Autre méthode pour apprendre la prononciation d'une seconde langue

Colleen C. Martin

Laboratoire de Phonétique et Phonologie (UMR 7018)
CNRS/Sorbonne Nouvelle
19, rue des Bernardins - 75006 Paris, France
colleen_martin@libertysurf.fr

ABSTRACT

The object of this study is to test the effect of synthetic voice manipulation for the purpose of second language (L2) learning. We manipulated the acoustic signal of the voice of a monolingual English speaker to make it correspond to a French model. After attempting to copy the pronunciation of the French model voice, the English speaker participated in a series of training sessions in which she attempted to imitate her own (modified) voice. After 5 days of pronunciation training, she once again attempted to imitate the pronunciation of the French stimuli. Judgements by French listeners showed that the pronunciations after training were generally preferred to those produced before training. It is thus possible to use the learner's own modified voice as a model for teaching L2 pronunciation.

1. INTRODUCTION

Alors que l'auditeur remarque un accent étranger dans sa globalité, il peut être pertinant pédagogiquement de démonter la portée de chacun des éléments constitutifs de l'accent global, en séparant les aspects segmentaux des aspects suprasegmentaux et en séparant ces éléments entre eux. Dans cette étude, nous proposons une approche progressive par laquelle l'apprenant peut se concentrer pendant un certain temps sur un élément isolé de son accent étranger, et ainsi travailler sur l'amélioration de cet accent étape par étape.

Selon la méthode traditionnelle, l'apprenant écoute une voix modèle et essaie de répéter ce qu'il entend. Evidemment, chaque locuteur a une qualité de voix qui lui est propre. L'apprenant est donc obligé de prendre en compte non seulement l'ensemble des détails segmentaux et prosodiques de la seconde langue, mais aussi la qualité de voix propre au locuteur. La tâche consistant à apprécier les différences entre phonèmes et allophones communs et différents dans les deux langues serait beaucoup plus facile si l'apprenant pouvait apprendre la prononciation de la L2 sur un modèle simplifié. Nous partons de l'hypothèse que l'apprenant, en entendant sa propre voix modifiée par synthèse partielle, pour rectifier une erreur à la fois, peut reconnaître et préciser l'erreur et donc la corriger.

Nous avons choisi deux erreurs phonétiques, bien documentées dans la littérature, que commettent les apprenants anglophones en parlant français. Toutes les voyelles françaises sont des monophtongues tandis qu'en anglais, il existe des diphtongues et des voyelles dites "monophtongues" qui peuvent être allongées et diphtonguées (Delattre [1], O'Shaughnessy [6]). Il est également bien établi dans la littérature que la présence ou l'absence de vibrations des cordes vocales durant les occlusives est l'indice essentiel qui permet de faire la distinction entre sonores et sourdes en français, et que l'anglais fait la différence entre les occlusives "voisées" et "non-voisées" par le VOT (temps de délai d'établissement du voisement ou Voice Onset Time), les "voisées" anglaises étant souvent dévoisées et les "non-voisées" (tense) étant fortement aspirées à l'initiale de mot et sous l'accent (Flege [2], Keating [3]).

Une étude précédente (Martin [5]) a démontré que les locuteurs natifs français estiment que la voix de l'apprenant anglophone, après modification de la durée vocalique ou de VOT, ressemble plus à la prononciation native que ne le fait la prononciation originale de l'apprenant. L'étude présente teste la possibilité d'utiliser ces stimuli modifiés comme outil pédagogique.

