

*Effets de la formation musicale
sur la production
de l'anglais*

Table du huitième chapitre

1) Recueil du matériel de parole

- 1.1. Type de discours
- 1.2. Passation
 - 1.2.1. Entretien avec un locuteur natif
 - 1.2.2. Conditions d'enregistrement
 - 1.2.3. Déroulement de l'entretien
 - 1.2.4. Matériel de parole

2) Etude n°3 « Production de la prosodie »

- 2.1. Plan expérimental
 - 2.1.1. Hypothèse
 - 2.1.2. Variables indépendantes
 - 2.1.2.1. La musique
 - 2.1.2.2. Le discours
 - 2.1.3. Variables dépendantes
 - 2.1.3.1. Forme de la courbe
 - 2.1.3.1.1. Caractéristiques des productions du locuteur natif
 - 2.1.3.1.2. Classement des contours en fonction de l'initiale du mot
 - 2.1.3.1.3. Les cinq modalités
 - 2.1.3.1.4. Classement des courbes
 - 2.1.3.2. Exagération de la courbe
 - 2.1.3.3. Localisation de l'accent
 - 2.1.4. Analyses acoustiques
- 2.2. Résultats
 - 2.2.1. Forme de la courbe
 - 2.2.1.1. Mots dont l'initiale est non voisée
 - 2.2.1.2. Mots dont l'initiale est voisée
 - 2.2.2. Amplitude de la courbe
 - 2.2.3. Localisation de l'accent

3) Etude n°4 « Production des noyaux vocaliques »

- 3.1. Plan expérimental
 - 3.1.1. Hypothèse
 - 3.1.2. Variables indépendantes
 - 3.1.2.1. La musique
 - 3.1.2.2. Le discours
 - 3.1.3. Variables dépendantes
 - 3.1.3.1. Timbre du noyau vocalique
 - 3.1.3.2. Durée du noyau vocalique
 - 3.1.4. Mesures acoustiques
 - 3.1.4.1. Mesures de durée
 - 3.1.4.2. Mesures de timbre
- 3.2. Résultats
 - 3.2.1. Timbre du noyau vocalique
 - 3.2.1.1. Voyelles

- 3.2.1.2. Diphtongues
- 3.2.1.3. Résumé
- 3.2.2. Durée du noyau vocalique
 - 3.2.2.1. Voyelles brèves
 - 3.2.2.2. Voyelles longues
 - 3.2.2.3. Diphtongues

4) Etude n°5 « Stabilité de timbre »

- 4.1. Plan expérimental
 - 4.1.1. Hypothèse
 - 4.1.2. Variables indépendantes
 - 4.1.2.1. La musique
 - 4.1.2.2. Le discours
 - 4.1.3. Variable dépendante : durée de la zone de stabilité de timbre de la voyelle
- 4.2. Résultats
 - 4.2.1. Résultats généraux
 - 4.2.2. Résultats détaillés
 - 4.2.2.1. Voyelles longues
 - 4.2.2.2. Voyelles brèves

5) Résumé des résultats et discussion générale

- 5.1. Production de la prosodie
 - 5.1.1. Forme de la courbe
 - 5.1.2. Localisation de l'accent
 - 5.1.3. Exagération de la courbe
- 5.2. Production des contrastes vocaliques
 - 5.2.1. Timbre du noyau vocalique
 - 5.2.1.1. Voyelles brèves
 - 5.2.1.2. Voyelles longues
 - 5.2.1.3. Diphtongues
 - 5.2.2. Durée du noyau vocalique
 - 5.2.3. Stabilité de timbre

Bibliographie du chapitre VIII

Dans l'apprentissage d'une langue étrangère, l'objectif consiste à se rapprocher le plus possible de la langue cible et, pour ce qui nous intéresse, de la prononciation native. Analyser, puis juger les productions des apprenants d'une langue donnée, c'est donc déterminer où ils se situent par rapport au « modèle » à acquérir et quelles stratégies sont mises en œuvre pour y parvenir ; c'est déterminer également si la LM produit des interférences avec la langue cible. Dans le chapitre V, nous avons décrit le système prosodique et vocalique de l'anglais « standard » (dans le sens du modèle à acquérir) par rapport à celui du français « standard », de façon à pouvoir prévoir ces interférences. Dans ce contexte, est-ce que la pratique musicale d'un enfant va lui permettre d'« être » mieux armé pour se confronter à une nouvelle langue et la reproduire ? Nous avons vu que les capacités musicales ont une influence sur la perception des contrastes vocaliques. Est-ce que le fait d'avoir une oreille musicienne va permettre de mieux reproduire ce que l'on entend mieux ? Et est-ce que les capacités à percevoir vont susciter une meilleure production des éléments musicaux de la langue ? Afin de le savoir, nous avons réalisé une étude sur la qualité des productions en anglais des enfants de notre corpus, en fonction de la « qualité » de leur oreille et de leur pratique musicale.

Nous posons donc comme hypothèse générale que la pratique musicale induit une meilleure production d'une langue étrangère : *« une oreille affinée par la pratique musicale induira une meilleure production d'une langue étrangère »*. Cette hypothèse peut être appréhendée en deux hypothèses particulières : une oreille affinée par la pratique musicale induira une meilleure production de la prosodie d'une langue étrangère ; une oreille affinée par la pratique musicale induira une meilleure production des contrastes vocaliques d'une langue étrangère.

1) Recueil du matériel de parole

1.1. Type de discours

Les enfants sont débutants en LV, on ne peut donc pas s'attendre à ce qu'ils s'expriment spontanément en anglais. Nous avons donc exclu de travailler sur du discours spontané. En revanche, les enfants apprennent un certain nombre de mots en classe et ce répertoire de mots est commun à de nombreux débutants. Ainsi, ils commencent par apprendre à compter, à nommer les animaux familiers, les couleurs et les objets de leur entourage. Ce répertoire constitue le premier vocabulaire qu'un apprenant enfant mémorise dans une langue étrangère, et pour ce qui nous intéresse, en anglais.

Nous avons ainsi constitué une liste de 43 mots, tirés du vocabulaire des couleurs, des nombres, des animaux et des objets familiers (voir matériel de parole, 1.2.4.). Nous avons également interrogé les professeurs d'anglais des trois écoles pour connaître le vocabulaire que les enfants avaient appris depuis le début de l'année scolaire (débutée il y a 6 mois).

Dans des études précédentes (Dodane, 1997, 2000), nous avons observé des comportements différents chez les enfants lorsqu'ils reproduisaient des mots pour la première fois, par rapport au moment où ils répétaient des mots qu'ils avaient déjà assimilés. En « répétition immédiate », l'apprenant essaie de reproduire fidèlement le mot qu'il entend de la bouche du locuteur, car il ne connaît pas ce mot. Ce type d'imitation est nécessaire pour lui permettre de réussir à reproduire et à mémoriser le mot cible. En « répétition différée », il a déjà mémorisé le mot et il peut « jouer » avec lui, de manière à se l'approprier complètement, quitte à exagérer les caractéristiques sonores qui lui paraissent « nouvelles » par rapport à celles de sa LM. Ce comportement est positif, puisqu'il démontre que l'apprenant est en pleine phase d'apprentissage et d'appropriation. On peut rapprocher la notion de « répétition immédiate » de la notion d'« *imitation-copie* » et celle de « répétition différée », d'« *imitation-jeu* ». Selon Piaget (1937), ces deux types d'imitation sont indispensables à l'enfant car ils lui permettent de s'adapter à son environnement et ici, de s'approprier progressivement une langue qui lui est étrangère.

Cependant, il est impossible de déterminer a priori quels sont les mots que les enfants connaissent et ceux qu'ils ne connaissent pas. Leur entourage a pu en effet leur apprendre des mots supplémentaires par rapport au vocabulaire appris en classe et ils ont pu oublier des mots qu'ils avaient appris en début d'année. Les enfants des trois classes sont donc censés connaître une partie des mots de cette liste, mais pas tous les mots. Nous avons donc établi notre classement de la manière suivante : si les enfants connaissent les mots de la liste, leurs productions sont classées dans la catégorie de discours « répétition différée » (RD). S'ils ne les connaissent pas et qu'ils doivent les répéter à la suite du locuteur natif, leurs productions sont classées dans la catégorie de discours « répétition immédiate » (RI).

1.2 Passation

1.2.1. Entretien avec un locuteur natif

Afin d'obtenir ces deux types de discours, nous avons choisi de placer les enfants dans une situation d'entretien avec un locuteur anglophone natif qui n'est pas l'enseignant habituel de la classe et que les enfants n'ont jamais rencontré auparavant. Cette situation d'entretien nous permet de conserver les mêmes conditions d'expérimentation pour les 3 classes (personne inconnue, la même pour les 3 classes). La situation d'entretien avec un anglophone est au départ une situation un peu angoissante pour l'enfant, qui a l'habitude de s'exprimer en classe et avec son professeur habituel. Les entretiens ont donc eu lieu juste après les tests de musique (chapitre VI) et les tests de perception en anglais (chapitre VII). Ainsi, les enfants, déjà habitués à l'expérimentatrice principale, ne se sont pas montrés trop déroutés par l'arrivée d'un nouvel inconnu, qui plus est, parlait une langue étrangère.

Le locuteur anglophone natif avait déjà fait l'expérience de ce genre d'entretien puisqu'il avait participé à l'élaboration de notre corpus de maîtrise, réalisé dans des conditions similaires (Dodane, 1996). Nous connaissions donc son comportement avec les enfants et c'est pour cette raison que nous avons de nouveau fait appel à lui. Son accent est non marqué, bien qu'il soit natif américain. Cependant, la question de l'accent de ce locuteur a relativement peu d'importance étant donné que notre étude ne porte pas sur cet accent, mais sur la capacité des enfants à reproduire. Le locuteur natif ne devait jamais s'exprimer en français pendant toute la durée de l'entretien. En cas de chevauchement ou d'inintelligibilité, il avait pour consigne de faire répéter les mots aux enfants, de manière à ce que ces mots soient exploitables pour l'analyse acoustique. Si leur voix était trop faible, il leur demandait de parler plus fort et si les enfants répondaient en même temps ou s'il y avait un bruit (la sonnerie au moment de la récréation par exemple), il leur faisait répéter le mot.

1.2.2. Conditions d'enregistrement

Dans les trois écoles, les enregistrements ont eu lieu dans une salle familière des enfants, la plus calme possible (salle de classe de chant pour les enfants de l'école de Dijon, salle de classe d'anglais pour les enfants de l'école de Choisey et salle de langue pour les enfants de l'école de Dole). En raison de la législation concernant la protection des mineurs, un enfant ne peut rester seul avec une ou plusieurs personnes adultes étrangères à l'établissement scolaire.

Comme pour les tests de perception, les enfants passent donc par deux, ce qui porte à quatre le nombre de personnes présentes dans la pièce : l'expérimentateur principal, le locuteur natif et deux enfants. Les enregistrements ont eu lieu au cours des mois de janvier et février 2001 alors que les enfants avaient commencé l'anglais depuis 6 mois. Ils avaient donc un niveau suffisant pour participer à un entretien avec un locuteur natif. Pour enregistrer les productions des enfants, nous avons utilisé un enregistreur mini-disque (de marque Sony, modèle « Portable Minidisc Recorder, M2-R30 ») avec un microphone (de marque Sony).

1.2.3. Déroulement de l'entretien

L'entretien se déroule en cinq étapes. La première commence par la présentation entre les enfants et le locuteur natif. Les enfants ont déjà appris à se saluer, à se présenter, à dire leur âge et comment ils vont. Le locuteur commence par se présenter, par exemple « *Hello, my name is Ronald* » puis il pose des questions aux enfants telles que « *What is your name ?* », « *How are you ? Are you fine ?* » ou encore « *How old are you ?* ». Toutefois, tous les enfants n'étaient pas capables de répondre à ces questions et ce matériel ne peut donc être étudié systématiquement dans notre corpus.

Après les présentations, viennent les couleurs. Le locuteur montre aux enfants des couleurs tirées d'un livre illustré (Beaumont et Desmoineaux, 2000). Il fait la même chose pour les animaux, tirés d'un autre livre illustré destiné à des enfants anglophones (Butterfield et O'Neill, 1997), dans lequel un système de pliage permet de découvrir les animaux, cachés derrière un tas de foin ou une porte d'écurie par exemple. Viennent ensuite les nombres, pour lesquels nous avons réalisé des supports cartonnés les représentant, de 1 à 13. Enfin, le locuteur termine avec des images tirées de l'Imagier du Père Castor en anglais (1997). Dans ce livre, les mots qui sont indiqués sous les images ont été occultés par des caches. Nous avons diversifié les supports de manière à introduire de la nouveauté dans un entretien qui pouvait s'avérer un peu long pour les enfants (les entretiens pouvaient durer jusqu'à 20 minutes environ).

Tout au long de l'entretien, le locuteur natif devait susciter la production en répétition différée des mots en montrant une image et en posant simultanément une question aux enfants du type « *What colour is it ?* », « *What is this animal ?* », « *Who is it ?* » ou encore « *What's this ?* », « *What is it ?* » ou simplement « *And this ?* » (voir exemple d'entretien, Volume II, annexe n°8.1. et **extrait sonore n°54**).

Ces questions étaient comprises par les enfants puisqu'elles étaient posées régulièrement par leur professeur d'anglais pendant les cours. Pour susciter la répétition immédiate, le locuteur pose les mêmes questions, mais après le temps de silence lui indiquant que l'enfant ne peut pas répondre, il dit le mot et demande de répéter après lui. Par exemple, « *And this ? What's this ?* (il attend, tout en désignant l'image) *No ? It's a duck. Can you repeat, a duck* » ou « *This... this animal. What is this animal ?* (il attend) *This animal is a parrot* » ou encore « *And this ?* (il attend) *Can you say what is this ?* (il attend) *No ?* (il attend) *It's a car* » et l'enfant doit répéter.

1.2.4. Matériel de parole

La liste est composée de 43 mots, dont 10 mots désignent des couleurs (« *green, pink, red, black, orange, blue, purple, white, yellow, brown* »), 7 mots, des animaux (« *dog, cow, sheep, bird, cat, rabbit, horse* »), 13 mots, des nombres (suite de 1 à 13) et 13 mots, des objets ou des êtres vivants familiers (« *boy, baby, doll, teddy, book, cake, fish, apple, parrot, duck, umbrella, car, banana* »). Parmi ces mots, 30 sont monosyllabiques, 10 sont bisyllabiques et 3 trisyllabiques (voir présentation du corpus, Volume II, annexe n°8.2.).

2) Etude n°3 « Production de la prosodie »

2.1. Plan expérimental

2.1.1 Hypothèse

Les locuteurs dont l'oreille a été entraînée par la pratique musicale rencontreront plus de facilités à reproduire les éléments musicaux de l'anglais.

2.1.2. Variables indépendantes

2.1.2.1. La musique : deux modalités

La variable indépendante « musique » est constituée de deux modalités, la modalité « musique » (G1) et la modalité « non musique » (G2) (voir chapitre VII, 1.1.2.) Cependant, par rapport aux études n°1 et n°2 sur la perception de l'anglais, nous avons dû réduire l'effectif du groupe G2 à 25 sujets, car l'un des entretiens n'était pas exploitable pour une analyse acoustique (mauvais rapport signal/bruit).

2.1.2.2. Le discours : deux modalités

Au cours de l'entretien avec le locuteur natif, les enfants sont amenés à produire deux types d'énoncés, des énoncés en répétition immédiate (RI) à la suite du locuteur ou des énoncés en répétition différée (RD) du vocabulaire appris en classe (voir 1.1.).

2.1.3. Variables dépendantes

2.1.3.1. Forme de la courbe

2.1.3.1.1. Caractéristiques des productions du locuteur natif

A cause de l'entraînement de leur oreille, les enfants musiciens devraient avoir plus de facilité à repérer les contours caractéristiques de l'anglais et donc, si nous supposons qu'une meilleure perception entraîne une meilleure production, ils réussiraient mieux à en reproduire la forme. Le contour typique qui affecte les mots monosyllabiques commençant par une initiale voisée en anglais est un contour qui prend la forme d'une cloche (montée, pic de Fo sur le noyau vocalique, puis descente). La figure n°83 nous en montre un exemple pour le mot « green »¹ (extrait sonore n°55).

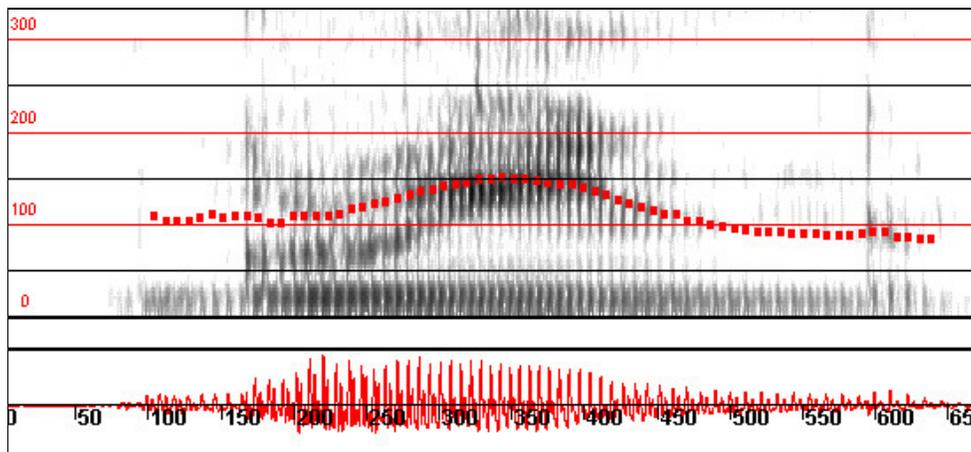


Figure n°83 : courbe mélodique en cloche sur le mot « green » par le locuteur natif
(représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

Sur chacun des 17 mots monosyllabiques de notre corpus qui commencent par une initiale voisée, (voir la liste des mots dans l'annexe n°X et les mesures réalisées, volume II, annexe 9.1.1.), le locuteur natif produit un contour en cloche.

¹ Durée de l'énoncé, 546 ms – points d'inflexion de la courbe « en cloche » : 105 Hz / 150 Hz / 85 Hz.

Ces mots peuvent commencer soit par une consonne voisée comme [g] dans « *green* », soit par une voyelle comme [ʌ] dans « *one* ».

Mais le locuteur natif ne produit pas que des « cloches » sur les mots monosyllabiques. Il produit également des contours descendants dans les mots commençant par une consonne non voisée. C'est le cas pour les 11 autres mots du corpus. La figure n°84 nous donne un exemple de ce type de contour, qui ici, affecte le mot « *ten* »² (**extrait sonore n°56**).

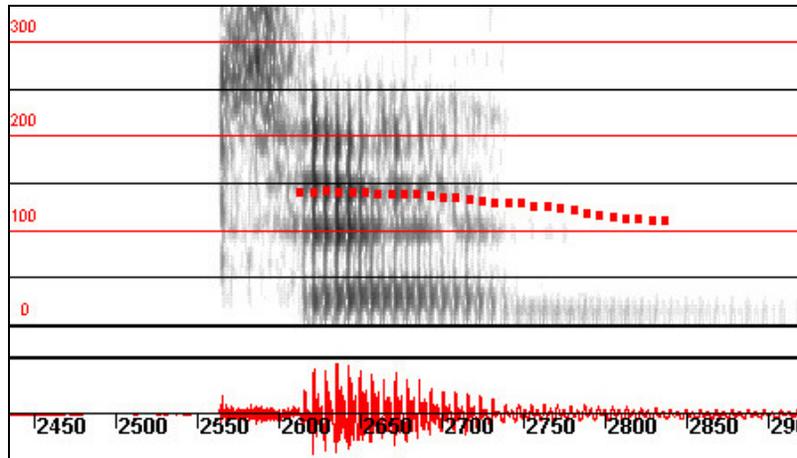


Figure n°84 : courbe mélodique descendant sur le mot « *ten* » par le locuteur natif (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

Il y a cependant deux exceptions à ce classement, le mot « *twelve* » et le mot « *three* » qui commencent tous les deux par une consonne non voisée, mais sont affectés d'un contour en cloche. La présence de ce contour s'explique par le fait qu'un son voisé est intercalé entre la consonne et la voyelle, dans ces deux mots « *w* » ou « *r* ».

2.1.3.1.2. Classement des contours en fonction de l'initiale du mot

En fonction de ces données, nous classerons les productions des enfants en deux catégories :

- la première catégorie « initiale non voisée » comprend les mots commençant par une consonne non voisée ; elle englobe 11 mots sur les 30 mots du corpus ;
- la deuxième catégorie « initiale voisée » comprend les mots commençant par une consonne voisée, par une voyelle ou par une consonne sourde suivie d'un son voisé ; elle englobe 19 mots sur les 30 mots du corpus.

² Durée de l'énoncé, 437 ms – points d'inflexion de la courbe « en descente » : 142 Hz / 110 Hz.

2.1.3.1.3. Les cinq modalités

Si les enfants restituent parfaitement les contours des 30 mots de notre corpus, ils devraient donc produire un contour en cloche pour les mots commençant par un son voisé et un contour descendant pour les mots commençant par une consonne non voisée. Mais, le contour en cloche est caractéristique de l'anglais et il est probable que certains enfants, à cause du crible prosodique, produisent des contours descendants comme ils le feraient dans leur langue maternelle. Par ailleurs, les enfants peuvent également produire des contours montants, des contours plats ou des contours en forme de « cloche inversée » correspondant à un contour de type haut-bas-haut (désormais HBH).

La variable dépendante « forme » de la courbe est donc une variable à 5 modalités : contour en forme de « cloche », de « cloche inversée », de « descente », de « montée » ou « recto-ono » (voir figure n°85 et figure n°86, correspondant à l'**extrait sonore n°57**).

