barres d'erreur indiquent l'erreur standard. ....	197
<b>Figure 65</b> : Impact de l'interaction Statut X Test sur le temps de réaction moyen (en %). La différence est marginalement significative d'après une analyse de variance à mesures répétées ( $F(1,16)=9,691$ , $p=0,007$ ; $F(1,6)=38,221$ , $p=0,001$ ). Les barres d'erreur indiquent l'erreur standard. ....	197
<b>Figure 66</b> : Schéma simplifié du Bonhomme Sonore pour les voyelles /i/, /y/ et /u/ (d'après Guimbretière & Kaneman-Pougatch, 1991 : 70). ....	208
<b>Figure 67</b> : Acuité des sons vocaliques du français standard (Lauret, 2011 : 110 ; d'après Guimbretière & Kaneman-Pougatch, 1991:8) à gauche et (Billières, 2005), à droite. « C » correspond à la caractéristique « clair(C+)/sombre(C-) » et « T » à la tension. ....	208
<b>Figure 68</b> : Corrélations arbitraires proposées entre la hauteur tonale des notes de musique et l'acuité des sons vocaliques postérieurs arrondis du français standard. A gauche : notation musicale ; à droite : spectrogramme (en abscisse : le temps et en ordonnée : la fréquence en Hz (min =0 Hz ; max = 5000 Hz) des quatre voyelles postérieures du français. ....	209
<b>Figure 69</b> : Les trois phases chronologiques du dispositif d'apprentissage .....	211
<b>Figure 70</b> : Fréquence (en Hz) de la F0 des continua d'input auditif pour les voyelles prononcées avec une F0 plate ou montante (à gauche) puis comparée avec des valeurs de référence jouées au piano (à droite). ....	215
<b>Figure 71</b> : Modulation juste perceptible de fréquence en fonction de la fréquence pour des modulations de fréquences sinusoïdales à une fréquence de modulation de 4 Hz. Les lignes pointillées sont des approximations affines de la courbe (d'après Zwicker & Fastl, 2003 :183 dans Charlemagne, 2010 : 35). ....	215
<b>Figure 72</b> : Représentation graphique des valeurs formantiques (en Hz) des stimuli vocaliques utilisés pour les inputs des groupes 2 et 3 et mesurées avec Praat d'après la valeur moyenne de la totalité du phonème. En abscisse les phonèmes et en ordonnée la fréquence (en Hz). ....	216
<b>Figure 73</b> : Trapèze vocalique renversé utilisé pour l'étude. ....	220
<b>Figure 74</b> : Chronologie des trois questionnaires visant à mesurer les représentations des apprenants pendant le dispositif d'apprentissage. ....	222
<b>Figure 75</b> : Graphique représentant les réponses à l'échelle d'évaluation portant sur l'auto-évaluation du niveau de français des apprenants pour les questionnaires 1 et 2. En ordonnée, le nombre de réponses données (max = 27). ....	224