2. MÉTHODE

Pour voir si notre méthode d'apprentissage peut avoir un effet sur les performances d'un apprenant anglophone n'ayant aucune connaissance du français, nous avons adopté une démarche qui sera détaillée par la suite. Le principe consiste à avoir 2 séances de test où l'apprenant anglophone produit des syllabes en essayant de copier le modèle français. Entre les deux tests, il y a 5 jours durant lesquels l'apprenant s'entraîne sur des stimuli synthétisés de sa propre voix corrigée selon des caractéristiques françaises (VOT, durée, cf. plus loin) et reçoit quelques explications sommaires sur les paramètres acoustiques de la correction. Enfin, les productions des deux phases de test sont jugées par des auditeurs français. Notre hypothèse est que l'entraînement aura un effet positif sur la prononciation de l'apprenant et que cette amélioration sera perçue par les juges francophones.

2.1. Séances de test et d'entraînement

Dans un premier temps, nous avons demandé à une locutrice américaine de répéter 30 monosyllabes prononcés par une voix modèle (F1). Cette première série de stimuli (C1) sera présentée, avec une autre série créée après des séances d'entraînement (C2), aux auditeurs francophones qui les jugeront pour déterminer laquelle des prononciations soumises est la plus proche du français. Pendant 5 jours, la locutrice anglophone a suivi un entraînement qui consistait à répéter deux corpus construits avec sa propre voix (enregistrée en anglais auparavant, cf. plus loin) modifiée a) pour la voyelle (AM1) et b) pour l'occlusive (AM2), afin de se rapprocher du modèle français. Finalement, la locutrice a encore entendu la voix modèle, et a de nouveau répété les 30 monosyllabes (C2).

Toutes les séances se sont déroulées sur une période de sept jours.

Table 1 : Déroulement de séances de test et d'entraînement.

Jour	Input	Output
1	F1	C1
2	AM1/AM2	entraînement
3	AM1/AM2	entraînement
4	AM1/AM2	entraînement
5	AM1/AM2	entraînement
6	AM1/AM2	entraînement
7	F1	C2

Les locutrices et les corpus de base

La voix modèle (F1) était celle d'une femme parisienne. Le corpus consiste en une série de syllabes CV composées des six occlusives françaises, /p, t k, b, d, g/ suivies chacune par les voyelles françaises, /i, ɛ, a, o, y/. Ces 30 logatomes monosyllabiques ont été lus dans une phrase cadre induisant une hyperarticulation de la syllabe cible : "J'ai dit ___, pas ___", seule la première occurrence est analysée.

Ce modèle F1 a servi, d'une part, à extraire les paramètres acoustiques qui nous serviront à la resynthèse, et d'autre part, de modèle à imiter par l'apprenant lors des phases de test. La durée moyenne du bruit d'explosion des consonnes sourdes est de 42 ms et la durée moyenne des voyelles de 168 ms.

Le corpus de base pour la resynthèse (A1) a été produit par une locutrice américaine (l'apprenant), monolingue et n'ayant jamais appris le français. Ce corpus de base provient d'une étude antérieure (MARTIN [7]) et consiste en 30 syllabes CV, contenant les six occlusives anglaises, /p, t, k, b, d, g/ suivies chacune par les voyelles anglaises /i, e, æ, o, u/, produites selon la même démarche que pour F1, dans la phrase cadre, "I said ___, not ___". La durée moyenne du VOT des consonnes sourdes est de 105 ms et la durée moyenne des voyelles de 250 ms.

Ce corpus original a seulement été utilisé comme base de manipulation pour créer deux nouveaux corpus composés des mêmes logatomes, modifiés une fois pour avoir la même durée vocalique que les logatomes du corpus F1 (AM1), et une fois pour avoir la même durée de VOT que ceux du corpus F1 (AM2). Ces corpus ont été utilisés pour l'entraînement de l'apprenant.

Avant de modifier, par resynthèse, le signal acoustique de notre locutrice américaine, nous avons, avec l'aide des logiciels *Praat* et *Winsnoori*, mesuré la durée de chaque segment et des quatre premiers formants de chaque voyelle à 1/3 et à 2/3 de sa durée. L'aspiration (et donc le VOT) était mesurée depuis le burst jusqu'à l'onset du F2 de la voyelle.