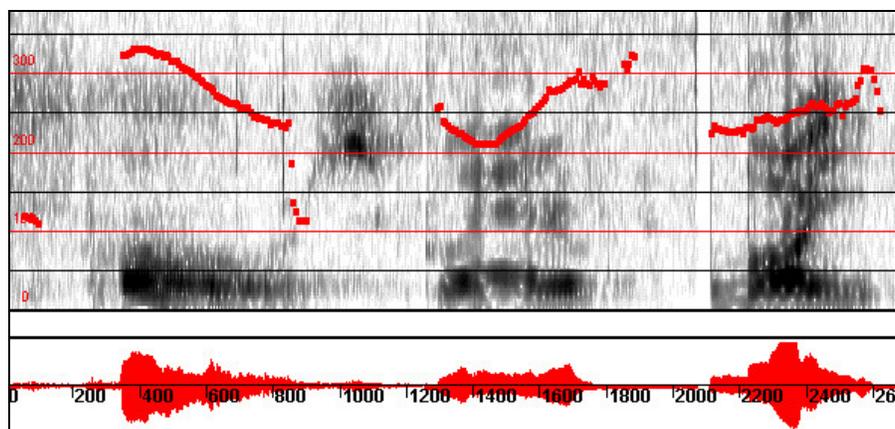


Figure n°85 : courbes mélodiques affectant les mots « horse », « twelve » et « boy » représentant successivement les modalités « cloche », « cloche inversée », « montée » (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

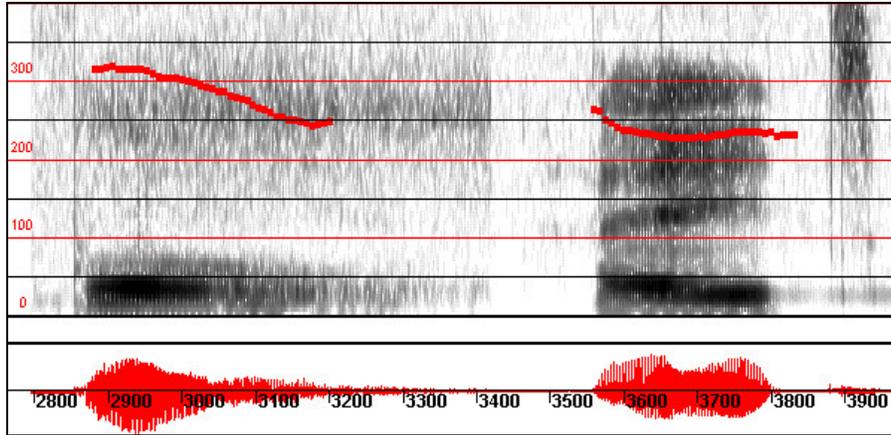


Figure n°86 : courbes mélodiques affectant les mots « cow » et « eight » représentant successivement les modalités « descente » et « recto-tono » (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

Pour cette variable, nous posons l'hypothèse opérationnelle que la proportion de contours en « cloche » sera plus importante dans les mots monosyllabiques à l'initiale voisée chez les enfants appartenant au « groupe musiciens » (G1) que chez les enfants du « groupe non musiciens » (G2). Par ailleurs, le contour en cloche est caractéristique de l'anglais et il est possible que certains enfants généralisent son utilisation pour donner une couleur plus « anglaise » à leurs énoncés, et ce, même dans les occurrences où il ne devrait pas apparaître. Nous posons donc comme hypothèse que les enfants du groupe G1 produiront plus de contours en cloche dans les mots monosyllabiques à l'initiale non voisée que les enfants du groupe G2. Le plan expérimental est exposé dans le tableau n°47.

Initiale de mot	Type de contour	G1		G2	
		RD	RI	RD	RI
Non voisée	« Cloche »				
	« Cloche inversée »				
	« Montée »				
	« Descente »				
Voisée	« Recto-tono »				
	« Cloche »				
	« Cloche inversée »				
	« Montée »				
	« Descente »				

Tableau n°47 : plan expérimental pour la variable « forme (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate et « RD » pour répétition différée »).

2.1.3.1.4. Classement des courbes

Nous avons volontairement limité le nombre de catégories à cinq, notre but n'étant pas de réaliser une analyse prosodique fine et détaillée, mais de faire apparaître des tendances générales dans la production des contours. Ces contours ont été répartis dans les différentes catégories en fonction de la forme de leur courbe mélodique apparaissant sur la représentation mélodique du logiciel de mesure et du jugement auditif de l'expérimentatrice. Pour le « groupe musiciens » (G1), 300 énoncés ont ainsi été collectés, parmi lesquels 26 énoncés n'ont pu être analysés (soit 8,66 % de la totalité des énoncés). Pour le « groupe non musiciens » (G2), 450 énoncés ont été collectés, dont 27 n'ont pas pu être analysés (soit 6 % de la totalité des énoncés). Le total des énoncés analysés se porte donc à 697, dont 274 pour G1 et 423 pour G2.

2.1.3.2. Exagération de la courbe

Des études précédentes (Dodane, 1997, 2000) ont montré que certains enfants musiciens exagéraient les paramètres prosodiques et spécialement la courbe mélodique lorsqu'ils restituaient des mots qu'ils avaient appris en classe (répétition différée, RD). Cette exagération pouvait être franchement caricaturale par rapport aux productions d'un locuteur natif. Est-ce que cette exagération est caractéristique d'un locuteur qui possède une oreille entraînée par la pratique musicale ? En effet, un enfant musicien, habitué à manipuler les paramètres musicaux au cours de sa formation musicale, pourrait reporter cette stratégie dans la restitution des contours mélodiques d'une langue étrangère, en particulier lorsque les sons de cette langue sont difficiles à reproduire.

Il serait donc intéressant d'examiner si les contours des enfants musiciens sont plus amples que ceux du locuteur natif dans notre corpus. Afin de simplifier la comparaison et de restreindre la complexité des contours, nous nous limiterons aux seuls mots monosyllabiques du corpus, la forme de la courbe étant nettement influencée par la localisation de l'accent dans les mots multisyllabiques. Nous avons vu précédemment que dans ce type de mots, le locuteur natif produisait deux types de contours, des contours descendants et des contours en cloche. Pour connaître l'amplitude des contours produits par le locuteur natif, une analyse a été réalisée sur 37 contours en cloche et 17 contours descendants (les résultats détaillés sont exposés dans le volume II, annexe n°9.2.1.) ; ces énoncés ont été extraits de deux entretiens différents entre le locuteur natif et quatre enfants prélevés dans le corpus.

Le nombre d'énoncés analysés est limité, le but n'étant pas de faire une analyse exhaustive des productions du locuteur natif, mais de celles des enfants (les contours exposés pour la variable précédente sont compris dans cette analyse). Néanmoins, elle nous permettra d'avoir une idée du portrait prosodique du locuteur natif.

Pour les contours en cloche, le contour « moyen » (pour 37 productions) est composé d'une inflexion montante de $8,21 \frac{1}{4}$ de tons, soit approximativement d'une sixte mineure et d'une inflexion descendante de $13,67 \frac{1}{4}$ de tons, soit approximativement d'une neuvième mineure (figure n°87).

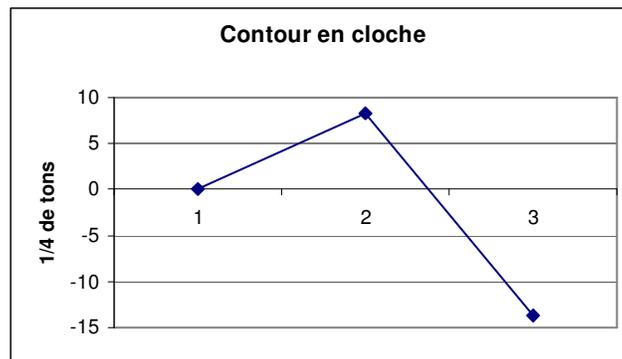


Figure n°87 : « contour moyen » construit à partir de 37 contours en cloche produit par le locuteur natif.

Le contour le plus ample a été relevé pour le mot « *blue* »³. La première inflexion du contour s'élève d'un intervalle de $21 \frac{1}{4}$ de tons, soit approximativement d'une septième mineure, et la deuxième inflexion s'abaisse d'un intervalle de $24 \frac{1}{4}$ de tons, soit approximativement une octave (figure n°88 et **extrait sonore n°58**).

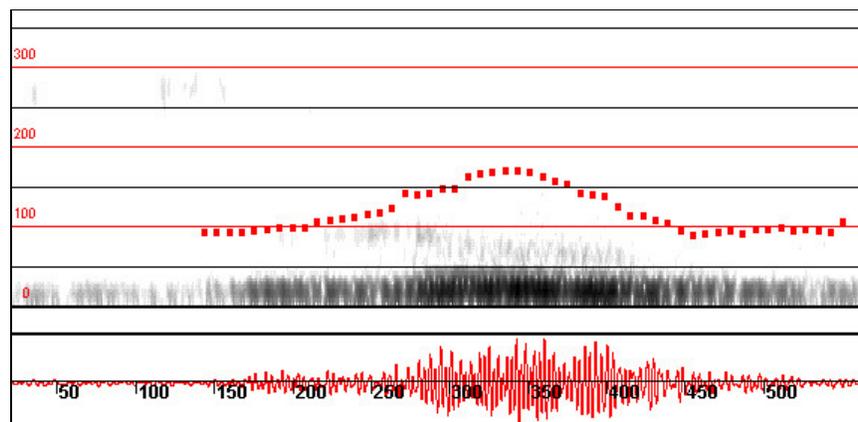


Figure n°88 : courbe mélodique sur le mot « *blue* » par le locuteur natif (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

³ Durée de l'énoncé, 538 ms – points d'inflexion de la courbe en cloche : 92 Hz / 170 Hz / 93 Hz.

Pour les contours descendants, l'intervalle moyen (pour 17 productions) est de $10,40 \frac{1}{4}$ de tons, soit approximativement d'une septième mineure (figure n°89). Le contour le plus ample a été relevé pour le mot « *pink* »⁴, qui s'abaisse d'un intervalle de $20 \frac{1}{4}$ de tons, soit approximativement une septième mineure (voir figure n°90 et **extrait sonore n°59**).

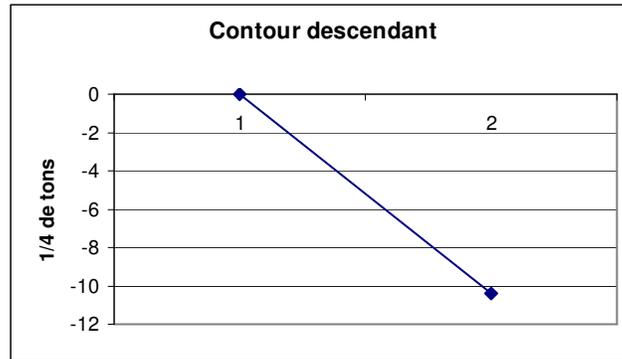


Figure n°89 : « contour moyen » construit à partir de 17 contours descendants produit par le locuteur natif.

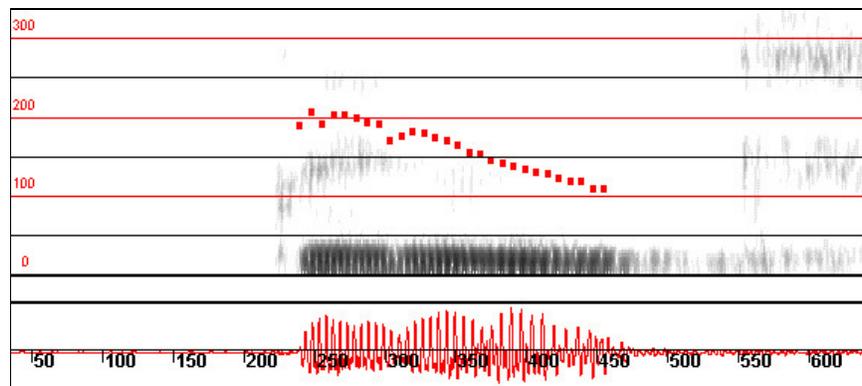


Figure n°90 : courbe mélodique sur le mot « *pink* » par le locuteur natif (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

Les contours du locuteur natif sont donc très amples et nous pensons que les enfants musiciens vont reproduire plus facilement cette amplitude de la courbe mélodique, voire l'exagérer.

En fonction des différents types de contours produits par les enfants dans notre corpus, la variable « exagération de la courbe » comprend quatre modalités : « cloche », « cloche inversée », « montée » et « descente » (nous avons retiré de l'analyse les contours « recto-ono »). Nous posons l'hypothèse opérationnelle qu'en situation de répétition différée (RD), l'amplitude de la courbe sera exagérée chez les enfants musiciens.

⁴ Durée de l'énoncé, 476 ms – points d'inflexion de la courbe en descente : 206 Hz / 110 Hz.

Le plan expérimental est exposé dans le tableau n°48.

	G1				G2			
	RD		RI		RD		RI	
	A	B	A	B	A	B	A	B
Cloche								
Cloche Inversée								
Montée								
Descente								

Tableau n°48 : plan expérimental pour la variable « exagération de la courbe » (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate et « RD » pour répétition différée » ; A désigne l'intervalle correspondant à la première inflexion de la courbe et B, l'intervalle correspondant à la deuxième inflexion de la courbe pour les contours en cloche et en cloche inversée).

Afin d'étudier l'amplitude de la courbe, nous avons relevé les intervalles (exprimés en ¼ de tons) séparant :

- Le point de départ et le point d'arrivée de la courbe dans les contours montants et descendants ;
- Le point de départ et le point d'arrivée de la première inflexion de la courbe dans les contours en cloche et en cloche inversée (intervalle A) ;
- Le point de départ et le point d'arrivée de la deuxième inflexion de la courbe dans les contours en cloche et en cloche inversée (intervalle B).

Le nombre d'énoncés analysés dans le corpus figure dans le tableau n°49, ci-dessous.

	G1		G2	
	RD	RI	RD	RI
Nombre d'énoncés	122	109	243	172
Total	231		415	
Total général	646			

Tableau n°49 : nombre d'énoncés analysés dans l'étude de la variable « amplitude de la courbe » (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

2.1.3.3. Localisation de l'accent

Dans le chapitre V (2.1.3.) et dans le catalogue d'interférences (Volume II, annexe n°3.2.1., « erreur de localisation de l'accent tonique »), nous avons vu que les francophones ont tendance à produire l'accent en finale de groupe rythmique, alors qu'en anglais, la tendance est à l'accentuation de la syllabe initiale. Est-ce que les enfants musiciens (G1) restitueront mieux la place de l'accent ?

Par ailleurs, il est probable qu'en répétition immédiate (RI), les accents soient mieux localisés qu'en répétition différée puisque les enfants répètent les mots tout de suite après le locuteur natif. En revanche, la répétition différée met en jeu la mémoire à long terme de l'enfant et il est possible que, dans cette situation, les interférences avec la LM soient plus puissantes. Le corpus est composé de 10 mots bisyllabiques et de 3 mots trisyllabiques (avoir annexe n°X). Dans les mots bisyllabiques, l'accent tombe sur la première syllabe comme dans le mot « *parrot* »⁵ par exemple (figure n°91 et **extrait sonore n°60**) ; dans les mots trisyllabiques, l'accent tombe sur la deuxième syllabe dans les mots trisyllabiques comme dans le mot « *eleven* »⁶ (figure n°92 et **extrait sonore n°61**).

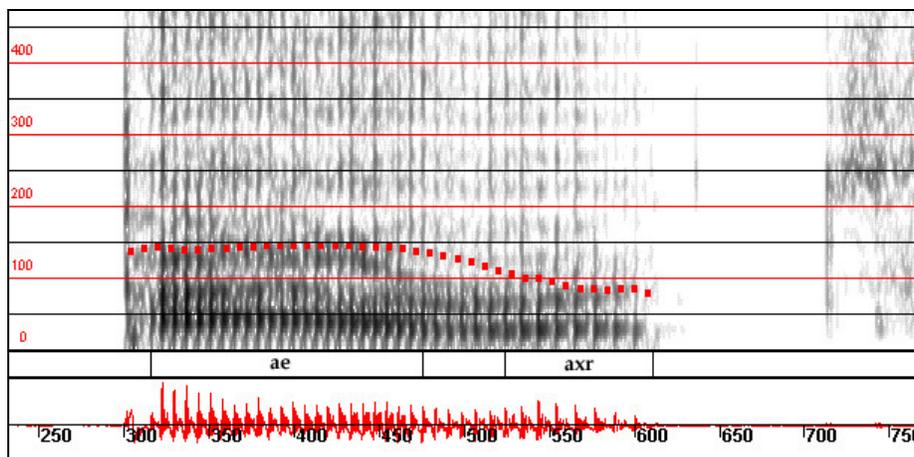


Figure n°91 : courbe mélodique affectant les mots « *parrot* », (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

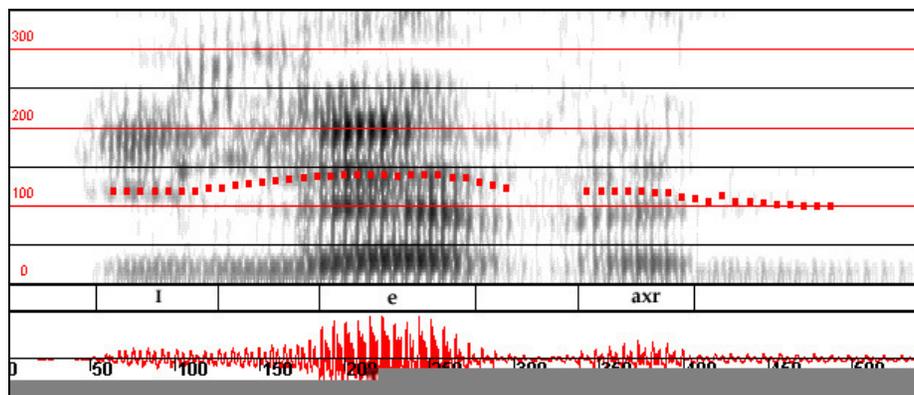


Figure n°92 : courbe mélodique affectant les mots « *eleven* », (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

⁵ Pic de Fo sur la voyelle [æ], d'une durée de 161 ms (la voyelle [ə] mesurant 74 ms) – Contour de Fo : 146 – 80 Hz.

⁶ Pic de Fo sur la voyelle [e], d'une durée de 139 ms (la voyelle [ɪ] mesurant 73,7 ms et la voyelle [ə], 67 ms) – Contour de Fo : 118-139-99 Hz.

Influencés par l'accentuation finale du français, les enfants pourront produire l'accent sur la deuxième syllabe dans les mots bisyllabiques. L'**extrait sonore n°62** nous fournit un exemple de ce type d'erreur, le mot « seven » étant accentué par l'enfant sur la deuxième syllabe. Pour la même raison, l'accent pourra être localisé sur la dernière syllabe des mots trisyllabiques, mais également sur la première. L'**extrait sonore n°63** nous fournit trois exemples de reproductions du mot « eleven », la première étant accentuée sur la première syllabe, la seconde, correcte, sur la deuxième syllabe et la troisième sur la dernière syllabe.

La variable « localisation de l'accent » comprend donc deux modalités pour les mots bisyllabiques : la modalité « première syllabe » (A) et la modalité deuxième syllabe (B) ; une troisième modalité se rajoute pour les mots trisyllabiques « troisième syllabe » (C). Nous posons comme hypothèse opérationnelle que la place de l'accent sera mieux localisée dans les productions des enfants musiciens. Le plan expérimental est exposé dans le tableau n°50.

Type de mot	Syllabe	G1		G2	
		RD	RI	RD	RI
Mots bisyllabiques	A				
	B				
Mots trisyllabiques	A				
	B				
	C				

Tableau n°50 : plan expérimental pour la variable « localisation de l'accent » (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, A pour première syllabe accentuée, B pour deuxième syllabe accentuée et C pour troisième syllabe accentuée).

Pour le groupe G1, la place de l'accent a été localisée dans 83 mots bisyllabiques, 17 n'ayant pu être analysés (soit 17 % de la totalité des énoncés) et pour le groupe G2, dans 122 mots bisyllabiques, 28 n'ayant pu être analysés (soit 22,95 % de la totalité des énoncés). Pour les mots trisyllabiques, la place de l'accent a été localisée dans 24 mots ont été analysés pour le groupe G1, 6 n'ayant pu être analysés (soit 25 % de la totalité des énoncés) et 40 mots pour le groupe G2, 5 n'ayant pu être analysés (soit 12,5%). Le nombre de mots analysés figure de manière détaillée dans le tableau n°51.

	G1		G2	
	RD	RI	RD	RI
Bisyllabiques	41	42	42	80
Trisyllabiques	4	20	18	22
Total	45	62	60	102

Tableau n°51 : nombre d'énoncés multisyllabiques analysés (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate).

2.1.4. Analyses acoustiques

Les analyses acoustiques ont été effectuées avec le logiciel Winsnoori⁷, la détection de la mélodie avec ce logiciel s'étant avérée particulièrement robuste pour l'analyse des voix d'enfants (certains enfants s'exprimant avec une voix très faible, d'autres avec des voix très aiguës) et dans les passages bruités (les enregistrements ayant eu lieu dans les écoles et non pas en chambre sourde).

Pour chaque mot du corpus, nous avons relevé :

- la fréquence correspondant à chaque point d'inflexion de la courbe mélodique (en Hz) ; à partir de ces mesures, les intervalles en ¼ de tons (Konopczynski et Mac Carthy, 1977) séparant deux points consécutifs de la courbe⁸ ont été déterminés ;
- la place de l'accent en combinant notre propre jugement auditif (notre oreille étant entraînée par la pratique musicale), à la localisation du pic de Fo et à la durée syllabique.

2.2. Résultats

2.2.1. Variable « Forme de la courbe »

Une fois analysés, tous les mots du corpus ont été répartis en fonction de leur appartenance à une catégorie spécifique de contours « cloche, cloche inversée, montée, descente, recto-ono », en fonction de leur initiale (« voisée ou non voisée »), en fonction du groupe (« G1, G2 ») et en fonction du type de discours (« RD, RI »). Une fois cette répartition achevée, la proportion des contours a été calculée en fonction du nombre total d'énoncés produits par les enfants dans chaque type de discours et dans chaque groupe. Cette proportion est exprimée en pourcentage (se reporter au Volume II, annexe n°9.1.2. pour les résultats exprimés en nombre d'énoncés). Les résultats sont exposés dans le tableau n°52 pour les énoncés dont l'initiale est non voisée (figure n°93 et n°94) et dans le tableau n°53 pour les énoncés dont l'initiale est voisée (figure n°95 et n°96).

⁷ Logiciel distribué par Babel Technologies (<http://www.babeltech.com>), mis au point par Yves Laprie, LORIA, Nancy. Version utilisée v.1.2. (1998-1999).

⁸ Les fréquences ont été converties en quart de tons selon la formule : $S = (\log(X) - \log(220)) / (\log(2)/24)$, X étant la valeur à convertir et 220, la fréquence de référence.

2.2.1.1. Mots dont l'initiale est non voisée

Initiale non voisée	G1		G2	
	RD	RI	RD	RI
« Cloche »	10	14	5,43	14,10
« Cloche inversée »	10	8	5,43	7,69
« Montée »	6,66	2	7,60	6
« Descente »	68,33	66	71,73	64,10
« Recto-tono »	5	10	9,78	10,25

Tableau n°52 : proportion de mots monosyllabiques commençant par un son non voisé répartis selon la forme de leur courbe mélodique, exprimée en pourcentage (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

Les résultats (tableau n°52) font apparaître une répartition semblable des différents types de contours pour G1 et G2 (figures n°93 et 94) en fonction des deux types de discours, RD et RI. La majorité des contours produits sont de type descendants (à plus de 60 % pour chacun des groupes dans chaque discours), ce qui correspond aux contours produits par le locuteur natif dans ce type de mots. Aucun effet notable de groupe, ni de discours n'a été relevé et l'ensemble des enfants de notre corpus restitue correctement le contour des mots monosyllabiques commençant par un son non voisé. En français, sur les mots monosyllabiques, les locuteurs ont également tendance à produire un contour descendant. Il est donc probable que, même s'il existait un effet de filtre prosodique du français vers l'anglais, il ne pourrait pas apparaître dans la restitution des contours descendants.