- Figure 76:** Graphique représentant les réponses aux questionnaires 2 et 3 à la question « j'ai perçu des changements dans ma prononciation depuis le début du cours de prononciation » (n=27). En abscisse, les questionnaires 2 et 3 et en ordonnée le nombre de réponses. .... 227
- Figure 77:** Graphique représentant les réponses à l'échelle d'évaluation portant sur l'auto-évaluation du niveau de français des apprenants pour les trois questionnaires. En ordonnée, le nombre de réponses données (max = 27). .... 228
- Figure 78:** Graphique des réponses aux questionnaires 2 et 3 à la question « je pense avoir pratiqué le français à l'oral (en dehors des cours)» (n=27/questionnaire. En abscisse, la fréquence et en ordonnée, le nombre de réponses. .... 230
- Figure 79:** Graphiques représentant les réponses concernant l'impact des parties théorique (à gauche) et pratique en groupe (à droite) sur leur prononciation générale du français (n=27). En abscisse, le degré de réponse (1 correspond à « tout à fait d'accord » et 5 « pas du tout d'accord ») et en ordonnée, le nombre de réponse. .... 231
- Figure 80:** Graphiques et tableaux (n=45) récapitulatifs de la progression globale perçue par les apprenants en fonction des trois groupes concernant les cinq niveaux questionnés au regard des questionnaires 1 & 2 (à gauche) et 1 & 3 (à droite). .... 234
- Figure 81 :** Graphiques et tableaux (n=9/groupe) récapitulatifs de la progression perçue par les apprenants en fonction des trois groupes concernant la prononciation des voyelles au regard des questionnaires 1 & 2 (à gauche) et 1 & 3 (à droite). (-) représente une différence négative donc une régression, (=) neutre donc une stagnation et (+) positive donc une progression entre les évaluations 2 et 3. .... 235
- Figure 82:** Graphiques représentant les réponses de 1 à 5 aux affirmations portant sur les liens entre notes de musique et trapèze vocalique lors du questionnaire 2 (à gauche) et 3 (à droite) n=9/affirmation). .... 237
- Figure 83:** Diagramme des durées moyennes (en ms) des phonèmes des deux continua analysés en fonction des tests 1 et 2 (T1 à gauche et T2 à droite) et des modalités de production (isolé/continuum). .... 241
- Figure 84:** Test d'identification Praat. Consignes (à gauche) et écran de réponse (à droite). .... 246
- Figure 85:** Graphique des degrés de confiance au test d'identification par les 10 locuteurs natifs du français en fonction du phonème et de la modalité de production. Les écarts-types sont indiqués. .... 248
- Figure 86 :** Moyenne d'identification correcte des phonèmes en fonction de la modalité de production selon la grille extrême du français standard. Les erreurs standards sont indiquées. .... 250
- Figure 87 :** Moyenne d'identification correcte des phonèmes en fonction de la modalité de production selon la grille avec neutralisation des archiphonèmes du français contemporain. Les erreurs standards sont indiquées. .... 251
- Figure 88:** Moyenne d'identification correcte des phonèmes en fonction de la grille d'évaluation pour chaque modalité de production. Les erreurs standards sont indiquées. .... 252
- Figure 89:** Apprenant-es concerné-es par au moins une confusion /u-y/ en fonction de la langue première et du groupe expérimental (à gauche) et tableau du pourcentage de confusion en fonction du groupe (à droite). .... 254

## LISTE DES TABLEAUX

<b>Tableau 1:</b> Caractéristiques principales des tâches sociales et d'entraînement. (D'après Bertin et al., 2010, chapitre 11 dans Miras & Narcy-Combes, 2014 : 22).....	63
<b>Tableau 2:</b> Tableau synthétique des apports complémentaires de l'émergentisme et des théories socioconstructivistes (Miras & Narcy-Combes, 2014 : 24). .....	66
<b>Tableau 3:</b> La compétence phonologique par le CECRL (Conseil de l'Europe, 2001 : 91-2). .....	73
<b>Tableau 4:</b> Comparaison des ajustements articulatoires caractéristiques de l'anglais (RP) et du français parisien (traduit en français par Wilhelm, 2012 : 16 d'après Honikman, 1964 : 79).....	78
<b>Tableau 5:</b> Pourcentage de réponses pour chaque catégorie dans les conditions auditivo-visuelles et mise en évidence de l'effet McGurk (McGurk & MacDonald, 1976 : 747).....	82
<b>Tableau 6:</b> Taxonomie des principales approches théoriques de la perception de la parole (traduit par l'auteur d'après Diehl, Lotto & Holt, 2004 : 1558).....	84
<b>Tableau 7:</b> Caractéristiques des 5 questionnaires de l'enquête quantitative selon les sujets auxquels ils s'adressent. ....	102
<b>Tableau 8:</b> Profil socioculturel et professionnel des sondés (n = 1293). Les résultats remarquables sont mis en évidence par un fond foncé et en gras.....	105
<b>Tableau 9:</b> Profil culturel des sondés. (n = 1293). Les résultats remarquables sont mis en évidence par un fond foncé et en gras. NT signifie non-traitable.....	106
<b>Tableau 10 :</b> Marge d'erreur calculée en fonction de la taille d'échantillon du panel. La formule utilisée est : $e = tp1 - pn$ . La population mère est considérée comme infinie car très grande. t correspond au coefficient de marge, calculé à partir d'un taux de confiance de 95% il est égal à 1,96. p correspond à la proportion des éléments de la population-mère et qui est égal à 0.05 car considéré comme inconnu. n correspond à la taille de l'échantillon.....	108
<b>Tableau 11:</b> Instruments joués régulièrement par les musiciens. Les résultats sont donnés pour chaque panel (MSS et MAS) et le total des deux groupes (TT). Les résultats marquants sont mis en évidence par un fond foncé et en gras. ....	109
<b>Tableau 12:</b> Profil spécifique des musiciens. Les résultats sont donnés pour chaque panel (MSS et MAS) et le total des deux groupes (TT). Les résultats marquants sont mis en évidence par un fond foncé et en gras. ....	109
<b>Tableau 13:</b> Profil spécifique des locuteurs-plurilingues. Pourcentages calculés sur un total de 704 répondants pour le total (TT), 573 locuteurs-plurilingues ayant une pratique régulièrement en contexte naturel (LPR) et 131 n'ayant pas une pratique régulièrement en contexte naturel (LPNR). Le pourcentage correspond au pourcentage de sondés pratiquant cette langue. Les résultats marquants sont mis en évidence par un fond foncé. ....	111
<b>Tableau 14:</b> Âge d'apprentissage de la première langue étrangère apprise. Pourcentages calculés sur un total de 704 répondants pour le total (TT), 573 locuteurs-plurilingues ayant une pratique régulièrement en contexte naturel (LPR) et 131 n'ayant pas une pratique régulièrement en contexte naturel (LPNR). Les résultats marquants sont mis en évidence par un fond foncé.....	111
<b>Tableau 15:</b> Profil spécifique des locuteurs-signeurs. Pourcentages calculés sur un total de 120 locuteurs-signeurs à l'exception du contexte de pratique de la langue des signes dont le pourcentage est calculé sur la base totale des réponses soit 183 à la question. Les résultats importants sont mis en évidence par un fond foncé. ....	112
<b>Tableau 16:</b> Exemple de correspondances entre phrases affirmatives dans le questionnaire à échelle d'attitude des 5 panels.....	113
<b>Tableau 17:</b> Les treize phrases affirmatives du questionnaire à échelle d'attitude des musiciens (MAS et MSS). ....	114