Les signaux anglais ont subi deux types de modifications : une série (corpus AM1) a été modifiée pour réduire la longueur et la diphtongaison vocalique (réduite en moyenne de 82 ms) et une deuxième série (corpus AM2) a été modifiée pour 1) réduire le VOT (réduite en moyenne de 63 ms) de la consonne sourde initiale et 2) prolonger, quand nécessaire, la partie voisée de la consonne sonore initiale (le bruit initial de la sonore réduit en moyenne de 21 ms et le voisement prolongé en moyenne de 108 ms. La locutrice anglophone n'avait voisé complètement que deux des segments sonores en A1.). Afin d'augmenter la durée du voisement des occlusives sonores, nous avons réalisé une réduplication de période pour le segment voisé. Pour ne pas donner un air haché aux voyelles modifiées, elles ont été coupées, non depuis la fin, mais dans la partie diphtonguée à un croisement par zéro, en laissant l'offset, ce qui crée un son plus naturel. Les réductions et réduplications de période ont été effectuées avec le logiciel *Sound Forge*.

Création de C1 et C2

A chaque séance, l'apprenant a écouté le stimulus deux fois avant de le répéter. Les stimuli de test ont été présentés en ordre aléatoire. L'enregistrement des imitations de la voix modèle de la part de la locutrice anglophone avant tout entraînement (C1) sera, à la suite de l'expérience, comparé à un enregistrement identique fait après les séances d'entraînement (C2).

Il a été donné à la locutrice une explication portant sur les modifications faites au signal acoustique de sa voix. Elle n'a pas eu de support graphique.

Toutes les séances d'entraînement ont été enregistrées. Ces enregistrements n'ont pas été présentées aux juges francophones et n'ont pas d'incidence sur nos résultats.

2.2. Jugements des auditeurs francophones

Les monosyllabes produits par la locutrice anglophone imitant la voix française, avant et après l'entraînement avec sa propre voix modifiée synthétiquement (C1 et C2), ont été enregistrés sur CD, et présentés en paires à trente-trois auditeurs parisiens adultes. Ceux-ci ont marqué leurs choix sur une feuille de réponse numérotée, un numéro étant attribué à chaque paire de stimuli. Nous avons demandé aux sujets de marquer A ou B selon le stimulus pour lequel l'accent français était le meilleur. Chaque paire de stimuli a été présentée deux fois, une fois dans l'ordre A B, et une fois dans l'ordre B A. Il y avait donc 60 réponses par test. Les paires de stimuli ont été présentées dans un ordre aléatoire aux auditeurs-juges, des locuteurs natifs de français standard de la région parisienne ayant une connaissance de la langue anglaise.

3. RÉSULTATS

Des réponses des 33 auditeurs, se dégage une préférence pour les stimuli produits post-entraînement (C2) par rapport à ceux produits avant l'entraînement (C1). En moyenne, 65% des réponses ont montré une préférence pour C2 avec une faible variation de 55% à 77% entre les locuteurs.

D'après ces résultats, les productions faites après les séances d'entraînement sont préférées par tous les auditeurs francophones. Cette tendance n'est pas très marquée mais un test de rangs de Wilcoxon montre que la différence est effectivement significative ($p<0,001$).

Selon un examen des corpus C1 et C2, nous avons détaillé certaines différences entre les deux corpus. Dans la table 2, nous montrons pour chaque aspect (durée d'aspiration des voyelles sourdes, durée de voisement des voyelles sonores, et instabilité (diphtongaison) pour chaque voyelle) dans quelle proportion C2 a été préféré par les auditeurs-juges et quel écart il y a entre la moyenne de ces mesures pour C1 et C2. La diphtongaison de voyelle est la différence en Hz entre les mesures des formants prises à ¼ et à ¾ de la voyelle, une monophtongue ayant une différence de 0Hz. Dans ce tableau, nous montrons la *différence* de la moyenne de diphtongaison pour la voyelle entre C1 et C2. La proportion de consonnes voisées préférées est égale à celle des non-voisées. La durée du voisement a augmentée et la durée du VOT des non-voisées a diminué. Toutes les durées vocaliques sont plus brèves sauf celle de /a/. Tous les formants de C2 sont plus stables, à l'exception du F1 pour /a/ et /o/, et du F2 pour /y/, qui sont plus stables en C1. La voyelle /y/ pose une exception : elle a été nettement préférée en C2, mais son F2 est le moins stable en C2.