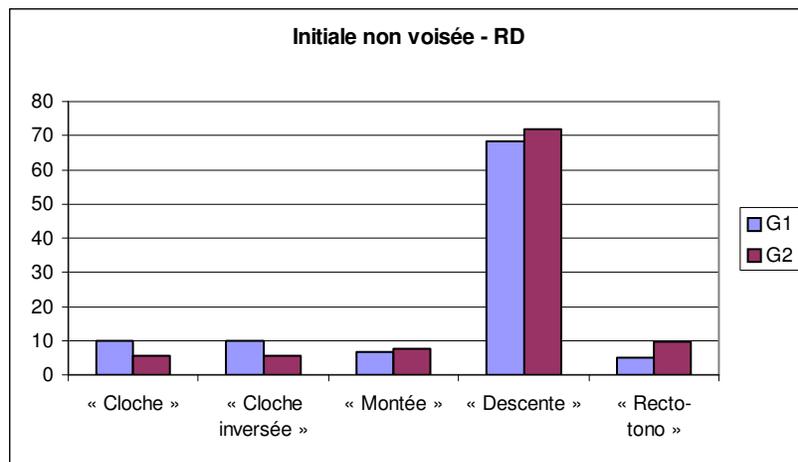


Figure n°93 : répartition en % des mots à l'initiale voisée selon la forme de leur contour mélodique en situation de répétition différée (RD) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

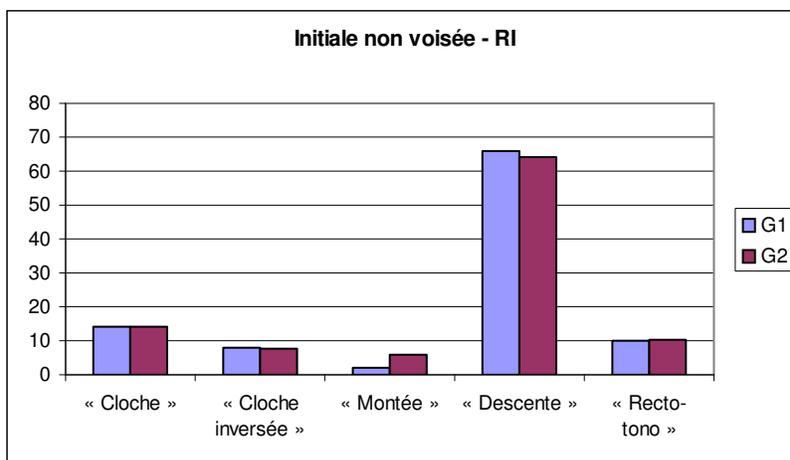


Figure n°94 : répartition en % des mots à l'initiale non voisée selon la forme de leur contour mélodique en situation de répétition immédiate (RI) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

Malgré les similarités, de légères différences entre les groupes peuvent être relevées. En RD, les enfants du groupe G2 produisent plus de contours recto-tono (9,78 %) que les enfants du groupe G1 (5 %). Par ailleurs, les enfants du groupe G2 produisent plus de contours descendants (71,73 %) et moins de contours en cloche (5,43 %) et en cloche inversée (5,43 %) que les enfants du groupe G1 (68,33 % de contours descendants ; 10 % de cloches et 10 % de cloches inversées). On peut se poser la question de savoir pourquoi les enfants musiciens produisent plus de contours en cloche sur les mots commençant par une initiale non voisée, alors qu'un contour descendant serait plus facile à réaliser et surtout, qu'il correspond mieux au modèle. Une réponse possible pourrait être que les enfants, ayant repéré les caractéristiques distinctives des contours en cloche de l'anglais, les appliquent plus souvent qu'il ne faudrait. De cette façon, ils doivent estimer que leurs énoncés sonnent « plus anglais ». Mais les différences sont cependant trop minimes pour que cette explication puisse être sûre.

En RI, les différences sont encore moins apparentes et les proportions de chaque catégorie sont remarquablement proches pour les deux groupes, à part le fait que les sujets de G2 (6 %) produisent plus de contours montants que les sujets du groupe G1 (2 %).

Si on ne relève pas de différences entre les productions des enfants des deux groupes en RI, une légère différence apparaît en RD allant en faveur de notre hypothèse initiale. Les enfants de G1 ont tendance à produire plus de contours en cloche que les enfants de G2 dans les mots à l'initiale non voisée, mais la différence n'est pas assez marquée pour vérifier notre hypothèse.

2.2.1.2. Mots dont l'initiale est voisée

<i>Initiale voisée</i>	G1		G2	
	RD	RI	RD	RI
« Cloche »	10	14	5,43	14,10
« Cloche inversée »	10	8	5,43	7,69
« Montée »	6,66	2	7,60	6
« Descente »	68,33	66	71,73	64,10
« Recto-tono »	5	10	9,78	10,25

Tableau n°53 : proportion d'énoncés de mots monosyllabiques commençant par un son voisé répartis selon la forme de leur courbe mélodique, exprimée en pourcentage (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

Comme dans les mots à initiale non voisée, la répartition des différents contours affectant les mots à initiale voisée est très proche pour les deux groupes en situation de discours différé (figure n°95 et tableau n°53). Comme précédemment, on relève une proportion de contours en cloche plus importante pour G1 (10 % de cloches et 10 % de cloches inversées) que pour G2 (5,43 % de cloches et 5,43 % de cloches inversées), ainsi qu'une proportion de contours recto-tono plus grande pour G2 (9,78 %) que pour G1 (5 %). Ces différences ne sont pas très importantes en terme de proportion, mais on les retrouve en RD dans les deux types de mots analysés.

Cependant, les résultats divergent entre les deux groupes en situation de répétition immédiate (figure n°96). Les résultats montrent que les courbes des enfants musiciens se répartissent de manière tranchée soit dans la catégorie « cloche » (46,05 %), soit dans la catégorie « descente » (42,10 %), à proportion relativement égale. Ils réussissent relativement bien à reproduire les contours en cloche caractéristiques de l'anglais, mais continuent à produire des contours descendants dans une forte proportion. Ces deux types de contours représentent 88,15 % de la totalité des énoncés produits par les enfants du groupe G1.

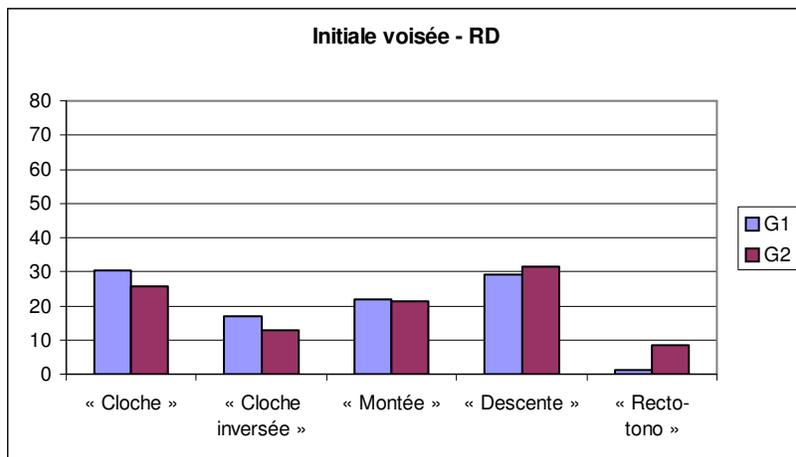


Figure n°95 : répartition en % des mots à l'initiale voisée selon la forme de leur contour mélodique en situation de répétition différée (RD) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

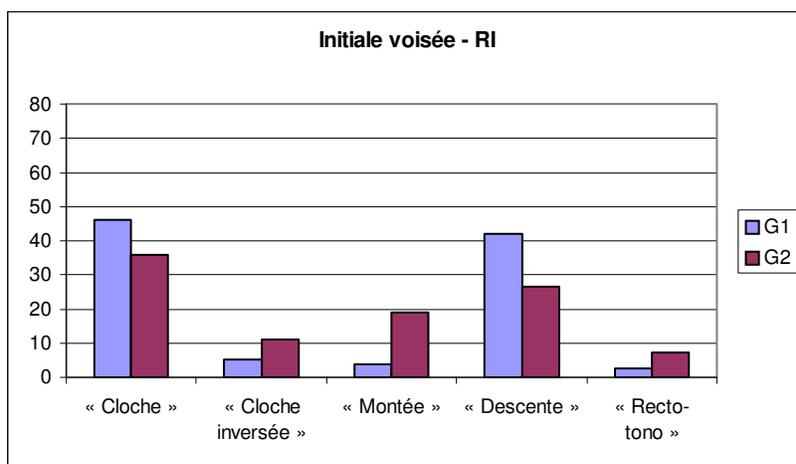


Figure n°96 : répartition en % des mots à l'initiale voisée selon la forme de leur contour mélodique en situation de répétition immédiate (RI) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

La répartition des courbes dans les différentes catégories est beaucoup plus variable chez les enfants non musiciens. Ils produisent moins de contours en cloche (36,02 %) et beaucoup moins de contours descendants (26,47 %). En revanche, ils produisent plus de contours montants que les enfants de G1 (19,11 % contre 3,94 %) et plus de cloches inversées (11,02 % contre 5,26 %). Il est étonnant de trouver des contours montants en situation de répétition immédiate. Il nous semble que, dans notre corpus, ces contours montants correspondent généralement à une demande de confirmation, comme si l'enfant demandait à son interlocuteur « *Est-ce que c'est bien ça ?* ». La partie finale des contours en cloche inversée est montante et correspond généralement, comme la montée simple, à une demande de confirmation.

La situation d'interaction semble donc jouer un rôle assez important lors de la production des contours montants. A la lumière de ces explications, il est étrange que les enfants du groupe G2 produisent plus de contours montants en situation de répétition immédiate. En effet, en répétition différée, les enfants peuvent ne pas se rappeler un mot et solliciter l'appui du locuteur natif pour qu'il les « rassure » en quelque sorte sur le choix du mot ; la montée exprime donc une incertitude. Mais, en répétition immédiate, le modèle est donné à l'enfant qui n'a plus qu'à le répéter. Cette proportion plus grande de contours montants pourrait traduire un « manque de confiance » à reproduire des mots inconnus ayant un contour complexe (plus difficile à restituer car non présent dans la LM).

Un autre point intéressant apparaît par rapport à la proportion de contours en cloche dans les deux situations de discours. Cette proportion est en effet plus élevée en RI (46,05 % pour G1 ; 36,02 % pour G2) qu'en RD (30,48% pour G1 et 25,64 % pour G2). Nous avons vu que le contour en cloche n'existe pas en français et qu'il pourrait être difficile à reproduire en raison du crible prosodique. Il serait plus facilement restitué à la suite du locuteur natif qu'en répétition différée, situation où l'effet de filtre agirait de manière plus marquée.

Notre hypothèse initiale se vérifie en situation de répétition immédiate dans les mots à l'initiale voisée, car les enfants musiciens produisent plus de contours en cloche que les enfants non musiciens. Et s'ils ne produisent pas de contours en cloche, ils produisent des contours descendants, mais pas de contours en cloche inversée, ni de contours montants. Leur comportement est donc plus tranché que celui des enfants non musiciens, révélant peut-être une plus grande « sécurité » à reproduire des événements sonores non habituels. Par ailleurs, la proportion de contours en cloche est plus élevée en RI qu'en RD pour les deux groupes, sûrement en raison du fait que les enfants reproduisent plus facilement les contours caractéristiques de l'anglais à la suite du locuteur natif.

2.2.2. Variable « Exagération de la courbe »

Après avoir calculé l'intervalle en $\frac{1}{4}$ de tons séparant le point initial du point final de chaque inflexion pour chacun des contours des énoncés (deux inflexions pour les contours en cloche et en cloche inversée ; une inflexion pour les contours montants et descendants), nous avons déterminé l'intervalle moyen pour chaque type d'inflexion.

Les résultats sont exposés dans le tableau n°54 et les figures n°97 et n°98 pour les contours en cloche et en cloche inversée et dans le tableau n°55 et les figures n°99 et n°100 pour les contours descendants et montants (voir également tableaux détaillés, avec la totalité des contours analysés, dans le volume II, annexe n°9.2.2.).

Contour	G1				G2			
	RD		RI		RD		RI	
	A	B	A	B	A	B	A	B
Cloche	4,18	5,89	4,37	7,15	3,87	7,87	3,35	7,33
Cloche inversée	5,99	5,57	5,52	3,60	5,65	4,25	5,36	4,30

Tableau n°54 : intervalles moyens correspondant à la première inflexion initiale et à la deuxième inflexion (B) des contours en cloche et en cloche inversée, exprimés en 1/4 de tons (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

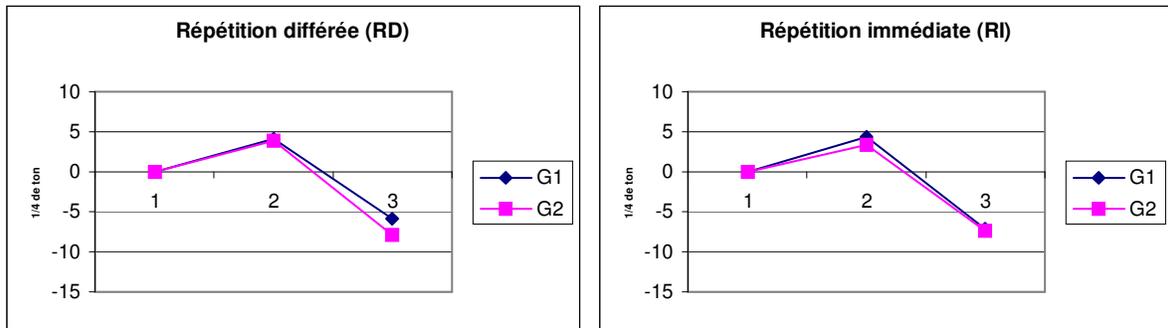


Figure n°97 : contour moyen « en cloche » en répétition différée (RD) et en répétition immédiate G1 pour « groupe musiciens », G2 pour « groupe non musiciens ».

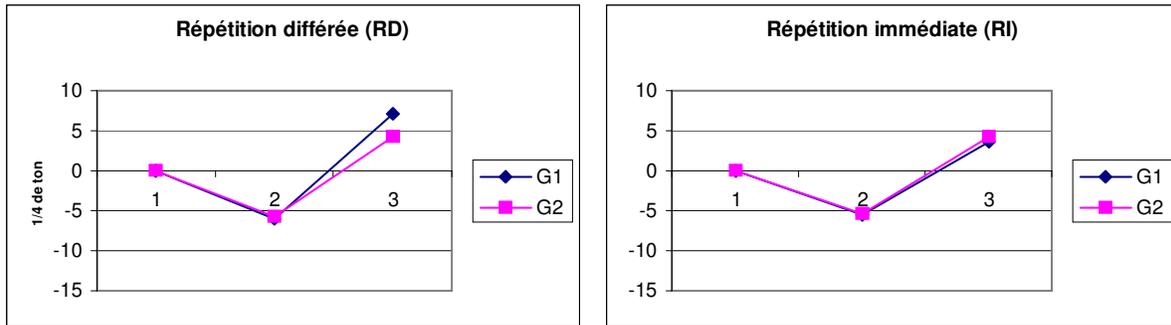


Figure n°98 : contour moyen « en cloche inversée » en répétition différée (RD) et en répétition immédiate pour G1 pour « groupe musiciens », G2 pour « groupe non musiciens ».

Contour	G1		G2	
	RD	RI	RD	RI
Descendant	7,06	6,14	5,56	7,08
Montant	6,38	4,53	5,44	4,17

Tableau n°55 : intervalles moyens correspondant aux contours montants et descendants, exprimés en 1/4 de tons (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

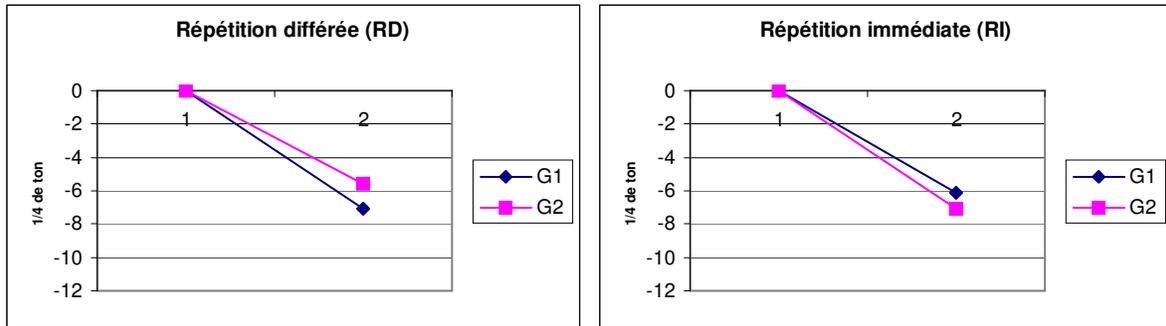


Figure n°99 : contour moyen « descendant » en répétition différée (RD) et en répétition immédiate pour G1, « groupe musiciens » et G2, « groupe non musiciens ».

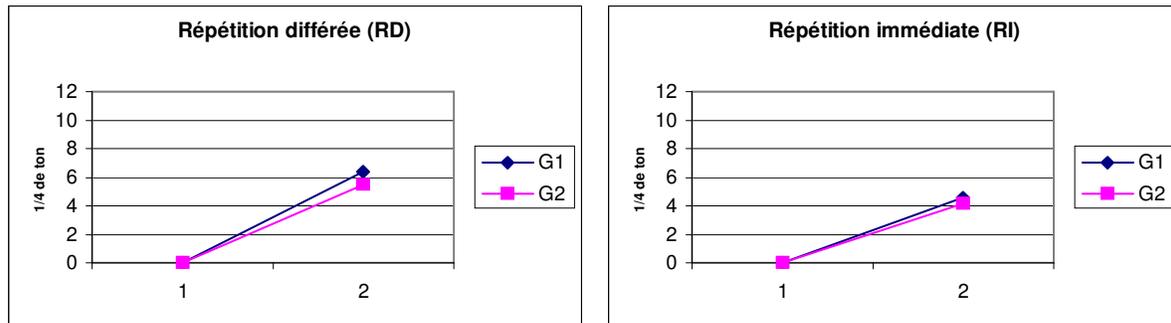


Figure n°100 : contour moyen « montant » en répétition différée (RD) et en répétition immédiate pour G1, « groupe musiciens » et G2, « groupe non musiciens ».

Une première analyse révèle que, pour l'ensemble des inflexions, les intervalles moyens les plus grands le sont beaucoup moins que ceux du locuteur natif. En effet, on relève pour ce locuteur une inflexion initiale de 8,21 1/4 de tons et une inflexion finale de 13,67 1/4 de tons pour le contour en cloche « moyen » (voir figure n°88) et une inflexion de 10,40 1/4 de tons pour le contour descendant « moyen » (voir figure n°89). Pour les enfants, le plus grand intervalle relevé est un intervalle moyen de 7,87 1/4 de tons pour la deuxième inflexion des contours en cloche (G2 en RD, voir tableau n°54). Ils ne produisent donc pas de contours plus amples que ceux du locuteur natif. Il est vrai que les moyennes lissent fortement les résultats et qu'elles font disparaître les particularités individuelles. Nous les avons donc complétées par le calcul des écarts-types pour les différents contours.

	G1				G2			
	RD		RI		RD		RI	
	A	B	A	B	A	B	A	B
Cloche	3,36	4,25	2,84	3,51	3,06	4,16	2,41	3,50
Cloche inversée	3,277	5,23	2,05	2,25	3,02	2,80	3,977	3,40

Tableau n°56 : écarts-types correspondant à la première inflexion initiale et à la deuxième inflexion (B) des contours en cloche et en cloche inversée, exprimés en 1/4 de tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

	G1		G2	
	RD	RI	RD	RI
Descendant	3,40	2,97	3,776	4,27
Montant	4,40	3,55	3,30	3,59

Tableau n°57 : écarts-types correspondant aux contours montants et descendants, exprimés en ¼ de tons (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

Pour le locuteur natif, l'écart-type est de 4,95 pour l'inflexion initiale et de 6,18 pour l'inflexion finale pour les contours en cloche, et de 5,24 pour les contours descendants. Chez les enfants, les écarts-types sont beaucoup plus réduits que ceux du locuteur natif : entre 2,05 et 3,97 pour l'inflexion initiale et entre 2,25 et 4,16 pour les contours en cloche et en cloche inversée (voir tableau n°56) ; entre 2,97 et 4,27 pour les contours descendants et entre 3,59 et 4,40 pour les contours montants (voir tableau n°57). Ces écarts-types réduits révèlent que, s'il y a des contours exagérés, ils ne sont pas nombreux et que la tendance générale des groupes est à la modération de l'ampleur de la courbe mélodique. Effectivement, si on regarde les résultats détaillés figurant dans les tableaux de l'annexe n°9.1.2. (Volume II), on peut voir que, pour l'inflexion initiale des contours en cloche par exemple, l'intervalle le plus élevé est de 12,05 ¼ de tons (soit un intervalle approximatif de quarte augmentée) et qu'il n'y a que 7 productions sur un total de 63 qui dépassent un intervalle de 10 ¼ de tons (soit un intervalle d'une quarte juste), tous types de discours et groupes confondus, alors que locuteur natif dépasse 8 fois l'intervalle de 10 ¼ de tons sur un total de 37 productions et qu'il va jusqu'à produire un intervalle de 21 ¼ de tons (soit un intervalle approximatif d'une septième mineure). Le comportement d'exagération ne caractérise aucun des deux groupes, il ne caractérise donc pas le « groupe musiciens ». Il semble plutôt relever de stratégies employées par quelques individus appartenant aussi bien à G1 qu'à G2. Par ailleurs, sur les représentations graphiques exposées dans les figures n°97, 98, 99 et 100, on peut remarquer que les différences entre les groupes sont minimales pour l'ensemble des résultats et que les « contours moyens » se superposent presque parfaitement, quel que soit le type de contour et quel que soit le type de discours.

On peut toutefois relever quelques différences, même si celles-ci sont légères. Par exemple, la deuxième inflexion du contour en cloche en RD est légèrement plus ample pour G2 que pour G1, mais seulement de 1,98 ¼ de tons (soit approximativement un demi-ton). A l'inverse, la deuxième inflexion du contour en cloche inversée en RD est légèrement plus ample pour G1 que pour G2, mais seulement de 1,32 ¼ de tons (voir tableau n°54). Les différences sont légèrement plus importantes pour les contours montants et descendants.

Ainsi, en ce qui concerne les contours descendants, en RD, l'intervalle moyen de G1 est plus grand d'1,5 ¼ de tons que celui du groupe G2 (voir tableau n°55). Nous ne continuerons pas plus avant, ces différences n'étant pas pertinentes. On peut faire la même remarque par rapport aux deux situations de discours.

En ce qui concerne les écarts-types, on peut relever des différences plus intéressantes entre les groupes. Ainsi, pour les enfants du groupe G1, l'écart-type est plus élevé en situation de RD qu'en situation de RI et ce, dans les quatre types de contours (tableau n°56 et n°57). Initialement, nous avons posé l'hypothèse que les musiciens exagéreraient plus souvent les contours en RD qu'en RI. Les écarts-types révèlent qu'il y a plus de contours amples en situation de RD, même s'ils ne sont pas assez nombreux pour affecter la configuration du contour moyen. Pour les enfants du groupe G2, les résultats sont moins cohérents et il n'y a pas d'uniformité pour les contours en cloche et en cloche inversée. En revanche, pour les contours montants et descendants, la situation est inversée par rapport à G1 : les écarts-types sont plus élevés en RI qu'en RD, ce qui veut dire qu'on relève plus de productions amples chez les enfants du groupe G2 en répétition immédiate pour ces deux catégories de contour.