<b>Tableau 18:</b> Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel MAS. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ ...	115
<b>Tableau 19:</b> Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel MSS. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ ...	116
<b>Tableau 20:</b> Les 13 phrases affirmatives du questionnaire à échelle d'attitude des locuteurs-plurilingues (LPR/LPNR).	118
<b>Tableau 21:</b> Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel LPR. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ ...	119
<b>Tableau 22:</b> Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel LPNR. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ ).	120
<b>Tableau 23:</b> Les 13 phrases affirmatives du questionnaire à échelle d'attitude des locuteurs-signeurs (LS).	122
<b>Tableau 24:</b> Résultats en pourcentages de réponses au questionnaire à échelle d'attitude du panel LS. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ ..	123
<b>Tableau 25:</b> Exemple de correspondances entre les questions et réponses dans le questionnaire à choix multiples des 5 panels. LO correspondant à langue orale, LS à langue des signes et LE à langue étrangère.	124
<b>Tableau 26:</b> Résultats en pourcentages de réponses au questionnaire à choix multiples du panel MAS. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .....	125
<b>Tableau 27:</b> Résultats en pourcentages de réponses au questionnaire à choix multiples du panel MSS. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .....	126
<b>Tableau 28:</b> Résultats en pourcentages de réponses au questionnaire à choix multiples du panel LPR. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .....	127
<b>Tableau 29:</b> Résultats en pourcentages de réponses au questionnaire à choix multiples du panel LPNR. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .....	128
<b>Tableau 30:</b> Résultats en pourcentages de réponses au questionnaire à choix multiples du panel LS. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .	129
<b>Tableau 31:</b> Résultats synthétiques en pourcentages de réponses au questionnaire à échelle d'attitude des panels MAS et MSS. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ .....	131
<b>Tableau 32:</b> Résultats synthétiques en pourcentages de réponses au questionnaire à choix multiples des panels MAS et MSS. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .	132
<b>Tableau 33:</b> Degré de différence entre les réponses au questionnaire à échelle d'attitude pour les affirmations : « un bon musicien doit seulement reproduire les intentions du compositeur » et « un bon musicien apporte sa personnalité dans l'œuvre qu'il interprète ». Le résultat correspond à la valeur absolue de la différence entre les deux attitudes d'un même sondé. Les données ont été calculées sur la base de « tout à fait d'accord » = 1 ; « plutôt d'accord » = 2 ; « moyenne d'accord » = 3 etc. Les formules utilisées ont été « SI », « NB.SI » et « ABS ». Un degré 2 de différence dans le tableau correspond par exemple à « tout à fait d'accord »-« moyennement d'accord ». 0 correspond à un degré similaire pour les deux affirmations et 4 correspond à deux réponses totalement opposées.	133
<b>Tableau 34:</b> Résultats synthétiques en pourcentages de réponses au questionnaire à échelle d'attitude des panels LPR et LPNR. En gras et surligné gris : pourcentages $X > 30\%$ ; en gras : pourcentages $30\% = \text{ou} > X > 20\%$ .....	135
<b>Tableau 35:</b> Résultats synthétiques en pourcentages de réponses au questionnaire à choix multiples des panels LPR et LPNR. En gras surligné gris : pourcentages $X = \text{ou} > 50\%$ ; en gras : pourcentages $30\% = \text{ou} < X$ .	136