Table 2 : Proportion de préférences et Différence de moyennes entre C1 et C2 par facteur.

Facteur	% Préféré	Ecart Durée	Diphtongaison
ptk	49 ,7	-15ms	
bdg	50,3	+3ms	
i	20	-34ms	
F1			16Hz
F2			91Hz
ε	19	-16ms	
F1			11Hz
F2			114Hz
a	15	+10ms	
F1			-11Hz
F2			0Hz
o	20	-11ms	
F1			-17Hz
F2			14Hz
y	25	-19ms	
F1			18Hz
F2			-143Hz

4. DISCUSSION

Avant tout, il faut reconnaître que pour donner une plus grande crédibilité à ce type d'étude, il faudrait en refaire avec plusieurs apprenants-locuteurs. L'étude sur laquelle celle-ci est fondée [5] ne nécessitait qu'un locuteur ; nous avons donc utilisé la même locutrice. Les résultats ici donnent donc une indication quant aux possibilités dans ce domaine, mais ne peuvent être considérés comme définitifs.

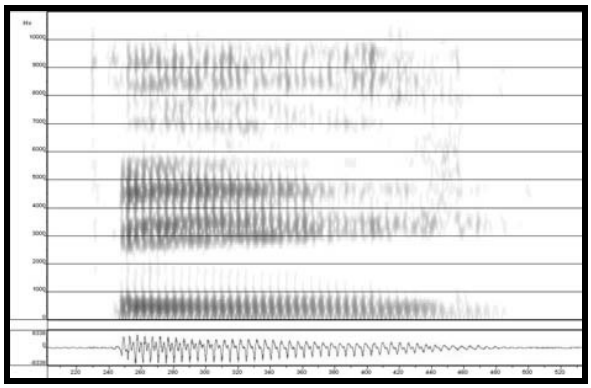
Les jugements des 33 auditeurs francophones ont confirmé notre hypothèse que l'apprenant d'une L2, ici, le français, peut améliorer son imitation de la prononciation d'une L2 significativement en s'entraînant avec sa propre voix modifiée synthétiquement pour corriger certaines erreurs dues à la phonologie de sa L1, ici, l'anglais.

Une comparaison des occlusives non-voisées et voisées de C2 préférées par les auditeurs-juges ne montre pas d'effet de facteur Voisement. A l'opposé, tandis que la seule voyelle monophtongue en anglais (a) a montré la plus faible différence de préférence entre C1 et C2, les trois voyelles diphtonguées (i, ε, o) ont été sélectionnées dans une plus grande proportion en C2 où elles ont été plus brèves et leurs F2 ont été plus stables. La seule voyelle qui différait en stabilité de formants *et* d'arrondissement (y) a été nettement préférée en tant que stimuli de C2 en dépit de l'instabilité de son F2, qui a baissé le plus en C2. Ceci est probablement dû à l'arrondissement de la voyelle. Il est donc tout à fait possible que les jugements des

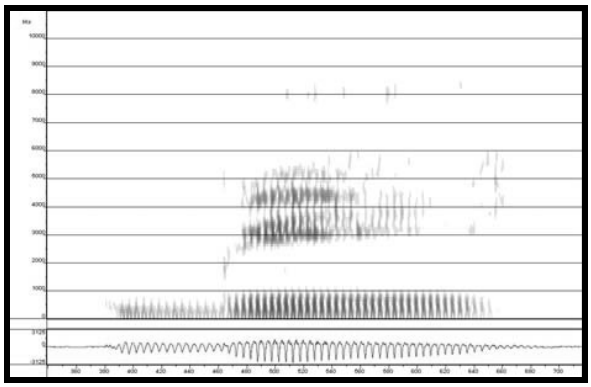
auditeurs francophones en faveur des prononciations post-entraînement soient dus uniquement à la voyelle.