Notre hypothèse initiale n'est pas vérifiée et l'amplitude de la courbe mélodique ne semble pas affectée par la variable indépendante « musique », mais il semble cependant qu'en situation de répétition différée, les enfants musiciens produisent des contours plus variables qu'en situation de répétition immédiate. Il serait donc particulièrement intéressant de procéder à une étude qualitative des productions des enfants, pour déterminer si certains d'entre eux manifestent des stratégies d'exagération des éléments prosodiques. Mais une telle étude ne concernerait que quelques enfants et la tendance générale qui caractérise notre échantillon est à la production de contours peu amples sans distinction de groupe. Même en situation de discours immédiat, les enfants ne reproduisent pas l'amplitude des contours caractéristiques du locuteur natif. Pourtant, ils perçoivent correctement les contours prosodiques, comme le suggèrent les résultats obtenus aux tests de perception dans l'étude n°1 (chapitre VII). En réalité, la stratégie d'exagération pourrait être plutôt liée à la personnalité de l'apprenant. En effet, nous savons que les apprenants francophones adultes n'osent pas exagérer les contours en anglais, par peur du « ridicule », car la production de contours très amples n'entre pas dans leurs habitudes prosodiques maternelles. Ils hésitent à jouer avec les paramètres prosodiques devant les autres élèves dans une classe de langue par exemple, alors qu'ils le font plus facilement individuellement.

De la même façon, le fait de « mettre l'accent » sur une expression anglaise à l'intérieur d'un discours en français est généralement mal toléré, jugé comme prétentieux. Il semble donc que les « préjugés linguistiques » exercent une influence considérable sur le comportement prosodique des locuteurs francophones et que ces préjugés semblent déjà affecter les apprenants âgés de 7 ans. Il serait intéressant d'étudier ce phénomène avec des apprenants plus jeunes afin de déterminer si cette « timidité prosodique » se renforce avec l'âge.

2.2.3. Variable « Localisation de l'accent »

2.2.3.1. Mots bisyllabiques

La proportion d'accents bien localisés, c'est-à-dire placés sur la première syllabe (A), est plus élevée chez les enfants du groupe G1 en RD (73,17 % en RD, voir tableau n°58) que chez les enfants du groupe G2 (63,46 % en RD). Pour RI, cette différence se réduit et la proportion des productions correctes avoisine les 90 % pour les deux groupes (95,23 % pour G1 et 91,25 % pour G2). Les enfants montrent donc d'excellentes capacités, qu'ils soient musiciens ou non, à restituer la place de l'accent en situation de répétition immédiate. En revanche, en situation de répétition différée, comme nous l'avions supposé au départ, la tâche paraît être beaucoup plus difficile, sûrement à cause de la mise en jeu de la mémoire à long terme. En répétition immédiate, l'enfant peut solliciter sa mémoire à court terme et oublier le mot quelques instants après l'avoir prononcé. Les effets de l'apprentissage musical semblent donc se manifester plutôt dans les tâches faisant appel à la mémoire à long terme qu'à la mémoire à court terme. Notre hypothèse initiale se vérifie donc uniquement en situation de RD, situation où l'accent est mieux localisé dans les productions des enfants musiciens que dans celles des enfants non musiciens.

Bisyllabiques	G1		G2	
	RD	RI	RD	RI
A	30	40	33	73
B	11	2	19	7
% A	73,17	95,23	63,46	91,25
% B	26,82	4,76	36,53	8,75

Tableau n°58 : nombre et proportion (exprimée en pourcentage) de mots bisyllabiques classés en fonction de la localisation de l'accent (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

2.2.3.2. Mots trisyllabiques

Dans les mots trisyllabiques de notre corpus, l'accent est localisé sur la deuxième syllabe. Les résultats sont assez contrastés selon les deux groupes (voir tableau n°59), mais nous prendrons cependant beaucoup de précautions en interprétant ces résultats, le nombre d'énoncés étant trop réduit pour en tirer des tendances généralisables. Ainsi en RD, nous n'avons pu relever que 4 énoncés pour G1. Nous n'en tiendrons donc pas compte. Pour le reste et à l'inverse des résultats concernant les mots bisyllabiques, les enfants du groupe G2 se montrent très bons pour restituer l'accent en situation de RI (82,60 % des mots correctement accentués contre 65 % pour G1). En situation de RD, les résultats sont nettement moins bons pour les deux groupes, plus mauvais là-aussi pour G1 (38,88 % des mots correctement accentués) que pour G2 (50 %). Lorsqu'ils se trompent, les enfants de G2 ont tendance à accentuer la première syllabe (50 % de mots accentués sur la première syllabe) et rarement la troisième syllabe (11,11 % de mots accentués sur la troisième syllabe). On constate la même chose pour le groupe G1 en situation de RI (35 % de mots accentués sur la première syllabe), ceux-ci ne produisant en revanche jamais d'accent en finale (0 % de mots accentués sur la troisième syllabe). Ces derniers résultats sont intéressants car ils révèlent que la position finale de l'accent en français ne semble pas influencer les résultats des enfants lorsqu'ils reproduisent des mots en anglais. Bien au contraire, leurs erreurs de localisation concernent la première syllabe. Or, nous savons que l'anglais connaît une tendance à l'accentuation en initiale de mot (cf. chapitre V, 2.1.1.1.). Ainsi, même s'ils se trompent, leurs énoncés « sonnent » quand même anglais. Cependant ces arguments ne peuvent rester qu'à l'état d'hypothèses, le nombre d'énoncés étant trop réduit (65 énoncés au total).

Trisyllabiques	G1		G2	
	RD	RI	RD	RI
A	1	7	9	2
B	2	13	7	19
C	1	0	2	2
% A	25	35	50	8,69
% B	50	65	38,88	82,60
% C	25	0	11,11	8,69

Tableau n°59 : nombre et proportion (exprimée en pourcentage) de mots trisyllabiques classés en fonction de la localisation de l'accent (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

On peut quand même conclure que pour les mots bi et trisyllabiques, le taux de réussite est beaucoup plus élevé en situation de RI qu'en situation de RD. La place de l'accent serait donc plus difficile à restituer en situation de discours différé.

Quoiqu'il en soit, notre hypothèse initiale ne se vérifie pas pour les mots trisyllabiques car les enfants du groupe G2 localisent mieux l'accent que les enfants du groupe G1.

3. Etude n°4 « Production des noyaux vocaliques »

3.1. Plan expérimental

3.1.1. Hypothèse

Nous posons l'hypothèse particulière « Contrastes vocaliques » qu'une oreille entraînée par la pratique musicale induit une meilleure production des contrastes vocaliques d'une langue étrangère.

3.1.2. Variables indépendantes

3.1.2.1. La musique : deux modalités

La variable indépendante « musique » est composée de deux modalités : le groupe d'enfants « musiciens » G1 et le groupe d'enfants « non musiciens » G2 (voir 2.1.2.1.).

3.1.2.2. Le discours : deux modalités

La variable indépendante « discours » est composée de deux modalités : discours immédiat RI et discours différé RD (voir 2.1.2.2.).

3.1.3. Variables dépendantes

3.1.3.1. Timbre du noyau vocalique

Nous l'avons vu, l'apprenant tente de se rapprocher du système cible. Cette tâche est entravée par les problèmes de crible phonologique. Dans l'étude n°2 sur la perception des contrastes vocaliques, nous avons vu que les enfants qui avaient une oreille entraînée par la pratique musicale se montraient moins sensibles au phénomène de crible phonologique que les enfants qui n'avaient pas d'oreille entraînée (chapitre VII). Est-ce que cette capacité à percevoir les contrastes vocaliques non natifs de l'anglais va également engendrer une meilleure production de ces contrastes, en vertu des liens étroits qui existent entre perception et production ?

Pour le savoir, nous étudierons la qualité des différents contrastes vocaliques produits par les enfants dans notre corpus et nous les comparerons aux voyelles et aux diphtongues produites par le locuteur natif de façon à déterminer s'il existe une différence entre les deux groupes et dans les deux situations de discours (l'effet de filtre pouvant être plus fort en RD qu'en RI).

Dans le chapitre V (3.1.3.2.3.), nous avons choisi de caractériser le timbre d'une voyelle par les rapports entre ses trois premiers formants. Le fait de travailler avec des rapports et non pas avec des fréquences absolues résout le problème de normalisation. En effet, à cause de la variabilité inter locuteurs, il est très difficile de comparer une voix d'homme, celle de notre locuteur natif, à des voix d'enfants. Cette variabilité est principalement due à une différence de taille entre le tractus vocal de l'adulte, et celui, plus court, de l'enfant. Cependant, l'utilisation des rapports entre la fréquence centrale des trois premiers formants peut réduire et presque éliminer les différences entre locuteurs (cf. travaux exposés dans le chapitre V). Ce mode de calcul nous permet de comparer les productions du locuteur natif à celles des enfants de notre échantillon.

La figure n°101 nous donne la « charte des rapports de formants » du locuteur natif (pour la description de cette représentation et sa lecture, se reporter au Volume II, annexe n°3.2.) et le tableau n°60, les valeurs des trois premiers formants et les intervalles F2-F1 et F3-F2 pour chaque voyelle représentée sur la charte.

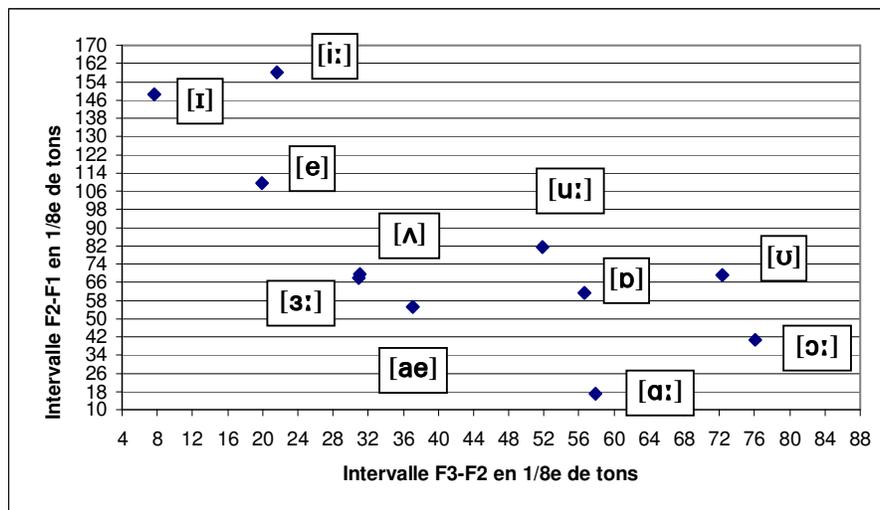


Figure n°101 : charte des rapports de formants du locuteur anglophone natif.

	F1	F2	F3	F1-F3	F2-F3
[ʊ]	343	931	2646	69,14	72,33
[æ]	686	1519	2597	55,04	37,13
[ɑ:]	923	1188	2722	17,01	57,88
[ʌ]	588	1568	2450	67,92	30,9
[ɜ:]	539	1470	2303	69,47	31,08
[ɪ]	343	2940	3283	148,7	7,64
[ɒ]	497	1208	2737	61,5	56,63
[i:]	245	2401	3283	158,05	21,66
[e]	462	2239	2986	109,29	19,93
[ɔ:]	490	882	2646	40,7	76,07
[u:]	392	1274	2695	81,62	51,88

Tableau n°60 : valeurs des trois premiers formants en Hz et des rapports F2-F1 et F3-F2 en 1/8^e de tons pour les 11 voyelles prononcées par le locuteur anglophone natif.

D'après les résultats de l'étude n°2 (chapitre VII), nous posons comme hypothèse opérationnelle que la capacité des enfants musiciens à mieux discriminer le timbre des contrastes vocaliques de l'anglais va engendrer une meilleure reproduction des contrastes vocaliques de l'anglais, c'est-à-dire des voyelles et des diphtongues, en raison des liens qui existent entre la perception et la production. La variable « timbre du noyau vocalique » comprend deux modalités : la modalité « voyelles » et la modalité « diphtongues ».

3.1.3.2. Durée du noyau vocalique

Nous avons posé comme hypothèse dans le chapitre 2.1.3.2. que les enfants musiciens, habitués à manipuler les éléments musicaux, pouvaient reporter cette stratégie dans la restitution de la prosodie d'une langue étrangère, en exagérant l'amplitude des contours mélodiques. Or, s'ils exagèrent ces contours, ils vont passer plus de temps à les réaliser et comme le noyau vocalique porte le contour mélodique, il devrait être également allongé. Non seulement, la durée absolue (désormais DA) du noyau sera plus élevée, mais sa durée relative (désormais DR) au sein de la syllabe sera plus grande.

Nous avons vu dans le chapitre V qu'il était possible de dresser une typologie des noyaux vocaliques de l'anglais en fonction de leur durée intrinsèque (3.1.3.2.1.). Cette typologie permet en particulier de différencier les voyelles brèves des voyelles longues. Quant aux diphtongues, elles sont encore plus longues en raison d'une trajectoire articulatoire beaucoup plus complexe (chapitre V, 3.2.2.1.1.).

La variable « durée du noyau vocalique » comprend donc trois modalités, la modalité « voyelles brèves », la modalité « voyelles longues » et la modalité « diphtongues » et nous posons comme hypothèse opérationnelle que la durée absolue et la durée relative des voyelles longues et des diphtongues seront plus élevées chez les enfants musiciens et spécialement en situation de répétition différée (RD). Il est probable que la différence entre les groupes soit moins marquée pour les voyelles brèves, mais celles-ci devraient cependant être mieux restituées par les enfants musiciens, c'est-à-dire que leur durée absolue et relative sera plus proche de celle des voyelles brèves prononcées par un locuteur natif. Le plan expérimental est exposé dans le tableau n°61.

	G1		G2	
	RD	RI	RD	RI
Voyelles brèves				
Voyelles longues				
Diphtongues				

*Tableau n°61 : plan expérimental pour la variable « durée du noyau vocalique »
(tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens »,
RD, répétition différée, RI, répétition immédiate).*

3.1.4. Mesures acoustiques

3.1.4.1. Mesures de timbre

Pour chaque voyelle de notre corpus, la fréquence des trois premiers formants F1, F2 et F3 a été relevée grâce à une coupe spectrale LPC et une coupe cepstrale éditée avec le logiciel Winsnoori (voir note n°7). Les coupes ont été appliquées au centre de la partie stable des voyelles.

Pour chaque diphtongue, la fréquence des trois premiers formants a été relevée sur chacune des deux cibles vocaliques. Les coupes ont été appliquées au centre de la partie stable de la cible, ou s'il n'y a pas de partie stable sur le point d'inflexion de la trajectoire des formants.

Les valeurs de F1, F2 et F3 ont ensuite été converties en huitièmes de tons selon la formule :

$$S = (\log(X) - \log(220)) / (\log(2) / 48)$$

X étant la valeur en Hz à convertir et 220, la fréquence de référence (fréquence choisie car elle est inférieure à la fréquence la plus basse que l'on puisse trouver pour F1 ; les chiffres obtenus après conversion seront donc toujours positifs).

Une fois la conversion faite, les intervalles en huitièmes de tons entre F1 et F2 (F2-F1) et entre F2 et F3 (F3-F2) ont été calculés.

Les 7 voyelles brèves et les 5 voyelles longues de l'anglais ont été analysées au sein de 35 mots de notre corpus (mots mono, bi et trisyllabiques) et ce, pour les productions des 25 enfants qui composent notre échantillon. 1225 prises de mesures ont ainsi été réalisées (au total, 49 occurrences différentes réparties dans 35 mots pour 25 enfants). La répartition des mots du corpus par voyelle étudiée est donnée dans le Volume II (annexe n°8.2.3.). Pour les noyaux vocaliques complexes, 5 diphtongues ont été analysées au sein de 10 mots de notre corpus (mots mono, bi et trisyllabiques). 275 prises de mesures ont été réalisées (au total 11 occurrences différentes réparties dans 10 mots pour 25 enfants).

Seuil de perception

Nous avons choisi les huitièmes de tons en fonction des seuils de perception donnés par la littérature. Il existe peu de travaux à ce sujet et ces seuils concernent souvent la variation d'un seul formant et non la variation simultanée de plusieurs formants. Ce seuil varie de 3 à 5 % de la fréquence du formant, mais une variation aussi large que 13 % a été démontrée par Nakagawa et al. (1982). Mermelstein (1978) a étudié la variation simultanée de la fréquence des formants à partir d'un seul point référence. Il relève un seuil de 3 % pour F2. Kewley-Port (1990) rapporte un seuil plus réduit, de 1 à 2 % pour la variation d'un seul formant et Hawks (1994), un seuil de 1,9 %. A titre d'exemple, par rapport à une valeur de 588 Hz correspondant au premier formant de la voyelle [ʌ], une variation de fréquence de 1 % représente une variation de $0,68 \frac{1}{8}^{\text{e}}$ de tons (593,88 Hz), une variation de fréquence de 2% représente une variation de $1,35 \frac{1}{8}^{\text{e}}$ de tons (599,76 Hz) et une variation de fréquence de 3 % représente une variation de $2,04 \frac{1}{8}^{\text{e}}$ de tons (605,64 Hz). Ces variations très faibles justifient le recours aux huitièmes de tons pour le calcul des intervalles entre les trois premiers formants. Mais ces auteurs utilisent une méthodologie qui place l'auditeur dans des conditions les plus favorables pour une tâche de discrimination. Les conditions d'écoute en parole naturelle sont plus dégradées, car soumises aux perturbations dues à l'environnement naturel.

De l'avis même de Hawks (1994 : 1080),

« discriminability of vowels in natural speech communication modes should be based on larger perceptual units ».

Par ailleurs, Hawks montre que le seuil de perception du changement simultané de la fréquence de plusieurs formants est plus petit que pour un seul formant, mais seulement pour certains patterns. La perception de la variation simultanée de plusieurs formants semble plus naturelle que la variation d'un seul formant :

« in natural speech, however, we would not anticipate shifting of only a single formant, but rather, the simultaneous shifting of multiple formants » (Hawks, 1994 : 1074).

Enfin, la discrimination de variations parallèles est significativement meilleure que pour des variations opposées et que pour la variation d'un seul formant (Hawks, 1994). Pour les mouvements parallèles, l'information provenant des formants séparés se combine pour améliorer la performance (Mermelstein, 1978). Il est donc probable que le seuil de perception de la variation des trois premiers formants au cours de la partie stable soit très réduit.

3.1.4.2. Mesures de durée

La durée absolue (DA exprimée en ms) de chaque noyau vocalique et de chaque syllabe a été relevée dans les 30 mots monosyllabiques qui composent le corpus (voir liste des mots dans le Volume II, annexe n°8.2.2.) et ce, pour les productions des 27 enfants de notre échantillon. 1620 prises de mesure ont ainsi été réalisées, dont 810 relevés de durée du noyau et 810 relevés de durée syllabique. Les étiquettes qui délimitent chacun des segments étudiés (vocaliques et syllabiques) ont été posées en couplant la représentation oscillographique à la représentation spectrographique à bandes larges du logiciel Winsnoori. La durée relative de la voyelle a été obtenue en calculant le pourcentage de la durée de la voyelle en fonction de la durée totale de la syllabe (DR exprimée en %).

3.2. Résultats

3.2.1. Timbre du noyau vocalique

Pour les productions des enfants, les résultats seront présentés par son, mais également par mot, car les contrastes vocaliques apparaissent au sein de contextes phonétiques très différents, mais surtout, par ce que chaque mot a un statut différent dans l'apprentissage.

Ainsi, des mots qui ressemblent aux mots français comme « *orange* » ou « *banana* » auront tendance à être prononcés à la française, c'est-à-dire avec un [ɔ] à l'initiale et la voyelle nasale [ɑ̃] en syllabe finale. Des mots très courants comme « *cat* » ou « *dog* » sont déjà connus des enfants et très souvent prononcés à la française dans leur environnement. En revanche, les mots inconnus posent moins de problèmes d'interférences. Les moyennes par son gommeraient les particularités qui nous intéressent dans le « parcours d'apprentissage » des enfants.

3.2.1.1. Voyelles

3.2.1.1.1. Voyelles brèves

[ɪ] : en syllabe accentuée, cette voyelle est très fermée (rapport F2-F1 pouvant atteindre 150 1/8^e de tons chez le locuteur natif (« *pink*, *eleven* »). Elle est plus fermée que la voyelle française de timbre correspondant [i] (120 1/8^e de tons). En syllabe non accentuée, la voyelle est beaucoup plus ouverte (F2-F1 = 100 1/8^e de tons), comme dans « *orange* » et « *rabbit* » (même aperture que [e]).

Pour les enfants, on ne relève pas de différences entre les groupes en RD ; ils produisent une voyelle équivalente [i] (120 1/8^e de tons). Il y a donc **assimilation** sans distinction de groupe. En revanche, en RI, G2 réalise une voyelle légèrement plus ouverte (113 1/8^e de tons) ; on relève également beaucoup de variabilité dans la réalisation de l'intervalle F2-F1 ($\sigma = 21,36$ pour G1 et $\sigma = 21,46$ pour G2 en RI ; $\sigma = 19,07$ pour G1 et $\sigma = 22,48$ pour G2 en RD). Cette variabilité reflète la différence d'aperture qui affecte les productions des enfants pour les deux groupes et dans les deux situations de discours.

Les résultats détaillés par mot font apparaître que les enfants, sans distinction de groupe ni de discours, produisent un [i] dans les mots « *six* », « *pink* » et « *fish* ». On relève une grande variabilité dans l'intervalle F2-F1 dans « *six* » ($\sigma = 22,28$ pour G1 et $\sigma = 23,22$ pour G2) et dans « *pink* » pour G1 ($\sigma = 20,27$). Dans « *teddy* », en RI, ils produisent une voyelle également proche de [i] quoique légèrement plus fermée pour G1 (136 1/8^e de tons). Si cette voyelle est aussi fermée dans une syllabe accentuée, c'est qu'ils ne reproduisent pas correctement la rythmique anglaise de ce mot. Dans « *baby* » en RI, G1 produit une voyelle équivalente à [i], mais G2, une voyelle plus légèrement plus ouverte (109 1/8^e de tons).

On note beaucoup plus de variabilité dans la réalisation de cette voyelle pour l'intervalle F2-F1 ($\sigma = 21,03$ pour G1 et $\sigma = 29,36$ pour G2) et pour l'intervalle F3-F2 pour G2 ($\sigma = 9,15$ pour G1 et $\sigma = 22,14$ pour G2). Dans « rabbit », en RI, la voyelle est légèrement plus ouverte (110 $1/8^\circ$ de tons pour G1 et 114 $1/8^\circ$ de tons pour G2), plus proche de la production du locuteur natif. On relève pour les deux groupes une grande variabilité de l'intervalle F2-F1 ($\sigma = 25,99$ pour G1 et $\sigma = 21,57$ pour G2). Dans le mot « eleven », la voyelle est plus fermée qu'un [i] français pour les deux groupes en RD (130 $1/8^\circ$ de tons pour G1 et 135 $1/8^\circ$ de tons pour G2), qu'en RI (106 $1/8^\circ$ de tons pour G1 et 99 $1/8^\circ$ de tons pour G2) où la voyelle se rapproche du [e] français (**assimilation**). Pour G2 en RD, on relève par ailleurs une grande variabilité dans l'intervalle F2-F1 ($\sigma = 14,19$ pour G1 et $\sigma = 24,03$ pour G2). Enfin, dans le mot « orange », les enfants des deux groupes produisent une voyelle de timbre similaire à la voyelle française [i] (116-21 $1/8^\circ$ de tons pour G1 et 116-23 $1/8^\circ$ de tons pour G2). En revanche, on note une différence entre les groupes en RI, où, tandis que les enfants de G1 produisent une voyelle de même timbre qu'en RD (126-23 $1/8^\circ$ de tons), les enfants de G2 produisent une voyelle moins fermée, mais surtout plus postérieure (87-44 $1/8^\circ$ de tons). La variabilité qui affecte l'intervalle F2-F1 est plus élevée en RD ($\sigma = 25,39$ pour G1 et $\sigma = 23,12$ pour G2) qu'en RI ($\sigma = 18,97$ pour G1 et $\sigma = 18,18$ pour G2).