<b>Tableau 36:</b> Degré de différence entre les réponses au questionnaire à échelle d'attitude pour les affirmations : « Ma prononciation est meilleure quand je parle spontanément à l'oral » et « Ma prononciation est meilleure quand je lis un texte ». Le résultat correspond à la valeur absolue de la différence entre les deux attitudes d'un même sondé. Les données ont été calculées sur la base de « tout à fait d'accord » = 1 ; « Plutôt d'accord » = 2 ; « moyenne d'accord » = 3 etc. Les formules utilisées ont été « SI », « NB.SI » et « ABS ». Un degré 2 de différence dans le tableau correspond par exemple à « tout à fait d'accord »-« moyennement d'accord ». 0 correspond à un degré similaire pour les deux affirmations et 4 correspond à deux réponses totalement opposées. ....	137
<b>Tableau 37:</b> Résultats comparatifs en pourcentages de réponses au questionnaire à échelle d'attitude pour deux affirmations sur le rôle du locuteur / musicien. En gras, les pourcentages supérieurs ou égaux à 20% et en gras surligné gris, les pourcentages supérieurs ou égaux à 50%.....	139
<b>Tableau 38:</b> Pourcentage de réponse des différents panels concernant la prise de conscience entre ce qui produit et ce qui est perçu. ....	141
<b>Tableau 39:</b> Résultats comparatifs en pourcentages de réponses au questionnaire à échelle d'attitude pour deux affirmations sur les effets de la timidité sur la qualité de l'oral/de jeu. En gras surligné gris : pourcentages X = ou > 50% ; en gras : pourcentages 30% =ou < X.....	142
<b>Tableau 40:</b> Synthèse des positionnements des quatre panels concernant les liens écrit-son dans un questionnaire à échelle d'attitude. Les signes correspondent à : (-) ou (+) tendance légèrement négative ou positive ; (- -) ou (++) tendance fortement négative ou positive ; (+/-) aucune tendance remarquable mais plutôt positive et (+/-) plutôt négative. ....	143
<b>Tableau 41:</b> Moyennes et écart type obtenues dans chaque groupe sur l'autoévaluation des sujets de leur niveau de français écrit, oral et de prononciation sur une échelle de 1 (très faible) à 7 (excellent). M : musiciens ; NM : non musiciens. ....	157
<b>Tableau 42:</b> Effectifs des apprenants en fonction des institutions dans lesquelles ils suivent une formation de français. M : musiciens ; NM : non musiciens.....	159
<b>Tableau 43 :</b> Effectifs des apprenants en fonction de l'âge d'apprentissage de la première langue étrangère (à gauche) et de premier apprentissage du français (à droite). M : musiciens ; NM : non musiciens.....	159
<b>Tableau 44 :</b> Effectifs des apprenants en fonction de la fréquence d'écoute de musique.....	160
<b>Tableau 45 :</b> Effectifs des apprenants en fonction du nombre d'années de pratique d'un instrument (à gauche), de l'âge de début de pratique (au centre) et la fréquence de pratique (à droite). ....	160
<b>Tableau 46:</b> Répartition des stimuli du test AXB pour la phase d'entraînement. Au total, huit triplets sont organisés de manière semi-aléatoire selon deux oppositions vocaliques et quatre répartitions possibles. ....	166
<b>Tableau 47:</b> Répartition des stimuli du test AXB pour la phase de test. Au total, 112 triplets sont organisés selon sept paires vocaliques, quatre répartitions possibles et quatre répétitions. ....	167
<b>Tableau 48:</b> Comparaison des valeurs formantiques (en Hz), pour F1, F2 et F3, obtenues pour les voyelles enregistrées par LOC1 et les valeurs de référence de Calliope (Tubach, 1989). L'écart est obtenu par la valeur absolue de la différence des F1 de Loc1 et Calliope. ....	170
<b>Tableau 49:</b> Comparaison des valeurs formantiques (en Hz), pour F1, F2 et F3, obtenues pour les voyelles enregistrées par LOC1 et les valeurs de référence de Gendrot & Adda (2005). L'écart est obtenu par la valeur absolue de la différence des F1 de Loc1 et Gendrot & Adda. ....	171
<b>Tableau 50:</b> Matrices de confusion par 8 locuteurs francophones natifs. En gris, les réponses au test en fonction du stimulus présenté (2 occurrences / phonème) et de la réponse donnée. Entre parenthèses, la moyenne des degrés de certitude sur la réponse entre 1 (pas du tout sûr) et 5 (tout à fait sûr). .	171
<b>Tableau 51:</b> Valeurs de F0 (en Hz), d'intensité (en dB) et de durée (en ms) pour les onze voyelles enregistrées par Loc1. Calcul de la moyenne, de l'écart type et des valeurs minimum et maximum. Le graphique représente la moyenne des trois valeurs acoustiques. ....	172