Enfin, dans cette étude, il faut aussi avouer que les séances d'entraînement ont sûrement eu un effet sur la perception phonétique de l'apprenant. Il est évident que pour reconnaître ses erreurs, l'apprenant améliore sa perception de la L2.

Un exemple illustrant une réalisation prise de C1 et de C2 comparées est donné dans la figure 1.



C1 /bi/ voisement b = 0ms, burst b = 18ms, i = 195ms



C2 /bi/ voisement b = 77ms, burst b = 8ms, i = 184ms

Figure 1 : Spectrogrammes montrant les réalisations de C1 et C2 pour le monosyllabe /bi/. Dans la version C2, l'occlusive est voisée, la voyelle plus brève, et les formants plus stables.

5. CONCLUSION

Nous avons vu qu'il est possible d'améliorer sa prononciation d'une seconde langue, même si l'apprenant n'a pas l'occasion de s'entraîner avec un locuteur natif. Nous avons démontré que l'apprenant, en travaillant avec sa propre voix synthétiquement modifiée afin de ressembler à la prononciation cible, peut, après quelques séances d'entraînement, améliorer, de manière significative, sa prononciation de la seconde langue.

Nos résultats pour cette étude sont positifs ; néanmoins, nous considérons que cette méthode est quand même moins efficace que ne le serait un pareil entraînement avec la voix d'un locuteur natif. Néanmoins, ce suggère une autre démarche possible quand l'environnement linguistique idéal ne se présente pas.

La méthode proposée ici offre deux avantages pour l'apprenant qui n'a pas l'opportunité d'apprendre avec un locuteur natif de la L2 : 1) L'apprenant a comme modèle sa propre voix, ce qui peut être plus facile pour lui à imiter que la voix d'un autre locuteur qui peut être très différente de la sienne (masculin ou féminin, grave ou aiguë, par exemple). 2) L'apprenant peut se focaliser sur un seul élément de son accent à la fois, ce qui l'aide à bien distinguer les caractéristiques de sa langue maternelle d'une part, et de la L2 d'autre part.

Les résultats qui ressortent ici sont encourageants, mais il faudrait refaire ce travail avec plus de sujets pour bien asseoir nos conclusions. Il faudrait en particulier faire des modifications successives automatiques de Fo, durée, intensité et formants de la voix de l'apprenant pour approcher celle du modèle, et pour réaliser une démarche de resynthèse plus rapide et plus économique en temps.

BIBLIOGRAPHIE

[1] P. Delattre, *Comparing the Phonetic Features of English, French, German and Spanish*, Heidelberg: Julius Groos Verlag, 1965.

[2] J. Flege et J. Hillenbrand, A differential effect of release bursts on the stop voicing judgments of native French and English listeners, *Journal of Phonetics,* vol. 15, pages 203-208, 1987.

[3] P. Keating, W. Linker et M. Huffman, Patterns in allophone distribution for voiced and voiceless stops, *Journal of Phonetics*, vol. 11, pages 277-290, 1983.

[4] H.S. Magen, The perception of foreign-accented speech, *Journal of Phonetics*, vol. 26, pages 381-400, 1998.

[5] C. Martin, Manipulation of foreign-accented speech: Improving English-accented French *Proceedings of the 15th International Congress of Phonetic Sciences*. pages 957-960. Barcelona, Spain 2003.

[6] D. O'Shaughessy, A study of French vowel and consonant durations, *Journal of Phonetics,* vol. 9, pages 512-521, 1981.