Ces résultats montrent que la voyelle anglaise [ɪ] est le plus souvent assimilée à la voyelle française [i] sans distinction de groupe, ni de discours et ce, même dans les syllabes non accentuées où la voyelle devrait être plus centralisée.

[ʊ] : au sein de notre corpus, il n'y a qu'une occurrence de cette voyelle dans le mot « book ». Le LN produit une voyelle très postérieure (F3-F2 : 72 $1/8^\circ$ de tons) et une apertures aussi marquée que pour la voyelle française [ɔ].

Dans les productions des enfants, on relève beaucoup plus de variabilité pour les deux intervalles en situation de RD. Dans cette situation, G1 réussit à produire un [ʊ] (F2-F1 : 64 $1/8^\circ$ de tons et F3-F2 : 71 $1/8^\circ$ de tons), tandis que G2 produit une voyelle légèrement plus antérieure (F3-F2 : 64 $1/8^\circ$ de tons) et plus ouverte (F2-F1 : 61 $1/8^\circ$ de tons), mais qui reste proche malgré tout de [ʊ]. En situation de RI, G1 produit un son de timbre similaire. En revanche, G2 produit une voyelle aussi postérieure que [ʊ] (F3-F2 : 70 $1/8^\circ$ de tons), mais plus ouverte (F2-F1 : 54 $1/8^\circ$ de tons), qui la rapproche de la voyelle française [o].

Par ailleurs, on relève beaucoup de variabilité dans l'intervalle F2-F1 pour G2 ($\sigma = 14,62$ pour G1 et $\sigma = 22,26$ pour G2), mais ce qui est plus frappant, une grande variabilité dans l'intervalle F3-F2 pour les deux groupes, spécialement pour G1 ($\sigma = 23,80$ pour G1 et $\sigma = 18,03$ pour G2). Cette variabilité, qui n'est pas habituelle pour l'intervalle F3-F2 sur l'ensemble des noyaux vocaliques de notre corpus, montre que le lieu d'articulation de la voyelle [ʊ] varie beaucoup dans les productions des enfants.

Ces résultats montrent que les enfants ne produisent pas l'assimilation avec la voyelle française [u], et que, s'il y a une différence de timbre, elle va dans le sens d'une exagération de type [bɒk] qui « sonne anglais ».

[ʌ] : dans le mot « duck », le locuteur natif produit une voyelle centrale (F2-F1 : $67 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $30 \frac{1}{8}^{\circ}$ de tons).

Les moyennes générales des enfants indiquent qu'en RD, G1 et G2 produisent une voyelle de timbre proche (pas de crible) et en RI, une voyelle légèrement plus postérieure (avec un intervalle F3-F2 plus élevé, $49,49 \frac{1}{8}^{\circ}$ de tons pour G1 et $44,43 \frac{1}{8}^{\circ}$ de tons pour G2 ; timbre proche de la voyelle française [ɔ]).

Dans le détail des trois mots qui contiennent cette voyelle, la voyelle de « one » est correctement reproduite pour G1 et G2 en RD (pas d'assimilation). En revanche, dans le mot « duck », en situation de RI, G1 produit une voyelle un peu plus fermée que la voyelle française [ɔ] (F2-F1 : $71 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $43 \frac{1}{8}^{\circ}$ de tons) et G2 une voyelle proche de la voyelle française [ɔ] (F2-F1 : $59 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $45 \frac{1}{8}^{\circ}$ de tons). Dans l'**extrait sonore n°64**, l'enfant n°6 produit ce type d'assimilation c'est-à-dire [dɒk] en situation de répétition immédiate, alors que l'enfant n°4 produit le mot correctement, c'est-à-dire [dʌk]. La prononciation du mot « umbrella » est un peu particulière pour le locuteur natif puisque celui-ci produit une voyelle plus ouverte et plus postérieure que le [ʌ] produit dans « duck » (F2-F1 : $40 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $49 \frac{1}{8}^{\circ}$ de tons), dont le timbre est proche de [ɑ] français. En RI, les enfants de G1 produisent une voyelle plus postérieure que celle du LN, dont le timbre est proche de la voyelle anglaise [ɒ] (F2-F1 : $56 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $56 \frac{1}{8}^{\circ}$ de tons).

Les enfants de G2 produisent une voyelle proche de [ɔ], peut-être [ɔ̃] (F2-F1 : 66 1/8° de tons et F3-F2 : 42 1/8° de tons), à cause du mot français « ombrelle » (**extrait sonore n°65** en RI, pas de correction, finale de mot prononcée « à la française » : « LN : *umbrella* – E1 et E2 : *umbrelle* – LN : *umbrella* – E1 : *umbrelle* ; *c'est une ombrelle ou un parapluie ? LN : *umbrella*, repeat – E2 : *umbrelle** »). Pour les deux groupes, la variabilité de l'intervalle F2-F1 est élevée, en particulier pour G1 ($\sigma = 25,02$ pour G1 et $\sigma = 18,63$ pour G2).

[ɒ] : dans les mots « orange » et « doll », le locuteur natif produit la voyelle [ɒ] de même aperture, mais plus antérieure que la voyelle française [o] (F2-F1 : 48 1/8° de tons et F3-F2 : 60 1/8° de tons). Dans le mot « dog », la voyelle est légèrement plus fermée, mais localisée au même endroit sur la charte des rapports de formants (F2-F1 : 61 1/8° de tons et F3-F2 : 56 1/8° de tons).

En RD, les enfants des deux groupes produisent des voyelles de timbre proche, voire similaire à [ɒ]. En RI, G2 produit le même type de voyelle et G1, une voyelle légèrement plus antérieure (F2-F1 : 69 1/8° de tons et F3-F2 : 48 1/8° de tons). On ne relève donc pas d'effet de filtre au profit de la voyelle française [ɔ].

Dans le détail des mots, la voyelle initiale [ɒ] de « orange » est correctement reproduite par les deux groupes en RD et en RI, par G2. En revanche, on relève un effet de filtre en faveur de la voyelle française [ɔ] pour G1. Pour les deux groupes, on relève une grande variabilité de l'intervalle F2-F1 ($\sigma = 24,98$ pour G1 et $\sigma = 18,34$ pour G2). Dans le mot « doll », la voyelle est correctement reproduite par les deux groupes en RI, ainsi qu'en RD pour G1. C'est cette fois G2 en RD qui produit une voyelle encore plus fermée que la voyelle française [ɔ] (F2-F1 : 76 1/8° de tons et F3-F2 : 39 1/8° de tons). On relève également une grande variabilité de l'intervalle F2-F1 pour les deux groupes ($\sigma = 21,04$ pour G1 et $\sigma = 22,11$ pour G2). Dans le mot « dog », l'effet de filtre est massif en RD et en RI pour les deux groupes (plus fermé que la voyelle [ɔ] pour G1 et légèrement plus postérieure que [ɔ] pour G2). Alors que le mot « doll » n'est pas courant et pas connu des enfants, le mot « dog » est très répandu. L'assimilation doit provenir de la prononciation « à la française » de ce mot.

[æ] : dans « rabbit » et « apple », la voyelle [æ] produite par le locuteur natif est au centre de l'axe F3-F2 (F2-F1 : 55 1/8° de tons et F3-F2 : 37 1/8° de tons pour « rabbit » ; F2-F1 : 38 1/8° de tons et F3-F2 : 36 1/8° de tons pour apple », légèrement plus ouverte). La voyelle de « parrot » est plus antérieure et se rapproche du [a] français dans « parrot ».

En moyenne, les enfants des deux groupes et dans les deux types de discours produisent la même voyelle que le locuteur natif dans « rabbit » et « apple ».

Dans le détail des mots, on ne relève pas de différence entre les productions des deux groupes pour « black » en RD, « cat » en RD, « rabbit » en RI. L'**extrait sonore n°66** nous donne un exemple de production du mot « cat » en répétition différée par deux enfants du groupe G2 (enfants n°16 et n°23). La même voyelle est produite par G1 en RD dans « apple ». En revanche, dans le mot « apple », la voyelle est beaucoup plus ouverte, proche de la voyelle française [ɑ] pour les deux groupes en situation de RI (F2-F1 37 1/8° de tons et F3-F2 : 44 1/8° de tons pour G1 ; F2-F1 : 45 1/8° de tons et F3-F2 : 43 1/8° de tons pour G2). Pour « parrot », on relève une différence entre les groupes : tandis que G1 produit une voyelle proche de [æ], G2 produit une voyelle proche de la voyelle française [ɑ].

[e] : on relève des différences de timbre pour cette voyelle chez le locuteur natif ; alors qu'elle est très fermée et très antérieure dans un mot comme « red » (F2-F1 109 1/8° de tons et F3-F2 : 19 1/8° de tons), elle est plus ouverte dans « eleven » (F2-F1 92 1/8° de tons et F3-F2 : 25 1/8° de tons, entre les deux voyelles françaises [e] et [ɛ]), elle est proche du [ɛ] français dans le mot « umbrella ».

Dans les productions des enfants, le timbre est similaire pour les deux groupes et dans les deux situations de discours. Le timbre de cette voyelle se situe entre les deux voyelles françaises [e] et [ɛ], comme la voyelle produite dans « eleven » par le locuteur natif.

Dans le détail des mots, on retrouve cette voyelle sans distinction de groupe ni de discours dans « ten », « yellow », « seven », avec peu de variabilité. On la retrouve également dans « red » en RD pour les deux groupes, avec une plus forte variabilité dans l'intervalle F2-F1 (σ =17,36 pour G1 et σ =25,23 pour G2). En revanche, le groupe G2 produit une voyelle proche de [ɛ] en RI ; il y a donc **assimilation**.

Dans le mot « twelve », la voyelle est proche de celle produite par le locuteur natif dans le même mot, c'est-à-dire plus postérieure et aussi ouverte que [ɛ]. On relève une forte variabilité dans l'intervalle F2-F1 pour G2 en RI ($\sigma = 14,15$ pour G1 et $\sigma = 20,28$ pour G2). En RI, dans « teddy », les deux groupes produisent une voyelle aussi fermée que le [e] du locuteur natif dans « red », sans doute en raison d'une harmonisation vocalique due à la présence de la voyelle [ɪ] en deuxième syllabe. Si de plus, cette deuxième syllabe est accentuée comme elle le serait en français, le phénomène d'harmonisation est renforcé. Dans « umbrella », la voyelle est également aussi fermée pour G1 en RI. En revanche pour G2, elle se situe entre les deux voyelles françaises [e] et [ɛ], comme dans la production du locuteur natif. Enfin, dans le mot « eleven », la voyelle est très fermée en RD pour G1 et G2, surtout si la première voyelle est correctement prononcée, c'est-à-dire [ɪ] (influence fermante). En RI cependant, la voyelle est plus ouverte, localisée au même endroit que la voyelle française [œ].

[ə] : la voyelle [ə] varie beaucoup chez le locuteur natif et son timbre est fortement influencé par le contexte phonétique dans lequel elle apparaît. Ainsi, dans des mots comme « purple » et « apple », elle est très postérieure, entre la voyelle anglaise [ɒ] et la voyelle [o], dans « apple » (F2-F1 : $64 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $66 \frac{1}{8}^{\circ}$ de tons) et elle est localisée au même endroit que la voyelle [o] dans « purple ». Cette postériorité provient de la présence de la latérale [l], qui provoque également une labialisation de la voyelle. Dans des mots comme « banana » et « umbrella », la voyelle [ə] est très centralisée (F2-F1 : $67 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $36 \frac{1}{8}^{\circ}$ de tons pour « umbrella » par exemple). Dans « parrot », la voyelle est légèrement plus fermée et plus antérieure, proche de la voyelle française [ɛ] (influence de l'articulation de l'occlusive dentale [t]). Dans le mot « eleven », la voyelle est aussi fermée que [ə] et aussi antérieure que [ɛ] (F2-F1 : $99 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $26 \frac{1}{8}^{\circ}$ de ton), sûrement influencée par les voyelles fermées se trouvant dans les deux syllabes précédentes.

En situation de répétition différée, les enfants des deux groupes produisent l'équivalent de la voyelle française [œ] (F2-F1 : $80 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $36 \frac{1}{8}^{\circ}$ de ton pour G1 et F2-F1 : $81 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $35 \frac{1}{8}^{\circ}$ de ton pour G2) ; il y a donc un effet de filtre assez marqué (voir **extrait sonore n°67**, « seven » prononcée par l'enfant n°6 du groupe G1).

En situation de répétition immédiate, la voyelle est plus ouverte et légèrement plus postérieure, proche de la voyelle française [ɔ] (F2-F1 : 66 1/8° de tons et F3-F2 : 40 1/8° de ton pour G1 et F2-F1 : 59 1/8° de tons et F3-F2 : 49 1/8° de ton pour G2) ; par rapport à la situation de discours précédente, l'effet de crible est donc moins marqué.

On retrouve cette voyelle dans les mots « parrot », « apple » et « purple » dont toutes les occurrences sont prononcées en RI. Il est à noter dans les deux derniers mots une très forte variabilité de l'intervalle F2-F1 pour G2 ($\sigma = 30,48$ dans « apple » et $\sigma = 22,15$ dans « purple »). En RI, dans « umbrella », le timbre de la voyelle est proche de la voyelle [æ] (F2-F1 : 55 1/8° de tons et F3-F2 : 37 1/8° de ton pour G1 et F2-F1 : 48 1/8° de tons et F3-F2 : 36 1/8° de ton pour G2) ; à cause de l'accentuation finale du français, les enfants doivent prononcer cette voyelle de façon pleine (on trouve aussi pour ce mot l'influence du mot français « ombrelle », **extrait sonore n°65**). Dans « seven », c'est la voyelle [œ] qui est produite ; il y a donc assimilation de timbre, d'autant plus marquée si l'enfant produit l'accentuation en finale comme en français. La première voyelle du mot « banana » est proche de [ɔ] en RI (F2-F1 : 64 1/8° de tons et F3-F2 : 34 1/8° de ton pour G1 et F2-F1 : 63 1/8° de tons et F3-F2 : 43 1/8° de ton pour G2). La seconde voyelle est produite avec le même timbre par G2. En revanche, elle est plus proche de [ʌ] pour G1 (F2-F1 : 70 1/8° de tons et F3-F2 : 35 1/8° de ton), mais également extrêmement variable (F2-F1 : $\sigma = 29,62$, ce qui représente l'un des écart-type les plus élevés de l'ensemble de nos mesures, avec celui de la voyelle [ə] de « apple »).

En RD, dans « eleven », la voyelle est proche de [œ] pour les deux groupes (F2-F1 : 88 1/8° de tons et F3-F2 : 33 1/8° de ton pour G1 et F2-F1 : 86 1/8° de tons et F3-F2 : 34 1/8° de ton pour G2) ; il en est de même en RI pour G2. En revanche, dans la même situation de discours, G1 produit une voyelle plus fermée dont le timbre se situe entre les deux voyelles françaises [e] et [ø] (F2-F1 : 97 1/8° de tons et F3-F2 : 27 1/8° de ton). Une forte variabilité affecte l'intervalle F2-F1 pour G1 et G2 ($\sigma = 18,64$ pour G1 et $\sigma = 22,58$ pour G2).

3.2.1.1.2. Voyelles longues

[i:] : le locuteur natif produit une voyelle avec un rapport F3-F2 moins important que la voyelle la plus antérieure de l'anglais [ɪ], mais avec un rapport G2-F1 beaucoup plus élevé, comme dans « green » (158-21 1/8^e de tons), « sheep » (164-13 1/8^e de tons) et « thirteen » (162-17 1/8^e de tons).

Dans les productions des enfants, on relève un effet de filtre massif en direction de la voyelle française [i] sans distinction de groupe, ni de discours (118-20 1/8^e de tons pour G1 et 121-24 1/8^e de tons pour G2 en RD ; 122-17 1/8^e de tons pour G1 et 127-18 1/8^e de tons pour G2).

Dans le détail des mots, c'est dans le mot « sheep » que la voyelle est produite avec le rapport F2-F1 le plus élevé (132-20 1/8^e de tons pour G1 et 125-17 1/8^e de tons pour G2 en RD), mais même pour G1, son timbre reste proche de la voyelle française [i]. Dans le mot « green » en RD, la voyelle produite est un [i] pour les deux groupes (voir **extrait sonore n°68**, « green » prononcé par l'enfant n°16 du groupe G2). En revanche, pour « green » en RI et dans tous les autres mots, c'est-à-dire « three » en RD, « thirteen » en RD et en RI, on relève une différence entre G1 et G2. En effet, les enfants de G1 produisent une voyelle plus ouverte que le [i] français qui se rapproche de la voyelle française [e] (pour « green » en RI, 110-23 1/8^e de tons pour G1 et 131-19 1/8^e de tons pour G2), alors que les enfants de G2 produisent un [i]. Par ailleurs, on relève une forte variabilité de l'intervalle F2-F1 pour G2 dans « green » en RI ($\sigma = 20,10$) et dans « thirteen » en RD ($\sigma = 25,14$). Dans « three » en RD et « thirteen » en RI, cette variabilité concerne les deux groupes ($\sigma = 20,52$ pour G1 et $\sigma = 19,73$ pour G2 dans « three » ; $\sigma = 25,85$ pour G1 et $\sigma = 19,05$ pour G2 dans « thirteen »).

[ɜ:] : dans les mots « bird », « thirteen » et « purple », le locuteur natif produit une voyelle légèrement plus postérieure et plus ouverte que la voyelle française [ɛ] (69-31 1/8^e de tons pour « bird » ; 78-30 1/8^e de tons pour « thirteen » ; 76-30 1/8^e de tons pour « purple »).

Dans les productions des enfants, on relève un effet de filtre massif en direction de la voyelle française [œ], c'est-à-dire plus fermée que la voyelle anglaise [ɜ:] produite par le locuteur natif (81-38 1/8^e de tons pour G1 en RD ; 82-34 1/8^e de tons pour G1 et 86-33 1/8^e de tons pour G2). En RD, G2 produit une voyelle encore plus fermée, proche de la voyelle française [ø] (94-31 1/8^e de tons).

Dans le détail des mots, on relève le même effet de filtre pour « bird » en RI, pour « thirteen » en RI. En RD, pour le même mot, on note une différence entre les deux groupes : tandis que G2 produit une voyelle proche de la voyelle française [ø] (101-33 1/8^e de tons), G1 produit une voyelle de même aperture que [ɜ:], mais plus postérieure (71-43 1/8^e de tons) (voir **extrait sonore n°69**, « bird » prononcé par l'enfant n°6 du groupe G1). En RI, dans le mot « purple », on remarque une différence de même type entre les deux groupes : G2 produit une voyelle équivalente à la voyelle française [œ] (81-37 1/8^e de tons), alors que G1 produit la même voyelle que dans « thirteen » en RD (71/41 1/8^e de tons). Enfin, en RD, G1 produit une voyelle proche de la voyelle française [e] (102-33 1/8 de tons) (pas d'occurrences de ce mot en RD pour G2).

[ɑ:] : dans les mots « car », le locuteur produit une voyelle proche de [æ] (58-38 1/8^e de tons) et dans le mot « banana », une voyelle plus fermée, proche de [ʌ] (70-36 1/8^e de tons dans « banana »). Dans le chapitre V, la locutrice anglophone avait produit une voyelle plus ouverte et plus postérieure (32-51 1/8^e de tons dans « banana »), proche de la voyelle française [ɑ] produite par la locutrice francophone (36-45 1/8^e de tons).

Dans les productions des enfants, la voyelle [ɑ:] n'apparaît qu'en situation de RI. Son aperture est équivalente à celle du locuteur natif, mais elle est plus postérieure (67-43 1/8^e de tons pour G1 et 59-42 1/8^e de tons pour G2). Dans le mot « car » et dans le mot « banana », on ne relève aucune différence entre les groupes, mais une différence de timbre entre ces deux mots, similaire à celle produite par le locuteur natif (57-46 1/8^e de tons pour G1 et 51-41 1/8^e de tons pour G2 dans « car » ; 79-38 1/8^e de tons pour G1 et 66-42 1/8^e de tons pour G2 dans « banana »). Dans chacun de ces deux mots, l'aperture est identique à celle du locuteur natif, mais la voyelle est aussi postérieure que la voyelle française [ɑ]. Dans « banana », on relève cependant une grande variabilité de l'intervalle F2-F1 pour G1 ($\sigma = 21,91$ pour G1 et $\sigma = 16,40$ pour G2).

[ɔ:] : dans le mot « horse », le locuteur natif réalise la voyelle la plus postérieure et la plus ouverte de toutes ses productions (40-76 1/8^e de tons) ; sur la charte des rapports de formants, elle est localisée à l'extrême droite, tout en bas de la représentation.

Pour les productions des enfants, nous exposerons directement les résultats par mot car toutes les occurrences en RD sont produites avec le mot « four » et toutes les occurrences en RI, avec le mot « horse ». Dans le mot « horse », les enfants produisent une voyelle légèrement plus postérieure et plus ouverte que la voyelle française [ɔ] (49-52 1/8^e de tons pour G1 et 51-53 1/8^e de tons pour G2), et encore plus postérieure dans « four » (49-60 1/8^e de tons pour G1 et 49-58 1/8^e de tons pour G2), mais jamais aussi postérieure que la voyelle produite par le locuteur natif. Dans ce mot, l'intervalle F2-F1 est assez variable pour G1 ($\sigma = 20,06$ pour G1 et $\sigma = 12,75$ pour G2). L'**extrait sonore n°70** nous donne un exemple d'une très bonne reproduction du mot « horse » par un des enfants du groupe G1 (voir **extrait sonore n°70**, enfant n°5).

[u:] : dans le mot « blue », le locuteur natif produit une voyelle de même aperture, mais plus antérieure (81-51 1/8^e de tons) que la voyelle française [u] (locutrice francophone chapitre V, 86-62 1/8^e de tons).

Les deux mots « blue » et « two » n'apparaissent dans notre corpus qu'en situation de discours différé. Dans les productions des enfants, on ne relève aucune différence entre les groupes (72-48 pour G1 et 71-50 pour G2). Le timbre de la voyelle est proche de celui de la voyelle produite par le locuteur natif (voir **extrait sonore n°71**, « two » prononcé par les enfants n°4 et 6 du groupe G1). Le détail des mots ne révèle pas de différence et c'est toujours la même voyelle qui est produite (75-48 1/8^e de tons pour G1 et 72-52 1/8^e de tons pour G2 dans « two » ; 74-48 1/8^e de tons pour G1 et 71-48 1/8^e de tons pour G2 dans « blue »). On relève une forte variabilité dans l'intervalle F2-F1 pour G1 par rapport à G2 ($\sigma = 24,47$ pour G1 et $\sigma = 13,12$ pour G2) dans le mot « two ».