<b>Tableau 52:</b> Caractéristiques des différentes tâches en fonction de la modalité et de la variable contrôlée. Les sigles représentent le document visuel (DV), le document audio de parole (DAP) et le document audio de musique (DAM).....	173
<b>Tableau 53:</b> Répartition des différents groupes en fonction de la tâche.....	174
<b>Tableau 54:</b> Questions du questionnaire de la pré-étude 1 portant sur la sélection des documents pour la phase tâche.....	177
<b>Tableau 55:</b> Les 28 termes sélectionnés permettant de définir la projection émotionnelle des différents documents.....	178
<b>Tableau 56:</b> Réponses relatives à la question portant sur le tempo perçu des cinq documents. Les réponses importantes (>20) sont sur fond gris.....	179
<b>Tableau 57:</b> Distance relative par rapport au nombre de réponses données pour DAP3. Le résultat est obtenu par la valeur absolue de la différence entre la réponse d'une case et celle correspondante à DAP3. Une valeur faible signifie que le nombre de répondants est proche de celui de DAP3. Le classement est donné pour les documents musicaux d'une part et les documents visuels d'autre part.....	179
<b>Tableau 58:</b> Réponses les plus importantes pour le premier mot donné parmi les cinq demandés à la question 2. Seuls les mots ayant obtenus 5 réponses ou plus sont affichés. En gras, les données qui correspondent aux réponses pour DAP3, notre document de référence. « Total réponses » correspond à la somme des réponses supérieures à 7 et « TOTAL MAX » au nombre total de réponses données pour chaque document.....	180
<b>Tableau 59:</b> Réponses les plus importantes pour tous les mots à la question 2. Seuls les mots ayant obtenus 15 réponses ou plus sont affichés. En gras, les données qui correspondent aux réponses pour DAP3, notre document de référence. « Total réponses » correspond à la somme des réponses supérieures à 7. Le nombre total de réponses données est de 305.....	180
<b>Tableau 60 :</b> Réponses relatives à la question portant sur le tempo perçu des deux documents visuels. Les tendances importantes sont sur fond gris.....	182
<b>Tableau 61 :</b> Réponses les plus importantes pour tous les mots à la question 2. Seuls les mots ayant obtenu 7 réponses ou plus sont affichés. « Total réponses » correspond à la somme des réponses supérieures à 7 et « TOTAL MAX » au nombre total de réponses données pour chaque document.....	182
<b>Tableau 62:</b> Récapitulatif de l'analyse statistique Test t pour échantillons indépendants concernant les différences entre les quatre groupes au regard du tempo évalué.....	184
<b>Tableau 63:</b> Nombre de mots donnés lors du questionnaire en fonction du groupe.....	185
<b>Tableau 64 :</b> Récapitulatif du taux de discrimination (en %) en fonction du groupe, du test et du statut musical. Les écart-types sont donnés entre parenthèses.....	188
<b>Tableau 65:</b> Récapitulatif du temps de réaction moyen (en ms) en fonction du groupe, du test et du statut musical. Les écart-types sont donnés entre parenthèses.....	190
<b>Tableau 66 :</b> Interaction ordonnés entre Test et Statut. Les moyennes sont données en ms.....	192
<b>Tableau 67 :</b> Récapitulatif de l'analyse statistique test t pour échantillons indépendants concernant les différences entre musiciens et non-musiciens respectivement pour les tests 1 et 2.....	192
<b>Tableau 68 :</b> Récapitulatif du taux de discrimination (en %) en fonction du Statut langagier, du test et du statut musical. Les écart-types sont donnés entre parenthèses.....	195
<b>Tableau 69 :</b> Récapitulatif du temps de réaction (en %) en fonction du Statut langagier, du test et du Statut musical. Les écart-types sont donnés entre parenthèses.....	195
<b>Tableau 70:</b> Interaction ordonnés entre Test et Statut. Les moyennes des musiciens, non-musiciens, test 1 et test 2 sont données en ms.....	198
<b>Tableau 71:</b> Mode de présentation du trapèze vocalique aux apprenants selon les trois groupes.....	211
<b>Tableau 72:</b> Types d'input visuel et auditif en fonction des trois groupes d'apprenants.....	214

<b>Tableau 73:</b> Tableau comparatif (en Hz) de la différence absolue de la valeur de f0 des voyelles et la hauteur tonale de référence des notes de musique « do, ré, mi et fa » pour un diapason à 440 Hz ainsi que les notes jouées au piano midi. ....	215
<b>Tableau 74:</b> Valeurs formantiques (en Hz) des stimuli vocaliques utilisés pour les inputs des groupes 2 et 3 et mesurées avec Praat d'après la valeur moyenne de la totalité du phonème. ....	216
<b>Tableau 75 :</b> Durée des phonèmes et de l'input total en fonction du continuum et de la F0 produite. Les données sont en seconde. ....	217
<b>Tableau 76 :</b> Exercices utilisés lors du travail en groupe avec les oppositions phonémiques travaillées, le type d'exercice selon la typologie proposée par le manuel et les références du manuel.....	221
<b>Tableau 77:</b> Comparaison à l'aide d'un Test t pour échantillons appariés et moyennes entre les questionnaires 1 et 2 pour les cinq caractéristiques évaluées, tous groupes confondus. ....	223
<b>Tableau 78:</b> Tableau montrant l'évolution positive (+), négative (-) ou neutre (=) entre les questionnaires 1 et 2 pour les notes extrêmes (1 et 5) et centrales (2, 3 et 4) dont 1 correspond à « peu satisfait-e » et 5 « très satisfait-e ».....	224
<b>Tableau 79:</b> Effectifs de la soustraction entre les réponses données aux questionnaires 2 et 1. Une valeur positive (1, 2 3) signifie qu'un apprenant à mis une meilleur évaluation au questionnaire 2 qu'au 1. Une valeur négative (-1, -2) signifie l'inverse. Nous avons comptabilisé pour chaque question le nombre d'occurrence par tendance (n=27).....	225
<b>Tableau 80:</b> Tableau comparatif de la progression perçue par les apprenants en fonction des trois groupes de manière globale (à gauche ; n=45) et concernant la prononciation des voyelles (à droite ; n=9). ....	225
<b>Tableau 81:</b> Comparaison à l'aide d'un test t pour échantillons appariés et moyennes entre les questionnaires 1 et 3 pour les cinq caractéristiques évaluées.....	227
<b>Tableau 82:</b> Tableau montrant l'évolution positive (+), négative (-) ou neutre (=) entre les questionnaires 1, 2 et 3 pour les notes extrêmes (1 et 5) et centrales (2, 3 et 4). ....	229
<b>Tableau 83 :</b> Comparaison de la progression mesurée entre les réponses données aux questionnaires 1 et 2 ainsi que 1 et 3. Une valeur positive (1, 2 3) signifie qu'un apprenant à mis une meilleur évaluation au questionnaire 2 qu'au 1. Une valeur négative (-1, -2) signifie l'inverse. Nous avons comptabilisé pour chaque question le nombre d'occurrence par tendance (n=135). ....	229
<b>Tableau 84:</b> Tableau des effectifs, moyennes et écart types (n=27) des réponses à trois questions portant sur le rapport aux symboles phonétiques dans le questionnaire 2.....	232
<b>Tableau 85 :</b> Moyennes et écarts-type (Ecart-T) des réponses au questionnaire d'attitude concernant les 5 niveaux évaluées en fonction des trois groupes et des trois questionnaires. ....	233
<b>Tableau 86 :</b> Résultats de l'analyse ANOVA à mesures répétées pour l'interaction entre le groupe et les auto-évaluations entre les questionnaires 1 et 2 ainsi que 1 et 3 pour les cinq niveaux interrogés.233	233
<b>Tableau 87:</b> Tableau représentant (à gauche) la moyenne et l'écart type des réponses données en fonction de chaque groupe à la question « Selon vous, comment votre prononciation générale en français a évolué pendant le semestre 1 ? » et (à droite) la progression mesurée à la même question entre les questionnaires 2 et 3. (-) représente une différence négative donc une régression, (=) neutre donc une stagnation et (+) positive donc une progression entre les évaluations 2 et 3. ....	234
<b>Tableau 88:</b> Tableau représentant les réponses sur les différences ressenties entre les deux enregistrements. Les données sont présentées pour chaque affirmation en fonction de chaque groupe (G1, G2, G3, n=9) puis le total pour chaque réponse possible (TT, n=27). Les réponses nulles ont été représentées par une case vide pour plus de clarté. En gras, les données les plus marquantes pour le total de chaque réponse possible. ....	236
<b>Tableau 89:</b> Durées et écarts-type moyens (en ms) des phonèmes des deux continua analysés en fonction des tests 1 et 2 et des modalités de production. Les durées moyennes des inputs sont également données.....	240