3.2.1.2. Diphtongues

Les résultats exposés dans le détail des mots ne seront pas traités en fonction de la variable « discours » puisque chaque mot contenant une diphtongue dans le corpus n'apparaissait que dans une seule situation de discours, soit en répétition différée, soit en répétition immédiate.

[aɪ] : dans le mot « nine », le locuteur natif produit une diphtongue dont le timbre de la première cible correspond à la voyelle [æ] (46-34 1/8^e de tons) et le timbre de la seconde cible, à la voyelle [i:] (160-14 1/8^e de tons), voyelle très fermée et très antérieure.

Comme pour le locuteur natif, les enfants des deux groupes produisent la première cible de la diphtongue [aɪ] avec un timbre proche de la voyelle [æ], et ce, dans les deux situations de discours (47-38 1/8^e de tons pour G1 et 54-37 1/8^e de tons pour G2 en RD ; 55-44 1/8^e de tons pour G2 en RI), bien qu'en situation de RI, les enfants du groupe G1 produisent une voyelle un peu plus postérieure (55-44 1/8^e de tons). La seconde cible est beaucoup moins fermée que celle du locuteur natif. Ainsi, dans les deux situations de discours, les enfants du groupe G1 produisent en moyenne une voyelle avec un timbre proche de la voyelle française [i] (122-21 1/8^e de tons en RI et 124-20 1/8^e de tons en RD), tandis que les enfants produisent une voyelle légèrement plus ouverte, proche de la voyelle française [e] (106-23 1/8^e de tons en RI et 114-22 1/8^e de tons en RD). Pour G2, on note une grande variabilité dans l'intervalle F2-F1 pour la première cible en situation de RI ($\sigma = 10,59$ pour G1 et $\sigma = 25,24$ pour G2 en RI) et pour la seconde cible dans les deux situations de discours ($\sigma = 10,38$ pour G1 et $\sigma = 28,28$ pour G2 en RI ; $\sigma = 13,35$ pour G1 et $\sigma = 23,01$ pour G2 en RD). Les cibles sont donc atteintes avec plus de précision par les enfants du groupe G1.

Dans le détail des mots « five » et « nine », on ne relève pas de différences entre les groupes. Ainsi dans ces deux mots, la première cible de la diphtongue correspond à la voyelle [æ] (47-38 1/8^e de tons pour G1 et 53-37 1/8^e de tons pour G2 dans « five » ; 47-37 1/8^e de tons pour G1 et 55-37 1/8^e de tons pour G2 dans « nine » en RD). Dans le mot « white », le timbre de la première cible est cependant légèrement plus postérieur, proche de la voyelle [ɒ] pour G1 (55-49 1/8^e de tons) et proche de la voyelle française [ɔ] pour G2 (54-39 1/8^e de tons). En ce qui concerne la deuxième cible, dans le mot « nine », son timbre correspond à celui de la voyelle française [i] pour les deux groupes (128-20 1/8^e de tons pour G1 et 119-20 1/8^e de tons pour G2). En revanche, on relève la même différence entre les groupes dans les mots « five » et « white » que celle qui apparaissait au niveau des moyennes générales. Les enfants du groupe G1 produisent une voyelle plus fermée, proche de la voyelle française [i] (122-21 1/8^e de tons pour G1 dans « white » et 122-20 1/8^e de tons pour G1 dans « five »).

Dans le mot « white », les enfants du groupe G2 produisent un intervalle F2-F1 très variable dans la première cible ($\sigma = 10,59$ pour G1 et $\sigma = 25,24$ pour G2). On relève la même variabilité pour la seconde cible ($\sigma = 10,38$ pour G1 et $\sigma = 28,28$ pour G2).

Ainsi, on trouve des productions allant de la simple voyelle, qui correspond à la voyelle [æ] (**extrait sonore n°72**, enfant n°15 ; 32-48 1/8° de tons) à une très bonne reproduction de la diphtongue (**extrait sonore n°73**, enfant n°16 ; 27-53 1/8° de tons pour la première cible et 115-27 1/8° de tons pour la seconde cible). Pour G1, les productions se rapprochent de celle de l'enfant n°6, qui réalise la meilleure production (**extrait sonore n°74** ; 38-52 1/8° de tons pour la première cible et 126-15 1/8° de tons pour la seconde cible). En revanche, dans les mots « five » et « nine », la différence entre les groupes est nettement moins marquée, même si on observe une plus grande variabilité pour G2.

[ei] : dans les mots « grappe » et « baby », le locuteur natif produit des diphtongues semblables et dont la seconde cible est très fermée, aussi fermée qu'un [i] et aussi antérieure qu'un [i:] (respectivement 136-22 1/8° de tons et 140-23 1/8° de tons). Il est normal que le point d'arrivée soit aussi fermé car la cible de départ est elle-même relativement fermée (proche de la voyelle anglaise [e] ; 101-14 1/8° de tons dans « grappe » 101-20 1/8° de tons dans « baby ») et il faut une trajectoire minimale pour que la diphtongue soit perçue en tant que telle.

Dans les productions des enfants, la première cible est similaire pour les deux groupes et dans les deux situations de discours. Son timbre est approximativement localisé entre la voyelle française [e] et la voyelle française [ɛ] (97-28 1/8° de tons pour G1 et 103-25 1/8° de tons pour G2 en RI ; 84-27 1/8° de tons pour G1 et 90-23 1/8° de tons pour G2 en RD). La seconde cible est proche du timbre de la voyelle française [i] pour les deux groupes en RD (118-22 1/8° de tons pour G1 et 123-23 1/8° de tons pour G2), ainsi que pour G2 en RI (127-23 1/8° de tons). Elle est en revanche légèrement plus antérieure pour G1 en RI avec un intervalle F3-F3 plus faible (124-15 1/8° de tons). En ce qui concerne la variabilité dans les productions des enfants, elle n'est pas très marquée sauf pour l'intervalle F2-F1 pour G1 en RD ($\sigma = 24,66$ en RD et $\sigma = 14,75$ en RI) et pour G2 en RD ($\sigma = 22,10$ en RD et $\sigma = 16,64$ en RI).

Dans le détail des mots, la diphtongue de « eight » est réalisée par tous les enfants. Les enfants du groupe G2 produisent une première cible dont le timbre se situe entre ceux des deux voyelles françaises [ɛ] et [e] (90-23 1/8° de tons), tandis que celui produit par les enfants du groupe G1 est plus proche de la voyelle française [ɛ] (84-27 1/8° de tons).

On relève plus de variabilité pour l'intervalle F2-F1 pour G1 que pour G2 ($\sigma = 24,66$ pour G1 et $\sigma = 15,74$ pour G2), à cause de la production divergente d'un des enfants, l'enfant n°8 (38-64 pour la première cible et 99-20 1/8^e de tons pour la seconde cible. Les autres enfants produisent des diphtongues très marquées comme l'enfant n°5 (38-64 pour la première cible et 99-20 1/8^e de tons pour la seconde cible ; **extrait sonore n°75**) et l'enfant n°7 (112-17 pour la première cible et 132-17 1/8^e de tons pour la seconde cible ; **extrait sonore n°76**). Dans cet extrait, la production de cet enfant est d'ailleurs précédée de celle, toute aussi bonne, d'un des enfants du groupe G2, l'enfant n°15 (110-11 pour la première cible et 129-19 1/8^e de tons pour la seconde cible). En ce qui concerne le timbre de la seconde cible, il correspond à celui de la voyelle française [i] pour G2 (123-23 1/8^e de tons) et il est légèrement plus ouvert pour G1, se situant entre le timbre des deux voyelles françaises [e] et [i] (118-22 1/8^e de tons). On relève plus de variabilité pour l'intervalle F2-F1 pour G2 ($\sigma = 16,17$ pour G1 et $\sigma = 22,10$ pour G2).

Si la diphtongue était plus ou moins correctement réalisée dans le mot « eight », il n'en va pas de même dans les mots « baby », « cake » et « grappe », où les enfants ont tendance à produire une voyelle à la place de la diphtongue. Cette tendance est particulièrement marquée dans le mot « grappe » (85,71 % de voyelles pour G1 et 100 % de voyelles pour G2), dans le mot « cake » (71,42 % pour G1 et 76,92% pour G2) et un peu moins dans le mot « baby » (33,33 % pour G1 et 57,14 % pour G2). Cette tendance est d'autant plus étonnante que ces trois mots soient produits en RI, c'est-à-dire tout de suite après le locuteur natif, alors que « eight » était produit en RD. Il semble que les effets de filtre peuvent être importants en RI, surtout si les mots sont assimilés à un mot français leur ressemblant. C'est le cas de « cake », terme utilisé en française pour désigner une sorte de gâteau ; il est prononcé [kɛk] (**extrait sonore n°77** ; dans l'**extrait n°78**, l'enfant n°16 produit également une voyelle, mais dont la durée est très allongée) On trouve le même type d'assimilation avec le mot « grappe » qui ressemble au mot français équivalent « grappe » prononcé [grap] (enfant n°1, **extrait sonore n°79**, en RD). En revanche, dans un mot comme « baby », il n'y a pas de correspondance avec le français si ce n'est avec le mot « bébé », mais celui-ci est très différent de la prononciation anglaise. Cependant, le mot « baby » est très utilisé par les français et on peut l'entendre très souvent prononcé de la manière suivante : [bebi]. La fréquence de cette prononciation a pu influencer la production des enfants, même en RI (enfant n°4, à la fin de l'**extrait sonore n°80**).

Si nous nous intéressons au timbre de cette voyelle, comment est-elle réalisée dans les trois mots ? Dans « baby » et dans « cake », elle équivaut à la voyelle française [e] (99-22 1/8° de tons pour G1 et 96-25 1/8° de tons pour G2 dans « baby » ; 105-22 1/8° de tons pour G1 et 105-18 1/8° de tons pour G2 dans « cake »), certains enfants produisant la voyelle [ɛ] (dans ce cas, l'assimilation au français est totale, comme dans l'exemple n°X). L'intervalle F2-F1 est relativement variable dans le mot « baby » (intervalle F2-F1 : $\sigma = 18,34$ pour G1 et $\sigma = 20,30$ pour G2) et dans le mot « cake » pour G1 (intervalle F2-F1 : $\sigma = 21,27$ pour G1 et $\sigma = 10,61$ pour G2). Cette variabilité s'explique par le fait que les enfants produisent soit une voyelle de timbre correspondant à [ɛ], soit une voyelle de timbre correspondant à [e], la différence entre les deux voyelles étant essentiellement une différence d'aperture (ce qui affecte l'intervalle F2-F1).

Enfin, dans le mot « grappe », le timbre de la voyelle est légèrement plus fermé, entre les voyelles françaises [e] et [i] (113-19 1/8° de tons pour G1 et 113-20 1/8° de tons pour G2). Comme dans les deux mots précédents, on relève une forte variabilité (intervalle F2-F1 : $\sigma = 18,34$ pour G1 et $\sigma = 20,30$ pour G2). Dans ce mot, il n'y a qu'un enfant qui produise une diphtongue, l'enfant n°2 appartenant au groupe G1 (**extrait sonore n°79**) et cette production est très proche de celle du locuteur natif (106-17 1/8° de tons pour la première cible et 139-13 1/8° de tons pour la seconde cible). Elle suit la production de l'enfant n°1, [grap]. Il est à noter que seuls 13 enfants sur 25 ont produit ce mot, dont 10 en G1 et 3 en G2). En ce qui concerne les diphtongues réalisées dans « cake », la première cible est proche de la voyelle française [e] dans les deux groupes (113-18 1/8° de tons pour G1 et 101-23 1/8° de tons pour G2) et la seconde cible est proche de la voyelle anglaise [ɪ] (136-14 1/8° de tons pour G1 et 137-13 1/8° de tons pour G2). On relève peu de variabilité sur la première cible, alors que la seconde cible est plus dispersée (intervalle F2-F1 : $\sigma = 17,46$ pour G1 et $\sigma = 16,87$ pour G2). Lorsque la diphtongue est réalisée, elle est correctement réalisée et proche de productions du locuteur natif. Dans le mot « baby », la diphtongue n'est pas produite exactement de la même façon selon les groupes. Ainsi, G1 produit une première cible dont le timbre est proche du timbre de la voyelle française [œ] et une seconde cible dont le timbre se situe entre ceux des deux voyelles françaises [e] et [i] (86-35 1/8° de tons pour la première cible et 115-16 1/8° de tons pour la seconde cible).

G2 produit une première cible dont le timbre est proche de celui de la voyelle française [e] (intervalle F2-F1 variable, $\sigma = 16,19$ pour G2 alors que pour G1, $\sigma = 9,05$) et une seconde cible de timbre proche de la voyelle française [i] (105-26 1/8° de tons pour la première cible et 117-24 1/8° de tons pour la seconde cible).

[au] : dans le mot « brown », le locuteur natif produit une diphtongue dont la première cible est plus ouverte que la voyelle [æ] (48-45 1/8° de tons) et la seconde cible, de même aperture, mais plus postérieure à la voyelle [ø] (101-10 1/8° de tons). Dans le mot « cow », la trajectoire est encore plus marquée puisque la première cible est proche de la voyelle [ʌ] (61-28 1/8° de tons) et la seconde cible est plus postérieure que la voyelle [ɔ:] (44-83 1/8° de tons).

Il n'y a que deux mots comportant la diphtongue [au] dans notre corpus, nous les traiterons donc directement sans passer par une moyenne générale, comme c'était le cas pour les précédentes diphtongues. Dans le mot « cow », la diphtongue n'est pas aussi marquée qu'elle l'était dans la production du locuteur natif, bien au contraire. La première cible équivaut à la voyelle [ʌ] pour G1 (60-35 1/8° de tons) et à la voyelle [ɔ] pour G2 (62-41 1/8° de tons). La seconde cible est plus postérieure pour G1, équivalente à la voyelle anglaise [ɒ] (52-53 1/8° de tons) ; elle est très peu différenciée de la première cible pour G2 (61-45 1/8° de tons) ce qui signifie qu'elle n'est pratiquement pas réalisée. Un des enfants produit la suite [ka] en RI, puis finit par se corriger après avoir entendu à nouveau le locuteur natif (enfant n°16 de G2 ; **extrait sonore n°81**). Dans cet extrait, on remarque à quel point la trajectoire du locuteur natif est marquée par rapport à celles des deux enfants. L'enfant n°23 (G2) réussit à produire l'une des plus belles diphtongues relevée pour ce mot (70-37 1/8° de tons pour la première cible et 63-55 1/8° de tons pour la seconde cible ; **dans l'extrait sonore n°82**, sa production est précédée par celle de l'enfant n°3, de G1).

Dans le mot « brown », la trajectoire de la diphtongue est marquée par une augmentation d'aperture (c'est-à-dire une montée de F2 et du rapport F2-F1), alors que le lieu d'articulation ne change guère.

Ainsi, pour les enfants de G1, le timbre de la première cible est proche de la voyelle anglaise [ʌ] et le timbre de la seconde cible, aussi fermé que le timbre de la voyelle française [œ] (65-32 1/8^e de tons), mais aussi postérieure que la voyelle anglaise [ɔ:] (80-42 1/8^e de tons). La trajectoire est moins marquée pour les enfants de G2, puisque le timbre de la première cible est proche de la voyelle française [ɔ] (62-46 1/8^e de tons) et le timbre de la seconde cible est proche de la voyelle française [œ] (74-38 1/8^e de tons). On relève également une très forte variabilité dans l'intervalle F2-F1 pour G2 alors que l'écart type est déjà élevé pour G1 ($\sigma = 30,83$ pour G2 et $\sigma = 16,19$ pour G1). Certains enfants ne produisent aucune diphtongue, mais une voyelle plus postérieure que [ɔ] pour G1 (55-46 1/8^e de tons) et une voyelle équivalente à [ɔ] pour G2 (44-39 1/8^e de tons). L'enfant n°4 (G1) produit la suite [brɔŋ] suivie par l'enfant n°6 (G2) qui produit la suite [braŋ] (73-54 1/8^e de tons pour la première cible et 58-55 1/8^e de tons pour la deuxième cible). Cette dernière production est l'une des meilleures réalisations de cette diphtongue (à la fin de l'**extrait sonore n°83**) avec celle de l'enfant n°23 du groupe G2 (71-66 1/8^e de tons pour la première cible et 69-48 1/8^e de tons pour la seconde cible) (**extrait sonore n°84**).

[ɔɪ] : dans le mot « boy », le locuteur natif produit une diphtongue dont le timbre de la première cible est proche de la voyelle [ɔ:] (40-78 1/8^e de tons) et la deuxième cible, proche de la voyelle anglaise [e], mais plus antérieure, comme la voyelle [i:] (110-14 1/8^e de tons).

En répétition immédiate, les enfants produisent également une diphtongue, mais comme pour la voyelle [ɔ:], ils produisent une voyelle trop antérieure sur la première cible, mais quand même plus postérieure que la voyelle française [ɔ] (42-64 1/8^e de tons pour G1 et 51-56 1/8^e de tons pour G2), surtout pour les enfants de G1. Pour la deuxième cible, la voyelle produite par les deux groupes est aussi fermée que celle du locuteur natif, mais plus postérieure (108-23 1/8^e de tons pour G1 et 106-26^e de tons pour G2), tendant vers la voyelle française [i]. On relève une grande variabilité dans l'intervalle F2-F1 pour la seconde cible par rapport à la première cible, surtout dans les productions des enfants de G2 ($\sigma = 21,15$ pour G1 et 31,67 pour G2).

On peut dire qu'il y a un léger effet de filtre sur les deux cibles, mais les enfants essaient d'atteindre une première cible aussi postérieure que celle du locuteur natif, même s'ils n'y parviennent pas (il y aurait un effet de filtre marqué s'ils avaient produit la voyelle française [ɔ]). La seconde cible tend vers la voyelle française [i], elle est plus postérieure que la voyelle produite par le locuteur natif. Cependant, leur timbre n'est pas très éloigné. Les deux meilleures productions ont été relevées dans chacun des deux groupes (38-99 1/8^e de tons et 144-13 1/8^e de tons pour l'enfant n°3 du groupe G1, **extrait sonore n°85** ; 44-73 1/8^e de tons et 153-22 1/8^e de tons pour l'enfant n°23 du groupe G2, **extrait sonore n°86**).

[əʊ] : dans le mot « yellow », le locuteur natif ne produit pas de diphtongue, mais la voyelle [ɔ:] (39-87 1/8^e de tons). En répétition différée, les enfants ne produisent pas non plus de diphtongue, mais une voyelle un peu plus postérieure que [ɔ] (66-49 1/8^e de tons pour G1 et 58-53 1/8^e de tons pour G2). Sur les chartes de rapports de formants (Volume II, annexe n°10.3.3.), on peut voir que pour G1, toutes les productions correspondent à [ɔ], tandis qu'un des enfants produit une voyelle proche de celle du locuteur natif (44-82 1/8^e de tons). Pour G2, les productions se répartissent plus progressivement entre les deux sons, un des enfants atteignant également la voyelle [ɔ:] (45-75 1/8^e de tons). La production de ces deux enfants intervient au cours du même entretien, l'enfant du groupe G2 (enfant n°15) étant influencé par la production de l'enfant du groupe G1 (enfant n°7) (production des deux enfants dans l'**extrait sonore n°87**, l'enfant n°7 figurant en premier).

3.2.2. Durée du noyau vocalique

3.2.2.1. Voyelles brèves

Dans notre corpus, les six voyelles brèves de l'anglais [ɒ], [ɪ], [æ], [e], [ʌ] et [ʊ] apparaissent au sein de 13 mots monosyllabiques (« *dog, doll, pink, six, fish, black, cat, red, ten, twelve, book, one, duck* »). La structure syllabique de ces mots est de type CVC. Il n'y a pas de consonne finale allongeante, sauf dans le mot « *twelve* ». La durée de la voyelle brève [ə] n'a pas été relevée car cette voyelle n'apparaît que dans les mots multisyllabiques en anglais (dans les syllabes atones).

Nous avons vu dans le chapitre V (3.1.2.1.1.1.) que les voyelles brèves de l'anglais se caractérisent par une durée absolue brève, généralement inférieure à 210 ms (comprise entre 180 et 200 ms pour les voyelles brèves de l'A.N.A. d'après les données de Lehiste et Peterson, 1960) et une durée relative faible par rapport aux voyelles longues et aux diphtongues (de 24,47 % à 29,52 % dans les syllabes CVC). Or, d'après nos résultats, les enfants produisent ces voyelles avec une durée absolue encore plus réduite et ce, pour les deux groupes et dans les deux situations de discours (tableau n°62, moyennes allant de 133,25 ms à 162,25 ms). Nous avons relevé des durées similaires dans le chapitre V pour les voyelles françaises [i], [e], [ɛ], [a], [u], [ɔ] et [œ] (inférieures à 139 ms) et nous avons posé l'hypothèse que les francophones, influencés par leurs voyelles maternelles, raccourcissaient les voyelles brèves de l'anglais (chapitre V, 3.2.1.1.4.), notamment les voyelles de timbre correspondant (à cause du phénomène de crible phonologique, assimilation aux voyelles natives et réalisation de la durée des voyelles natives ; par exemple [ɪ] à [i] ; [ʌ] à [œ] et [ʊ] à [u]). Cette hypothèse semble se vérifier d'une manière générale pour les enfants de notre corpus. A l'échelle de la syllabe, la durée relative n'est en revanche pas modifiée.

Si nous regardons les résultats de manière plus détaillée (moyennes par voyelle disponibles dans le Volume II, annexe n°10.3.), les voyelles les plus courtes de l'anglais, c'est-à-dire [ɪ], [ʌ] et [ʊ] sont produites de manière encore plus brève par les enfants francophones, notamment la voyelle [ʌ] (respectivement 98,7 ms et 93,33 ms pour G1 et G2 en RD). Pour les deux groupes, la durée absolue est cependant plus élevée en situation de RI, notamment pour G2 (162,25 ms). Il paraît logique qu'en RI, la durée des voyelles de l'anglais soit plus facilement restituable, car l'enfant répète tout de suite après le locuteur natif.

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	134,05	148,72	133,26	162,25
DR (en %)	27,51	28,30	28,84	31,48

Tableau n°62 : résultats pour les voyelles brèves dans les mots monosyllabiques (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Nous avons posé comme hypothèse initiale que les enfants du groupe « musiciens » (G1) produiraient des voyelles brève avec une durée absolue et relative proche des voyelles prononcées par un locuteur natif. Cette hypothèse ne se vérifie pas et l'ensemble des enfants produit des voyelles encore plus courtes que ne le sont les voyelles anglaises.

3.2.2.2. Voyelles longues

Les 5 voyelles longues de l'anglais [i:], [ɑ:], [ɔ:], [u:] et [ɜ:] apparaissent au sein de 9 mots monosyllabiques dans notre corpus (« *car, horse, four, blue, two, bird, green, sheep, three* »). La structure syllabique de ces mots est soit de type CVC, soit de type CV. Les voyelles longues de l'anglais se caractérisent par une grande durée absolue (chapitre V, 3.2.1.1.2.) pouvant aller de 233 à 325 ms dans les syllabes de type CVC (entre 240 et 260 ms pour les voyelles longues de l'A.N.A. d'après les données de Lehiste et Peterson, 1960). Dans le même type de syllabe, la durée absolue des voyelles longues produites par les enfants de notre corpus correspond à ces données. Ils respectent la différence de durée entre ces voyelles et les voyelles brèves (les voyelles longues étant en moyenne deux fois plus longues que les voyelles brèves). Leur durée relative correspond à la durée la plus longue relevée pour la locutrice native dans le chapitre V. En effet, la proportion moyenne de la voyelle longue au sein de la syllabe varie pour les enfants de 42 à 46 % (tableau n°63), alors que la durée relative la plus longue était de 42,09 % (voyelle [ɑ:]) pour la locutrice dans les syllabes CVC et pouvait descendre jusqu'à 27,9 % (voyelle [i:]).