- Tableau 90** : Moyenne de la différence de la F0 (en Hz) des différentes paires vocaliques en fonction des continua antérieur (ANT) et postérieur (POST) lors de l'enregistrement 1. En haut, les moyennes en fonction de chaque groupe d'apprenant et en bas (à gauche et à droite), pour chaque apprenant. En gris et gras, les résultats supérieurs à 0. Un nombre positif indique une augmentation positive moyenne entre chaque phonème du continua. La F0 a été calculée par Praat sur la moyenne du phonème. .... 242
- Tableau 91**: Analyse perceptive de la réalisation de la F0 attendue d'après l'input musical lors des enregistrements 1 et 2. Une croix indique que l'apprenant a réalisé les quatre phonèmes du continua avec une F0 montante proche de celle de l'input auditif. En gris, les apprenants ayant réalisé au moins un continua avec la forme attendue lors de l'enregistrement 1. .... 243
- Tableau 92**: Réponses aux questionnaires 2 et 3 des apprenants du groupe 3 concernant l'intérêt de l'association note de piano / phonème corrélée au nombre de continua produits avec la F0 attendue. « 1 » correspond à « tout à fait d'accord » et « 5 » à « pas du tout d'accord ». .... 244
- Tableau 93**: Réponses aux questionnaires 2 des apprenants du groupe 3 concernant les différences psycho-affectives entre les deux enregistrements et leur rapport au chant corrélée au nombre de continua produits avec la F0 attendue. .... 245
- Tableau 94** : Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français en fonction de l'apprenant et du groupe pour l'enregistrement 1. .... 247
- Tableau 95** : Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français en fonction du phonème et de la modalité de production..... 248
- Tableau 97** : Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français pour les six apprenants retenus lors de l'enregistrement 2..... 248
- Tableau 98** : Moyennes et écarts-types des degrés de confiance au test d'identification par les 10 locuteurs natifs du français pour les six apprenants retenus lors de l'enregistrement 2 en fonction du phonème et de l'enregistrement. L'évolution entre les deux enregistrements est indiquée par un symbole mathématique..... 249
- Tableau 99**: Matrices de confusion des 27 apprenant-es en fonction de la modalité de production (en haut) et global (en bas) et de la grille archiphonème. Les résultats sont donnés en pourcentages. Seules les données supérieures à 0 sont indiquées, celles supérieures à 10% sont surlignées en gris et les identifications correctes signalées en gras. .... 252
- Tableau 100**: Matrices de confusion des 27 apprenant-es en fonction de la modalité de production et de la grille extrême. Les résultats sont donnés en pourcentages (n=540/ligne). Seules les données supérieures à 0 sont indiquées, celles supérieures à 10% sont surlignées en gris et les identifications correctes signalées en gras. .... 253
- Tableau 101**: Matrices de confusion globale des 27 apprenant-es globale et de la grille extrême. Les résultats sont donnés en pourcentages (n=1080/ligne). Seules les données supérieures à 0 sont indiquées, celles supérieures à 10% sont surlignées en gris et les identifications correctes signalées en gras. .... 254
- Tableau 102** : Langues premières des apprenants en fonction des trois groupes (Gr.) et présence phonologique ou non des sons /u/ et /y/ dans les différentes langues d'après UCLA Phonological Segment Inventory Database (UPSID) (Maddieson, 1984). Le /y/ apparait également pour le coréen dans UPSID mais il semblerait que cette existence soit remise en question surtout concernant le coréen contemporain (Schuh, 2013)..... 255
- Tableau 103**: Comparaison des matrices de confusions entre les modalités « continuum » et « isolé » en fonction du groupe et selon la grille « archiphonème ». Les résultats sont donnés en pourcentage. Un nombre négatif indique qu'en moyenne cette confusion est plus apparue à l'isolée et un nombre positif indique que cette confusion est plus apparue en continuum. En gris foncé, les données supérieures à 10% et en gris clair, inférieures à -10%..... 255

**Tableau 104:** Comparaison des matrices de confusion pour les phonèmes /e-ε/ et /o-ɔ/ entre les modalités « continuum » et « isolé » en fonction du groupe et selon la grille « extrême ». Les résultats sont donnés en pourcentage. Un nombre négatif indique qu'en moyenne cette confusion est plus apparue à l'isolée et un nombre positif indique que cette confusion est plus apparue en continuum. En gris foncé, les données supérieures à 10% et en gris clair, inférieures à -10%..... 256

**Tableau 105:** Résultats au Test t pour échantillons appariés entre les pré- et post-test pour les deux modalités de production avec la grille « extrême » et « archiphonème ». ..... 258

**Tableau 106 :** Nombre d'apprenants montrant une progression, stagnation ou régression entre les deux enregistrements en fonction de la grille et de la modalité. Les moyennes (Moy.) correspondent au nombre moyen d'identifications (n=6)..... 258

**Tableau 107:** Différence entre les matrices de confusion du post-test et du pré-test des 6 apprenants en fonction du type de grille archiphonème (en haut) et extrême (en bas). Les résultats sont donnés en pourcentages. Seules les données différentes de 0 sont indiquées. Un pourcentage positif indique que l'opposition a plus été identifiée au post-test qu'au pré-test (en gris clair) et un pourcentage négatif indique que l'opposition a moins été identifiée au post-test qu'au pré-test (en gris foncé). ..... 259



## **Approche plurielle des liens musique-parole pour la didactique de la prononciation du français en contexte de langue étrangère/seconde.**

Cette thèse s'intéresse aux utilisations de la musique instrumentale pour la didactique de la prononciation du français comme langue étrangère/seconde (FLES). Elle s'inscrit dans une perspective plurielle et un cadre théorique émergentiste et socioconstructiviste. La thématique est abordée par trois études (1) psychosociale, (2) psychoacoustique et (3) didactique permettant de questionner à plusieurs niveaux les apports potentiellement développementaux de la musique instrumentale pour la prononciation du français en contexte FLES.

Le but de cette thèse est une contribution à la didactique des langues afin d'apporter des réponses empiriques à des pratiques pédagogiques présentes depuis longtemps mais peu questionnées scientifiquement. Elle permet également de prendre du recul sur les récents résultats en neurocognition et notamment leurs applications pour un développement langagier.

L'étude 1 psychosociale, par questionnaire, montre que la didactique de la musique peut nous aider à amener les apprenants à se focaliser sur le son plutôt que sur le sens des unités sonores langagières. L'étude 2 psychoacoustique révèle que les musiciens ont de meilleurs temps de réaction que les non-musiciens chez des apprenants du français pour un test AXB d'oppositions fines des phonèmes vocaliques du français. Grâce à l'étude 3, didactique interventionniste et quasi-expérimentale, nous avons montré qu'une association entre notes de piano et aperture suggère une potentialité développementale des phonèmes vocaliques du français mais que cette dernière n'est pas supérieure à d'autres approches plus conventionnelles. Cependant, elle semble être un apport cognitif pour certains apprenants même si cet apport montre une tendance bimodale. Ces trois études permettent de montrer que la musique instrumentale peut trouver sa place dans la « boîte à outils » de l'enseignant.

**Mots clés :** *didactique des langues, français langue étrangère, musique instrumentale, développement langagier, didactique de la prononciation, perception et production.*

## **Pluralistic approach of the links between music and speech in the learning and teaching of the pronunciation of French as a foreign/second language.**

This work deals with the use of instrumental music in the teaching and learning of French as a foreign language. It is part of a pluralistic theoretical framework and an integrated approach of emergentism and social cultural theories. The topic is discussed through three studies: (1) a psychosocial study, (2) a psychoacoustic study and (3) an applied linguistics study – questioning at different levels the potentially developmental intake of instrumental music for French pronunciation as a foreign language.

The aim is a contribution to applied linguistics in order to bring empirical data to long-lasting pedagogical practices but also to question recent neurocognitive results in an applied perspective.

The first study, which is psychosocial and was led by questionnaires, shows that music learning and teaching can help us to have students focus on sound rather than meaning. The second study, which is psychoacoustic, reveals that, among learners of French, musicians have better reaction times than non-musicians to an AXB discrimination test of close French vowels. The third study, which is quasi-experimental and in intervention, suggests that associating piano notes and vowel aperture doesn't lead to bigger developmental intake compared to more conventional approaches. However, it seems to point out to the fact that some learners experience cognitive input even if such impact evokes bimodal tendencies. These three studies prove that instrumental music can be a tool to assist teachers.

**Keywords:** *applied linguistics, didactics, French as a foreign language, instrumental music, pronunciation, perception & production.*

UNIVERSITE SORBONNE NOUVELLE - PARIS 3

ED 268 – Langage et langues : description, théorisation, transmission

Didactique des Langues, des textes et des cultures - DILTEC (EA2288)

Laboratoire de Phonétique et Phonologie - LPP (UMR7018)

1, rue Censier, 75005 Paris