En ce qui concerne la différence entre les deux groupes, on relève un comportement inverse de G1 par rapport à G2. Nous avons déjà noté que les voyelles brèves étaient plus longues en situation de RI pour G2. Ce phénomène se confirme pour les voyelles longues (durée moyenne de 261,55 ms en RI contre 218,91 ms en RD). En revanche, pour G1, c'est en situation de RD que les voyelles longues sont plus longues (262 ms) par rapport à la situation de RI (229,05 ms). Cependant, pour les deux groupes, la durée relative des voyelles longues est plus élevée en RD qu'en RI (46,4 % pour G1 et 46,94 % pour G2 en RD contre 42,08 % pour G1 et 42,48 % pour G2 en RI).

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	262	229,05	218,91	261,55
DR (en %)	46,4	42,08	46,94	42,48

Tableau n°63 : résultats pour les voyelles longues dans les mots monosyllabiques de type CVC (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée,

RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Les voyelles longues qui apparaissent dans des syllabes de type CV sont plus longues chez les enfants du groupe G1 que chez les enfants du groupe G2 (tableau n°64). C'est en situation de répétition immédiate que la différence est la plus marquée (332 ms pour G1 contre 105,33 ms pour G2), mais il y a trop peu d'occurrences en RI pour déterminer une tendance générale, les mots « *three* », « *blue* » et « *two* » étant connus des enfants et produits majoritairement en situation de répétition différée (voir tableaux figurant dans le Volume II, annexe 10.3.).

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	192,62	332	175,62	105,33
DR (en %)	49,54	47,29	48,06	30,89

Tableau n°64 : résultats pour les voyelles longues dans les mots monosyllabiques de type CV (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Nous ne commenterons pas les résultats par mots (disponibles dans le Volume II, annexe 10.3.), mais nous prendrons deux exemples pour illustrer deux comportements différents. La durée absolue de la voyelle longue [i:] ainsi que sa durée relative semblent plus courtes dans les mots « *sheep* » et « *green* » produits par les enfants (Volume II, annexe n°10.3.) que ce qui avait été relevé pour la locutrice francophone dans le chapitre V. Ce « raccourcissement » pourrait indiquer un phénomène d'assimilation avec la voyelle française de timbre correspondant [i] et c'est effectivement ce qu'indiquent les mesures de timbre réalisées dans la partie précédente (voir 3.2.1.1.2.). A l'inverse, la voyelle longue [ɑ:] est très allongée dans le mot « *car* » pour les deux groupes (309 ms pour G1 et 380 ms pour G2 en RI). Or, cette voyelle sonne « anglais », d'autant plus qu'elle est suivie de la consonne [r] qui n'existe pas en français. En répétition immédiate les enfants perçoivent bien ce son et essaient de le restituer, notamment en amplifiant sa durée et en diphtonguant (**extrait sonore n°88**). Le mot « *car* » est connu en France, même par les non anglicistes et il est habituellement prononcé à la française, [kaɛr]. Sa « version » anglaise surprend l'un des deux enfants : « *on aurait dû penser à [kar], mais [kaewə]* » (**extrait sonore n°89**). En revanche, les caractéristiques spécifiquement « anglaises » de la voyelle [i:] sont beaucoup moins saillantes que celles de la voyelle [ɑ:], donc moins repérables, ce qui la rend plus vulnérable au phénomène de crible phonologique.

3.2.2.3. Diphtongues

Parmi les 8 diphtongues que compte le système vocalique anglais, les 4 diphtongues [aɪ], [eɪ], [ɔɪ] et [aʊ] apparaissent dans notre corpus au sein de 6 mots monosyllabiques (« *white, five, nine, eight, cake, brown* »). La structure syllabique de ces mots est de type CVC. Dans le chapitre V, nous avons vu que les diphtongues figurent parmi les noyaux vocaliques les plus longs de l'anglais avec une grande durée absolue (jusqu'à 510 ms). Dans les productions des enfants, les diphtongues sont également les noyaux vocaliques possédant les durées absolues les plus élevées par rapport aux voyelles brèves et aux voyelles longues. On ne relève aucune différence entre G1 et G2 (voir tableau n°65). En revanche, le discours semble affecter la durée absolue et relative des diphtongues de la même manière dans les deux groupes. Ainsi, on relève une durée absolue en RD pour G1 (293,15 ms) et pour G2 (295,86) plus élevée qu'en RI (248,26 ms pour G1 et 243,40 ms pour G2). De même, la durée relative est plus élevée en RD pour G1 (53,66 %) et pour G2 (52,59 %) qu'en RI (41,06 % pour G1 et 40,47 % pour G2).

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	293,15	248,26	295,86	243,40
DR (en %)	53,66	41,06	52,59	40,47

Tableau n°65 : résultats pour les diphtongues dans les mots monosyllabiques de type CVC (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Pour les diphtongues présentes dans des syllabes de type CV ([ɔɪ] dans « *boy* » et [aʊ] dans « *cow* »), on ne relève pas de différence entre les groupes (voir tableaux figurant dans le Volume II, annexe n°10.3.). Comme ces deux mots ne sont pas connus des enfants et ont été en majorité produits en situation de répétition immédiate, nous ne pouvons en tirer une tendance générale sur la différence entre les deux types de discours. Cependant, nous pouvons relever la durée absolue la plus élevée de tous les noyaux vocaliques, avec une durée relative proche des 80 %.

En ce qui concerne la différence entre les groupes, nous avons posé comme hypothèse que les enfants du « groupe musiciens » (G1) produiraient des voyelles longues et des diphtongues d'une durée absolue et relative plus grande que les enfants du « groupe non musiciens » (G2), spécialement en situation de répétition différée (RD). Cette hypothèse est en partie confirmée puisque les voyelles produites par G1 sont plus longues en RD dans les syllabes de type CVC.

Quant aux voyelles produites dans les syllabes de type CV, elles sont plus longues pour G1 en RD, mais aussi en RI. La tendance du groupe G1 à allonger les voyelles en situation de RI est confirmée pour les voyelles longues dans les syllabes de type CVC. Les diphtongues constituent les noyaux vocaliques les plus spécifiques de l'anglais par rapport au français ; ils sont également les plus allongés et cet allongement est correctement reproduit par l'ensemble des enfants de notre corpus. Aucune différence pertinente n'a été relevée entre les groupes. En revanche, la variable indépendante « discours » influence la durée des diphtongues, puisque celle-ci est plus allongée en situation de répétition différée.

4. Etude n°5 « Stabilité » de timbre

4.1. Plan expérimental

4.1.1. Hypothèse

Les locuteurs dont l'oreille a été entraînée par la pratique musicale rencontreront plus de facilités à restituer l'instabilité de timbre des voyelles anglaises.

4.1.2. Variables indépendantes

4.1.2.1. La Musique : deux modalités

La variable indépendante « musique » est composée de deux groupes, le groupe d'enfants musiciens G1 et le groupe d'enfants non musiciens G2 (voir 2.1.2.1.).

4.1.2.2. Le discours : deux modalités

La variable indépendante « discours » est composée de deux modalités : la modalité de discours immédiat RI et la modalité de discours différé RD (voir 2.1.2.2.).

4.1.3. Variable dépendante : durée de la zone de stabilité de timbre de la voyelle

Dans le chapitre V (3.1.3.2.2.), nous avons vu que les voyelles brèves de l'anglais se caractérisent par une zone de stabilité de timbre étendue et des zones de transition réduites, alors que les voyelles longues se caractérisent par une zone de stabilité de timbre réduite par rapport à des zones de transition étendues. A cause du phénomène de crible phonologique, certaines voyelles longues peuvent être assimilées à des voyelles brèves et en conséquence, être produites de la même façon, c'est-à-dire avec une trop grande stabilité de timbre.

Par ailleurs, à cause de l'excès de tension articulatoire qui caractérise la production de sa LM et qu'il va transposer à l'anglais (rythmicité syllabique), un apprenant francophone éprouvera des difficultés à restituer l'instabilité de timbre des voyelles longues de l'anglais. En revanche, il ne devrait pas rencontrer de difficultés avec les voyelles brèves car celles-ci se caractérisent par une grande phase de stabilité.

Afin d'étudier le paramètre de stabilité, il faut déterminer pour chaque type de son, c'est-à-dire pour les voyelles longues, mais également pour les voyelles brèves, la proportion de stabilité et la proportion d'instabilité. Par exemple, dans le mot « *horse* » prononcé par le locuteur natif (figure n°102 et **extrait sonore n°90**), la proportion de stabilité de la voyelle [ɔ:] est de 20,45 %, alors que la proportion d'instabilité est de 79,54 % (avec une transition initiale de 10,25 % et une transition finale de 66,23 %).

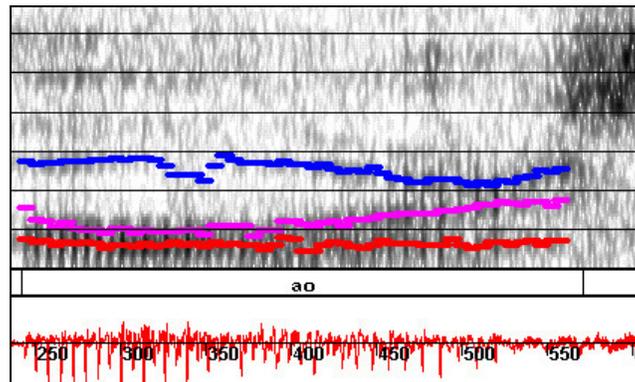


Figure n°102 : représentation spectrographique du mot « horse » avec tracé des trois premiers formants de la voyelle [ɔ:].

A l'inverse, dans le mot « *duck* » (figure n°103), la proportion de stabilité est de 88,17 %, alors que celle d'instabilité est de 11,82 % (absence de phase de transition initiale et présence d'une phase de transition finale réduite, de 11,82 %).

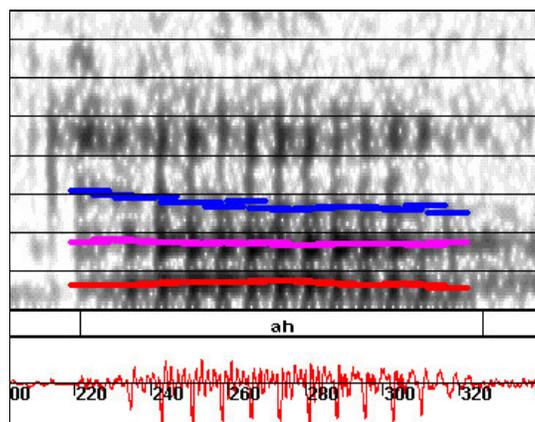


Figure n°103 : représentation spectrographique du mot « duck » avec tracé des trois premiers formants de la voyelle [ʌ].

La variable « stabilité de timbre » comprend donc deux modalités : la modalité « stabilité » et la modalité « instabilité ». La proportion de stabilité correspond à la durée relative de la zone de stabilité de timbre et la proportion d'instabilité, à la somme de la durée relative de la zone de transition initiale et de la durée de la zone de transition finale.

Nous posons l'hypothèse opérationnelle que les locuteurs dont l'oreille a été entraînée par la pratique musicale restitueront mieux l'instabilité de timbre des noyaux vocaliques longs de l'anglais, car ils pourraient reproduire plus facilement la rythmique spécifique de l'anglais et ils sont moins sensibles au phénomène de crible phonologique (cf. résultats de l'étude n°2 en perception). Quant aux voyelles brèves, elles seront produites avec une grande stabilité de timbre par les enfants des deux groupes. Le plan expérimental de l'étude est présenté dans le tableau n°66.

	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
Répétition différée (RD)				
Répétition immédiate (RI)				

Tableau n°66 : tableau expérimental de l'étude n°5 (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens »).

Matériel analysé

Notre corpus est constitué de 43 mots prononcés par 25 enfants (l'un des entretiens n'ayant pu être analysé en raison d'un fort bruit de fond), soit un total de 1075 mots. Les noyaux vocaliques étudiés sont les 5 voyelles longues de l'anglais, c'est-à-dire [i:], [ɜ:], [u:], [ɔ:] et [ɑ:] et les 7 voyelles brèves de l'anglais, c'est-à-dire [ɪ], [e], [æ], [ɒ], [ʌ], [ʊ] et [ə]. Cette dernière voyelle sera étudiée à part car elle dépend du phénomène d'accent. Sur la totalité des noyaux vocaliques appartenant à ces mots, nous avons lancé une détection automatique des trois premiers formants avec le logiciel Winsnoori (voir note n°X et annexe n°X pour un exemple de fichier de sortie édité avec ce logiciel). Cependant, cette détection a échoué pour un grand nombre de fichiers et seulement 260 fichiers ont pu être édités, soit 24,18 % de la totalité des mots composant le corpus. Outre des problèmes de qualité du signal pour certains enregistrements (rapport signal/bruit élevé), le problème principal a été le traitement des voix d'enfants. La détection automatique est en effet beaucoup plus facile pour des voix graves, en particulier les voix d'hommes.

Ainsi, toutes les productions du locuteur natif ont pu être analysées, même lorsque le signal était bruité. En revanche, la détection était beaucoup plus difficile pour les productions des enfants : leurs voix sont bien sûr beaucoup plus aiguës, mais de plus, certains enfants s'expriment avec une voix faible car ils sont intimidés par la situation d'entretien. Par ailleurs, nous avons relevé des problèmes spécifiques de détection concernant les voyelles dont les second et troisième formants sont élevés, comme [i:] et [I] ou concernant les voyelles compactes dont les premier et second formants sont très rapprochés, comme [u] et [u:].

A cause de ces problèmes de détection, la fiabilité des 260 fichiers retenus a été vérifiée scrupuleusement grâce aux mesures manuelles réalisées sur la totalité du corpus dans le cadre de l'étude n°4 (détail des fichiers, tableau n°67). Ils ont ensuite été traités par le programme informatique créé spécialement à cet effet (pour une description de la méthode et une description détaillée de la conception du programme, se reporter au Volume II, annexe n°11.1.). Celui-ci fournit la durée relative des différents événements composant le noyau vocalique, c'est-à-dire la transition initiale, la zone de stabilité de timbre et la transition finale (exprimée en pourcentage par rapport à la durée totale du noyau vocalique).

	G1		G2	
	RD	RI	RD	RI
Nombre de fichiers	46	61	89	85
Total	107		174	

Tableau n°67 : nombre de fichiers par mots et par discours (avec RD pour répétition différée et RI pour répétition immédiate).

3.2. Résultats

3.2.1. Résultats généraux

Les voyelles longues de l'anglais se caractérisent par une plus grande zone de variabilité de timbre que les voyelles brèves et cette caractéristique est respectée par les enfants des deux groupes en situation de RD. En effet, la proportion d'instabilité est de 54,02 % pour G1 et de 51,09 % pour G2, alors que pour les voyelles brèves, elle est de 43,31 % pour G1 et de 35,69 % pour G2 (voir tableau n°67). En situation de RI, on observe la même tendance pour G1 avec une proportion d'instabilité de 50,17 % pour les voyelles longues et de 46,55 % pour les voyelles brèves. En revanche pour G2, dans la même situation de discours, les voyelles longues sont produites avec trop de stabilité (57,27 % de stabilité) et se rapprochent de la structure des voyelles brèves (42,72 % d'instabilité pour les voyelles longues contre 39,49 % d'instabilité pour les voyelles brèves).

On remarque donc que dans toutes les situations de discours, les voyelles longues sont produites avec une proportion d'instabilité plus grande pour G1 sauf pour G2 en RI, où elles sont produites avec trop de stabilité. L'hypothèse initiale est donc confirmée, les enfants musiciens restituent mieux l'instabilité de timbre qui caractérise les voyelles longues anglaises.

Les voyelles brèves sont produites avec une proportion de stabilité plus grande que les voyelles longues et cette structure est respectée par les enfants du groupe G2 (64,30 % en RD et de 60,50 % en RI). En revanche, nous relevons un **résultat inattendu** pour les enfants du groupe G1. Ils produisent les voyelles brèves avec une proportion d'instabilité plus élevée par rapport aux enfants du groupe G2 et ce, dans les deux situations de discours (46,55 % d'instabilité pour G1 contre 39,49 % pour G2 en RI et 43,31 % d'instabilité pour G1 contre 39,49 % pour G2 en RD). Nous pensions relever des résultats identiques pour les deux groupes car les voyelles brèves devraient être plus facilement reproduites que les voyelles longues. La stabilité qui les caractérise correspond en effet à la stabilité de timbre des voyelles françaises. Une interprétation possible de ces résultats est que les enfants musiciens ont repéré que l'instabilité de timbre caractérisait certaines voyelles anglaises. En effet, ils la reproduisent correctement au sein des voyelles longues. Cette caractéristique n'existe pas en français et elle contribue à donner à l'anglais sa « couleur » sonore spécifique. Il est possible qu'ils généralisent cette instabilité à des sons qui ne devraient pas l'être pour que leurs énoncés « sonnent » anglais. Nous avons déjà évoqué cette hypothèse à deux reprises dans l'étude n°3, pour les résultats concernant la variable « forme de la courbe » et la variable « localisation de l'accent ». En effet, les enfants du groupe G1 ont tendance à produire plus de contours spécifiquement anglais (les contours en cloche) sur des mots qui ne devraient pas être affectés par de tels contours (les mots commençant par une initiale non voisée). De la même manière, dans les mots trisyllabiques qui portent l'accent sur la seconde syllabe, les erreurs de localisation de l'accent produites par les enfants du groupe G1 affectent toujours la première syllabe et jamais la dernière. Or l'anglais se caractérise justement par une tendance à l'accentuation de la première syllabe. Le comportement des musiciens consisterait donc plutôt à « généraliser » des caractéristiques spécifiques à l'anglais plutôt qu'à les « exagérer ».

		G1		G2	
		Stabilité	Instabilité	Stabilité	Instabilité
RD	Voyelles brèves	56,68	43,31	64,30	35,69
	Voyelles longues	45,97	54,02	48,90	51,09
RI	Voyelles brèves	53,44	46,55	61,45	38,54
	Voyelles longues	49,82	50,17	60,50	39,49

Tableau n°68 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles brèves et des voyelles longues de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition immédiate (RI) et de situation différée (RD).

4.2.2. Résultats détaillés

4.2.2.1. Voyelles longues

En RD, la voyelle [i:] est marquée par une grande instabilité pour les deux groupes, surtout pour G2 (81,23 % pour G2 et 66,66 % pour G1 ; voir tableau n°69). En RI, cette instabilité est conservée pour G2 (66,72 %) et plus modérée pour G1 (47,53). Or cette voyelle figure parmi les voyelles les plus stables des voyelles longues de l'anglais (avec 63,79 % de stabilité chez la locutrice francophone étudiée dans le chapitre V). De même, en RD, la voyelle [ɜ:] est produite avec beaucoup d'instabilité par G2 (65,13 % d'instabilité), alors qu'elle est beaucoup plus stable pour G1 (38,55 % d'instabilité). Or, comme la voyelle [i:], cette voyelle est produite de manière stable par la locutrice anglophone (voir chapitre V). En revanche, la stabilité de cette voyelle est plus grande en RI (47,53 % d'instabilité pour G1 et 33,26 % d'instabilité pour G2). Pour ces deux voyelles, c'est G2 qui les produit avec le plus d'instabilité, surtout en RD.

Les voyelles [u:] ; [ɔ:] sont très instables pour G1 en RD et en RI. La même instabilité affecte la voyelle [u:] en RD pour G2 ; en revanche, on remarque une très grande stabilité en RI (93,82 % de stabilité, voir tableau n°70). A l'inverse, la voyelle [ɔ:] est marquée par une instabilité plus grande en RI (52,73% d'instabilité) qu'en RD (42,35 % d'instabilité). En RI, la voyelle [ɑ:] est instable pour G1 (57,57 % d'instabilité). On obtient les mêmes résultats pour G2 en RI (55,18 % d'instabilité). En revanche, on observe une grande stabilité en RD pour G2 (88,74 % de stabilité). Pour G1, la comparaison n'est pas possible en raison de l'absence d'occurrences pour cette voyelle en RD.

Répétition différée (RD)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[i:]	33,33	66,66	18,76	81,23
[ɜ:]	61,44	38,55	34,87	65,13
[u:]	44,22	55,77	44,49	55,50
[ɔ:]	44,90	55,08	57,64	42,35
[ɑ:]			88,74	11,25
Moyenne totale	45,97	54,02	48,90	51,09

Tableau n°69 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles longues de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition différée.

Répétition immédiate (RI)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[i:]	52,46	47,53	33,26	66,72
[ɜ:]	58,41	41,58	67,19	32,80
[u:]			93,82	6,18
[ɔ:]	46	53,99	47,25	52,73
[ɑ:]	42,41	57,57	44,81	55,18
Moyenne totale	49,82	50,17	57,27	42,72

Tableau n°70 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles longues de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition immédiate.

3.2.2.2. Voyelles brèves

Les voyelles brèves de l'anglais se caractérisent par une durée absolue réduite et surtout une grande stabilité de timbre (plus de 50 % de la durée de la voyelle pour la locutrice anglophone dans le chapitre V). Les voyelles [ɪ], [e], [æ] et [ɒ] sont très stables en RD pour G2. Pour G1, il n'y a que la voyelle [ɒ] qui soit produite avec beaucoup de stabilité (76 %, voir tableau n°71). Les autres voyelles sont plus instables et en particulier, la voyelle [æ] (38,04 % de stabilité contre 64,44 % pour G2). Quant à la voyelle [ʌ], elle est instable pour les deux groupes (57,82% d'instabilité pour G1 et 52,53 % d'instabilité pour G2).

En situation de RI, les voyelles brèves sont très stables pour G2 sauf la voyelle [ɒ] (57,19 % d'instabilité, voir tableau n°72). On peut remarquer que la voyelle [ʌ] est plus stable, mais pas aussi stable que les autres voyelles (51,26 % de stabilité, alors que pour [æ], 58,02 % de stabilité, mais surtout pour [ɪ], 75,13 % de stabilité et pour [e], 80,51 % de stabilité). Pour G1, on observe les mêmes résultats que G2 pour la voyelle [æ] (59,02 % de stabilité). On relève également plus de stabilité pour G1 en RI qu'en RD.

La voyelle [ʌ] est plus instable pour G1 que G2 (55,14 % d'instabilité contre 48,73 % pour G2). Pour la voyelle [ɒ], elle est plus stable pour G1 (50,23 % de stabilité) que pour G2 (42,31 %). On remarque plus d'instabilité pour G2 en RI par rapport à RD.

La voyelle [ə] est spécifique à étudier parce que sa stabilité dépend beaucoup de l'accentuation. En effet, dans les syllabes non accentuées, on observe une réduction du timbre et une grande instabilité. Mais, les enfants produisent ce son avec trop de stabilité. En effet, on observe dans la réalisation de ce son une proportion plus forte de stabilité que d'instabilité (la proportion de stabilité allant de 52,56 à 55,76 %) et ce, pour les deux groupes et dans les deux situations de discours.

Répétition différée (RD)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[ɪ]	56,92	43,0644,24	79,24	20,75
[e]	55,74	44,24	68,02	31,98
[æ]	38,04	61,95	64,44	35,55
[ɒ]	76,00	23,98	67,59	32,40
[ʌ]	42,16	57,82	47,46	52,53
[ʊ]			58,58	41,41
[ə]	52,56	47,43	64,80	35,19
Moyenne totale	56,68	43,31	64,30	35,69

Tableau n°71 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles brèves de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition différée.

Répétition immédiate (RI)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[ɪ]	49,13	50,86	75,13	24,86
[e]	61,64	38,35	80,51	19,49
[æ]	59,02	40,97	58,02	41,97
[ɒ]	50,23	49,75	42,31	57,68
[ʌ]	44,84	55,14	51,26	48,73
[ʊ]	55,70	44,28		
[ə]	53,48	46,51	55,79	44,20
Moyenne totale	53,44	46,55	60,50	39,49

Tableau n°72 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles brèves de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition immédiate.

5. Résumé des résultats et discussion générale

5.1. Production de la prosodie

Dans l'étude n°3, nous avons posé comme hypothèse que les locuteurs dont l'oreille a été entraînée par la pratique musicale rencontreraient plus de facilités à reproduire les éléments musicaux de l'anglais, c'est-à-dire les éléments prosodiques. Cette hypothèse se vérifie partiellement pour deux variables sur les trois variables étudiées.

5.1.1. Forme de la courbe

En ce qui concerne la forme de la courbe, il apparaît en effet qu'en situation de répétition immédiate, la proportion de contours spécifiquement anglais est plus grande chez les enfants musiciens. Inversement, la proportion de ces contours dans les productions des non musiciens est plus faible, révélant peut-être un phénomène de crible prosodique avec le français en situation de répétition immédiate (réalisation de contours descendants à la place des contours en cloche). Pour des enfants de 7 ans, il ne suffit donc pas de répéter des mots après un locuteur natif pour pouvoir les reproduire correctement. En revanche, le fait d'avoir un oreille entraînée par la pratique musicale leur permet de mieux repérer les caractéristiques prosodiques non natives et donc de mieux les reproduire. Par ailleurs, les enfants musiciens généralisent la production des contours en cloche aux énoncés où ces contours ne devraient pas apparaître, c'est-à-dire dans les mots monosyllabiques à l'initiale non voisée (normalement affectés d'un contour descendant). Les enfants se montrent donc « bons imitateurs », non pas dans le sens où ils imitent fidèlement la réalité, mais dans le sens où ils sont capable d'isoler les principales caractéristiques d'une langue et de les incorporer à leurs productions pour avoir au final des énoncés qui « sonnent » anglais, en tout cas aux oreilles des français. Ce faisant, ils se comportent exactement comme le ferait un caricaturiste lorsqu'il restitue les traits caractéristiques du visage d'une personne sur un dessin ou sa façon de parler dans une imitation vocale ou sa façon de bouger dans une imitation corporelle et gestuelle. Si les contours en cloche contribuent à la couleur « anglaise » de l'énoncé, alors pourquoi ne pas les généraliser ?

5.1.2. Localisation de l'accent

On retrouve un phénomène similaire dans la restitution de l'accent dans les mots trisyllabiques. Nous avons vu qu'il était difficile de tirer une conclusion quant à la différence entre G1 et G2 et ce, à cause du petit nombre d'occurrences analysées. Cependant, lorsque les enfants des deux groupes se trompent et localisent l'accent sur une mauvaise syllabe, c'est le plus souvent la première syllabe qui est concernée et non la syllabe finale, ce qui aurait traduit une interférence avec la rythmique du français. Or, l'accentuation initiale constitue l'une des caractéristiques distinctives de l'anglais par rapport au français (cf. Chapitre V, 2.1.1.2.). Comme pour les contours en cloche, les enfants pourraient avoir identifié cette caractéristique distinctive et la généraliser. Ainsi, même s'ils font des erreurs, leurs productions ressemblent à des mots anglais. Cette stratégie de généralisation est utilisée par les enfants des deux groupes, sûrement à cause du fait que l'accent en anglais est un phénomène sonore saillant. Il n'y a donc pas besoin d'avoir une oreille particulièrement entraînée pour le repérer. Nous avons vu dans le chapitre VII que l'accent était perçu facilement dans le test de perception T5 et ce, par l'ensemble des enfants de notre échantillon (étude n°1). Les résultats concernant leur restitution de l'accent dans les mots bisyllabiques confirment cette interprétation. En effet, les enfants des deux groupes le localisent quasiment parfaitement (plus de 90 % de productions correctes dans les deux groupes) en situation de répétition immédiate. En revanche, la tâche semble beaucoup plus difficile en répétition différée et ce sont les musiciens qui s'en acquittent le mieux. Dans ce type de répétition, c'est la mémoire à long terme qui est mise en jeu puisque ces enfants restituent des énoncés qu'ils ont appris en classe. La capacité des musiciens à reproduire correctement des énoncés stockés à long terme pourrait provenir d'une mémoire musicale plus optimale. La mémoire musicale de la parole est définie par Llorca (1992 : 47-48) comme :

« celle qui repose seulement sur les données concrètes, données phonétiques et prosodiques d'une séquence entendue, indépendamment de leur valeur linguistique : il s'agit purement de la mémoire des sons perçus et de leur agencement ».

La mémoire musicale est très importante dans une LE, puisqu'elle permet la mémorisation d'une forme sonore avant l'accès à son interprétation linguistique et à son sens. La prosodie joue un rôle essentiel dans cette mémorisation :

« On découvre que l'intégration des intonations et des schémas rythmiques se fait davantage par des souvenirs sonores que par des connaissances théoriques. Par exemple, on ne retient vraiment la place de l'accent dans un mot anglais que lorsqu'on a le souvenir de sa prononciation par une personne particulière, avec la voix de cette personne et l'intervalle mélodique qu'elle a placé entre la syllabe accentuée et la syllabe inaccentuée » (Llorca, 1992 : 60).

Les musiciens semblent donc capables de conserver des « souvenirs sonores » précis des énoncés qu'ils ont appris, probablement parce qu'ils ont été entraînés à mémoriser des extraits sonore au cours de leur formation musicale. En revanche, les énoncés seraient moins bien stockés par les non musiciens et pourraient être plus facilement altérés par des effets de filtre avec la langue maternelle. Il est donc nécessaire d'opérer un « rafraîchissement » fréquent au contact du modèle natif pour éviter ces phénomènes d'assimilation « à long terme ».

5.1.3. Exagération de la courbe

Contrairement aux deux autres variables, notre hypothèse ne se vérifie pas pour la variable « exagération de la courbe ». Non seulement les enfants musiciens n'utilisent pas de stratégie d'exagération, mais de plus, leurs contours moyens sont étonnamment identiques à ceux des enfants non musiciens. En effet, les contours des deux groupes se superposent presque parfaitement et ce, sans distinction de situation de discours. Par ailleurs, ces contours moyens sont beaucoup moins amples que les productions du locuteur natif, même en répétition immédiate. Pourtant, nous avons trouvé, dans un corpus rassemblant les productions de 10 enfants du même âge (Dodane, 1997, 2000), un enfant musicien, Anthony, qui manifestait une stratégie d'exagération très marquée. Il est possible que cette stratégie ne se manifeste que chez des enfants ayant un très bon niveau musical, ce qui était le cas de cet enfant. Il est également possible que d'autres facteurs interviennent dans la formation de ce type de stratégie comme la personnalité de l'apprenant par exemple. En effet, est-ce que l'apprenant va oser exagérer ses productions en anglais ou au contraire, va-t-il les « tasser » par peur d'avoir l'air ridicule devant ses camarades ? L'âge pourrait également se révéler déterminant, les enfants plus jeunes n'ayant pas peur de jouer avec les sons ; le poids social est déjà identique à 7 ans pour tous les enfants mais il est possible qu'une étude avec des enfants plus jeunes fasse apparaître des différences entre les groupes. Dans la présente étude, nous avons choisi d'étudier les productions des enfants par groupe, « groupe musiciens » et « groupe non musiciens ».

Ainsi, les contours étudiés sont des contours moyens, c'est-à-dire qu'ils gommant les différences entre les individus. Avec cette méthode, si l'un des enfants de notre corpus manifeste une stratégie d'exagération, nous ne pouvons pas le remarquer. Il serait donc intéressant, dans le futur, de compléter notre analyse par une étude qualitative des productions de chaque enfant de notre corpus. Cette étude sera facilitée par le fait que nous disposons dorénavant des contours moyens produits par l'ensemble des enfants, contours que nous pourrions comparer aux productions individuelles. Si nous n'avons pas fait apparaître de stratégie d'« exagération », nous avons en revanche fait apparaître, et de manière plus marquée chez les enfants musiciens que les enfants non musiciens, une stratégie de « généralisation » des caractéristiques prosodiques spécifiquement non natives de l'anglais.

5.2. Production des contrastes vocaliques

Dans l'étude n°4, nous avons posé l'hypothèse que le fait d'avoir une oreille entraînée par la pratique musicale amènerait les apprenants à mieux percevoir et donc mieux restituer le timbre et la durée des contrastes vocaliques de l'anglais, c'est-à-dire les voyelles brèves, les voyelles longues et les diphtongues.

5.2.1. Timbre du noyau vocalique

5.2.1.1. Voyelles brèves

L'ensemble de nos résultats ne fait guère apparaître de différences entre les groupes G1 et G2 ; on relève quelques différences en faveur de l'un ou l'autre groupe mais pas de tendance générale qui pourrait les différencier. S'il y a une assimilation, elle est réalisée la plupart du temps par les deux groupes pour une situation de discours déterminée ou pour les deux situations de discours. On relève plusieurs types de fautes :

- Soit on relève une **assimilation de la voyelle anglaise à une voyelle française** comme c'est le cas pour [ɪ] et [i] (bien que le groupe G1 ait tendance à produire une voyelle plus fermée, donc plus proche de la production du locuteur natif dans les mots « *teddy, orange, baby* »).

- On relève également des **assimilations provoquées par la transposition de la rythmique française à l'énoncé anglais**, en particulier dans les mots multisyllabiques : dans « umbrella », la voyelle [ʌ] devient [ɒ] et la voyelle [ə] devient [æ] ; dans « seven », la voyelle [ə] devient [œ]. La voyelle [ə] devrait être centrale car elle apparaît au sein d'une syllabe non accentuée (réduction de durée et de qualité). Or, les petits francophones ont tendance à produire des syllabes de durée identique à cause de la régularité syllabique et une syllabe allongée en syllabe finale à cause de la rythmique du français. Le timbre de la voyelle est donc trop marqué.
- **La prononciation de certains mots anglais semble influencée par leur prononciation « à la française »** (mots courants et connus même des non anglicistes), comme dans le mot « dog » (G1 et G2) et le mot « orange » (G1 en RI) où la voyelle [ɒ] est assimilée à la voyelle française [ɔ].
- **La voyelle est correctement reproduite**, comme c'est le cas de la voyelle [ɒ] dans « orange » (G1) et « doll » (G1 et G2). De la même manière, la voyelle [æ] est correctement produite, sauf dans « apple » (G1 et G2) et dans « parrot » (G2). C'est également le cas pour la voyelle [e] sauf dans le mot « red » (G2) (« twelve, comme LN). La voyelle [ʌ] est bien reproduite dans le mot « one » en RD pour les deux groupes.
- La voyelle peut être correctement reproduite, mais ses **caractéristiques non natives sont exagérées**. C'est le cas de la voyelle [ʊ] qui est prononcée comme la voyelle française [o] (pas d'assimilation avec la voyelle française [u] comme on aurait pu le prévoir), et de la voyelle [ʌ] dans le mot « duck » en RI qui est prononcée comme la voyelle française [ɔ].

5.2.1.2. Voyelles longues

Comme pour les voyelles brèves, on ne relève pas de différences marquées entre les groupes dans la prononciation des voyelles longues :

- La **voyelle anglaise [u:] est bien réalisée** et proche de la production du locuteur natif. La voyelle [ɔ:] n'est pas aussi postérieure que celle du locuteur natif, mais

elle est plus postérieure que la voyelle française [ɔ]. On ne peut donc pas dire qu'il y a assimilation car les enfants des deux groupes tentent d'« atteindre » le timbre de la voyelle produite par le locuteur natif, surtout dans le mot « horse ».

- La voyelle [ɑ:] est produite comme la voyelle [æ] par les enfants, mais le locuteur natif les prononce exactement de la même manière. En revanche, dans les mots « car » et « banana » en situation de répétition immédiate, les enfants produisent une voyelle plus postérieure que celle du locuteur natif en situation de répétition immédiate.
- **On relève une assimilation de la voyelle anglaise à une voyelle française dans deux cas :** la voyelle anglaise [i:] est assimilée à la voyelle française [i] et la voyelle anglaise [ɜ:] est assimilée à la voyelle française [œ]. Cependant, on remarque que les enfants musiciens, même s'ils sont sensibles à cette assimilation, produisent des voyelles dont le timbre est légèrement plus proche du timbre des voyelles produites par le locuteur natif.

5.2.1.3. Diphtongues

On relève quelques différences entre les groupes dans la production des diphtongues [aɪ] et [aʊ] ; les autres diphtongues sont produites de façon relativement homogène :

- Dans la diphtongue [aɪ], bien qu'on relève un effet de filtre sur la seconde cible pour les deux groupes ([i] ou [e]), les **deux cibles sont atteintes avec plus de précision par les enfants de G1**. De même, **ils réalisent mieux la diphtongue [aʊ]** que les enfants du groupe G2. Mais, même en fonction de ces différences, la trajectoire de l'ensemble des diphtongues produites par les enfants est beaucoup moins marquée que celle du locuteur natif.
- Dans la diphtongue [eɪ], on relève un **effet de filtre sur chaque cible** : la première cible est assimilée aux voyelles françaises [e] ou [ɛ] et la seconde cible, à la voyelle française [i]. Le même phénomène se produit pour la diphtongue [ɔɪ], où la première cible est assimilée à la voyelle française [ɔ] et la deuxième cible, à la voyelle française [i].

- Dans les mots « baby », « cake » et « grappe », **les enfants ont tendance à produire une voyelle à la place de la diphtongue**. Il est possible qu'ils soient influencés par la prononciation française de ces mots, comme nous l'avons déjà relevé pour certaines voyelles brèves. Dans les mots « baby » et « cake », le timbre de la voyelle est proche de celui de la voyelle française [e] et dans le mot « grappe », il se situe entre celui des deux voyelles françaises [e] et [i]. Dans le mot « yellow », tous les enfants produisent une voyelle équivalente à la voyelle française [ɔ], mais légèrement plus postérieure. Cependant, le locuteur natif ne produit pas non plus de diphtongue, mais la voyelle [ɔ:].

L'ensemble de ces résultats ne laisse guère apparaître de différences entre les groupes de musiciens et de non musiciens, mis à part quelques diphtongues et quelques voyelles qui sont mieux reproduites par les musiciens. Cependant, comme nous l'avons déjà évoqué plus haut, l'analyse par groupe atténue les différences individuelles et il est possible qu'une analyse des productions par individu contribue à mettre à jour des différences entre musiciens et non musiciens, notamment si nous limitons cette analyse à la comparaison des productions des enfants ayant les meilleures oreilles musiciennes à celles des enfants ayant les plus mauvaises oreilles musiciennes de notre corpus.

Si nous avons échoué à faire apparaître une différence entre musiciens et non musiciens, les productions relativement homogènes des enfants révèlent des tendances générales qui pourraient caractériser les productions en anglais des enfants francophones (et peut-être des apprenants francophones d'une manière plus générale pour certaines d'entre elles) :

- assimilation de certaines voyelles anglaises aux voyelles françaises de timbre proche et effets de filtre sur les cibles des diphtongues;
- assimilation provenant de la transposition de la rythmique du français à l'anglais ;
- assimilation provenant de l'influence de la « prononciation à la française » de certains mots anglais utilisés très couramment dans l'environnement de l'enfant ;
- exagération des caractéristiques non natives de certaines voyelles.

5.2.2. Durée du noyau vocalique

En ce qui concerne les voyelles brèves, nous pensons que les musiciens produiraient des voyelles d'une durée absolue et relative proche de celle des voyelles produites par le locuteur natif. Cette hypothèse ne se vérifie pas et on ne relève pas de distinctions entre les groupes. Bien au contraire, l'ensemble des enfants produit des voyelles encore plus courtes que ne le sont les voyelles anglaises. Ce phénomène de « réduction » de durée pourrait s'expliquer par le fait que les voyelles anglaises sont assimilées aux voyelles françaises de timbre proche. En effet, les voyelles françaises étant plus courtes que les voyelles anglaises « correspondantes », ces dernières sont « raccourcies ». Ce phénomène de réduction est moins marqué en situation de répétition immédiate, notamment pour les enfants du groupe G2, sans doute parce que les enfants restituent les mots tout de suite après le modèle. La prononciation du mot, parce qu'elle est stockée à court terme, n'a pas eu le temps de s'altérer au contact des représentations maternelles stockées à long terme. Il est donc très important de fournir à l'enfant un modèle natif, authentique, en répétition immédiate, puis de le répéter régulièrement au cours de l'apprentissage de manière à éviter les effets de « filtre » dus au stockage à long terme. Ce phénomène confirme qu'il ne suffit pas de placer un enfant au contact d'une LE, pour que sa prononciation soit « native ». Les effets de filtre sont déjà présents à l'âge de 7 ans et il est nécessaire de prévoir une « rééducation », même légère, lors de l'apprentissage. Par ailleurs, les effets d'une prononciation « à la française » chez le professeur de langue seraient désastreux car le modèle fourni à l'enfant en répétition immédiate serait déjà altéré.

Par rapport aux voyelles brèves, les voyelles longues et les diphtongues de l'anglais possèdent des caractéristiques spécifiques que l'on ne retrouve pas en français, ce qui les rend particulièrement saillantes. Leur grande longueur absolue et relative est correctement reproduite par l'ensemble des enfants du corpus. Cependant, on relève une différence entre les groupes car les enfants musiciens produisent des voyelles longues d'une durée encore plus grande que celles produites par les enfants non musiciens. C'est le cas pour les voyelles apparaissant dans les syllabes de type CVC et dans les syllabes de type CV dans les deux situations de discours, mais de manière plus marquée en situation de répétition dans ce dernier contexte, l'allongement apparaît également en situation de répétition immédiate).

D'après ces résultats, il semble que plus un noyau vocalique possède des caractéristiques spécifiquement « natives », plus il sera facilement reproduit par les enfants et moins les effets de filtre n'en altéreront la prononciation. Par ailleurs, sa durée sera allongée plus facilement en situation de répétition différée. Le fait d'avoir une oreille entraînée paraît faciliter ce processus pour les voyelles longues. En revanche, les noyaux proches de ceux de la langue maternelle de l'apprenant seront assimilés à ces derniers, comme c'est le cas pour les voyelles brèves de l'anglais, sans qu'il y ait de différence entre les groupes. L'exagération des caractéristiques « natives » pourrait donc être un principe efficace pour éviter les phénomènes de filtre. La prononciation du professeur doit donc non seulement être native, mais très exagérée.

5.2.3. Stabilité de timbre

Dans l'étude n°5, nous avons posé l'hypothèse que les locuteurs dont l'oreille est entraînée par la pratique musicale restitueraient mieux l'instabilité de timbre spécifique aux voyelles longues de l'anglais à cause d'une meilleure restitution de la rythmique de cette langue et d'une sensibilité moindre au phénomène de crible phonologique. Cette hypothèse se vérifie en situation de répétition immédiate, puisque les enfants musiciens produisent les voyelles longues avec plus d'instabilité que les enfants non musiciens. L'instabilité de timbre affectant les noyaux vocaliques de longue durée constituant l'une des caractéristiques spécifiques de l'anglais, il est possible que les enfants l'ait « repérée » et qu'ils la généralisent à des sons qui ne sont normalement pas concernés. Si c'est le cas, ils feraient preuve de la même stratégie que la stratégie de généralisation mise à jour pour les caractéristiques prosodiques spécifiques de l'anglais, c'est-à-dire les contours en cloche et l'accentuation initiale. Or, c'est justement le comportement qu'ils adoptent pour reproduire les voyelles brèves de l'anglais. Celles-ci sont produites avec une grande instabilité et ce, dans les deux situations de discours, alors qu'elles se caractérisent habituellement par une grande phase de stabilité.

Bibliographie chapitre VIII

- Baken, Ronald J. and Daniloff, Raymond G. (1991). *Readings in Clinical Spectrography of Speech*. San Diego: Singular, 566 p.
- Beaumont et Desmoinaux (2000). *J'apprends, je dessine, je joue : Les Couleurs*. Ed. Cerf Volant, 10 p.
- Butterfield et O'Neill (1997). *Peekaboo, I know !* Instructioal Fair TS Denison, 10 p.
- Dodane, Christelle (1996). *Traitement Comparé de la Musique et de la Langue*. Maîtrise. Université de Franche-Comté, Besançon.
- Dodane, Christelle (1997). *La Langue en Harmonie*. Diplôme d'Etudes Approfondies. Université de Franche-Comté, Besançon.
- Dodane, Christelle (2000). L'apprentissage précoce d'une langue étrangère : une solution pour la maîtrise de l'intonation et de la prononciation. In Guimbretière, 229-248.
- Guimbretière, Elisabeth (2000). *Apprendre, Enseigner, Acquérir : la Prosodie au Cœur du Débat*. Rouen : Publications de l'Université de Rouen, 312 p.
- Hawks, John W. (1994). Difference limens for formant patterns of vowel sounds. In *Journal of Acoustical Society of America*, 95/2, 1074-1084.
- Imagier en Anglais* (1997). Père Castor. Paris : Flammarion, 258 p.
- Kewley-Port, Diane (1990). Thresholds for formant frequency discrimination in isolated vowels. In *Journal of Acoustical Society of America*, 87/1, S159.
- Konopczynski, Gabrielle et Mac Carthy, Brian (1977). Les effets de la lecture à voix haute sur l'intonation. In *International Review of Applied Linguistic in Language Teaching*, 15/1, 230-238.
- Lehiste Ilse and Peterson Gordon E. (1960). Transitions, glides, and diphthongs. In Baken and Daniloff (1991), 286-295.
- Llorca, Régine (1992). Le rôle de la mémoire musicale dans la perception d'une langue étrangère. Dans *Revue de Phonétique Appliquée*, 102, 45-67.
- Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. In *Journal of Acoustical Society of America*, 85, 2114-2134.
- Nakagawa, T., Saito, S. and Yoshino, T. (1982). Tonal difference limens for second formant frequencies of synthesized Japanese vowels. In *Ann. Bull. RILP*, 16, 81-88.
- Piaget Jean (1937). *La Construction du Réel chez l'Enfant*. Paris : Delachaux et Niestlé, 399 p.
- Zenatti, Arlette (1980). *Tests Musicaux pour Jeunes Enfants, avec Applications en Psychopathologie de l'Enfant et de l'Adulte*. Issy Les Moulineaux : Editions Scientifiques et Psychologiques, 53 p.