Détection et correction automatique des déviations dans la réalisation de l'accent lexical anglais par des apprenants français

Guillaume HENRY, Anne BONNEAU, Vincent COLOTTE

Speech Group
LORIA/CNRS and INRIA
Campus Scientifique - BP 239
Vandoeuvre-lès-Nancy 54506, France
{Guillaume.Henry, Anne.Bonneau, Vincent.Colotte}@loria.fr

ABSTRACT

The work presented here is developed within a project devoted to the acquisition of English prosody by French learners, using speech technology modifications, and knowledge about L1 and L2 prosody. Our goal is to provide learners with relevant feedback in an automatic manner by comparing the prosodic cues of their realizations to that of a model. We focus on the English lexical accent and propose methods to correct automatically the learner's realizations. We present our strategy and illustrate it through a concrete example.

1. INTRODUCTION

Nos travaux se placent dans le cadre de l'action « Assistance à l'apprentissage des langues » du Plan Etat Région et concernent plus particulièrement l'apprentissage de la prosodie anglaise par des apprenants français.

Depuis quelques années, nous voyons l'émergence de logiciels dédiés à l'apprentissage de la composante orale des langues assisté par ordinateur. Parmi ceux-ci, Winpitch LTL [10], utilisé par des enseignants de langue de l'université, propose une visualisation en temps réel de la courbe mélodique de l'apprenant et possède des fonctions de modifications du signal. BetterAccent [8] propose une comparaison visuelle des tracés prosodiques des apprenants avec ceux d'une référence et indique, à partir de la réalisation modèle, les indices que l'apprenant doit reproduire correctement. Le module prosodique de SLIM [5] propose une évaluation automatique de la durée relative des syllabes au sein d'un mot, afin d'améliorer la réalisation (par un locuteur italien) de l'accent lexical anglais (comparaison avec une référence humaine). Cependant, si on excepte le module d'analyse de la durée de SLIM, ces logiciels ne proposent pas d'évaluation automatique de la réalisation de l'apprenant.

Notre but est de fournir aux apprenants non seulement une visualisation des indices prosodiques, mais également une évaluation automatique de sa production, ainsi qu'une correction auditive. Pour ce faire, nous exploitons des outils d'analyse et de traitement du signal développés dans notre équipe ainsi que des connaissances sur la prosodie de la langue maternelle (L1) et de la langue seconde (L2). L'évaluation est donnée sous la forme d'indices visuels et de petits textes.

Nous avons choisi de nous concentrer dans un premier temps sur la réalisation de l'accent lexical. Ce papier présente notre démarche, que nous résumons ci-dessous, ainsi que quelques tests. Pour établir une évaluation, la production de l'apprenant est comparée à un modèle (provenant d'une référence humaine, pour l'instant). Après une segmentation en syllabes et en phonèmes des deux productions, l'accent lexical est automatiquement localisé. Ensuite, une comparaison entre les indices prosodiques de l'apprenant français et les indices prosodiques du modèle est réalisée dans le but d'affiner l'évaluation. Enfin, on procède à la correction automatique des réalisations. Il s'agit en fait de modifier automatiquement les paramètres prosodiques de la réalisation de l'apprenant dans le but de s'approcher le plus possible de la réalisation du modèle, tout en gardant le timbre et la hauteur de voix de l'apprenant.

Nous présentons tout d'abord le corpus et les outils dont nous disposons (partie 2), puis la méthode utilisée (partie 3). Enfin, nous proposons une première évaluation de cette méthode (partie 4).

2. CORPUS ET OUTILS

2.1. Corpus

Dans le cadre du Plan Etat-Région, nous sommes en collaboration avec des enseignants d'anglais de l'université de Nancy 2 et du secondaire. Ce partenariat a débouché sur la mise au point d'une base d'exercices progressifs, et d'un corpus associé. Le corpus est composé de mots isolés transparents, de phrases (quelques centaines) et de petits textes (quelques dizaines). Il a été enregistré par deux enseignants d'anglais (un homme et une femme) de langue maternelle anglaise. Nous avons sélectionné dans un premier temps une liste de mots isolés transparents qui mettent bien en lumière les problèmes d'accentuation des locuteurs français.

2.2. Outils de traitement du signal dédiés à l'apprentissage des langues

Des fonctions de modification du signal de parole utilisant une version améliorée de TD-PSOLA [4] ont été

développées et incluses dans Winsnoori, logiciel dédié à la visualisation, au traitement et à l'analyse de la parole [9]. Concrètement, l'utilisateur a la possibilité de modifier le contour mélodique d'une phrase ou d'une partie d'une phrase, de resynthétiser le signal et de sauvegarder la modification effectuée. Il en est de même avec le débit : l'utilisateur peut non seulement appliquer un ralentissement du débit de la parole afin d'améliorer l'intelligibilité du signal, mais aussi modifier la durée de chaque segment. Ces deux traitements peuvent s'effectuer soit simultanément, soit séparément. Ils permettent aux utilisateurs de modifier les paramètres prosodiques de leur réalisation afin de se rapprocher du modèle tout en gardant leur timbre et leur registre de voix. Cette correction semble bénéfique dans le processus d'apprentissage d'une langue étrangère [2].

3. METHODE

3.1. Exploitation de la prosodie de L1 et L2

Lors du processus d'apprentissage d'une langue étrangère, la langue maternelle (L1) influence considérablement les réalisations des apprenants dans une langue seconde (L2) [1]. Cette influence est particulièrement importante pour les langues anglaises et françaises qui sont classées dans deux catégories prosodiques différentes, le français étant considérée comme « syllable-timed » et l'anglais comme « stress-timed ».

L'évaluation que nous proposons exploite donc les difficultés spécifiques des apprenants français dans leur réalisation de la prosodie de l'anglais. La réalisation de l'accent lexical anglais est particulièrement difficile pour les français puisque sa place est libre, alors que celle du français est fixe, et qu'il est très marqué acoustiquement, contrairement à l'accent français. Plus précisément, l'accent lexical français est essentiellement caractérisé par un allongement de la durée de la dernière syllabe du mot ou d'un groupe de mots. L'accent lexical anglais est généralement marqué par une forte augmentation de l'intensité, accompagnée d'une modification de la hauteur et d'un allongement de la durée des noyaux vocaliques.

En outre, en anglais, les voyelles des syllabes inaccentuées sont souvent réduites. Ceci est une caractéristique de la langue anglaise que les locuteurs français ont du mal à assimiler puisqu'en français les voyelles inaccentuées conservent leur timbre. Enfin, les occlusives sourdes sont aspirées sous l'accent en anglais.

Lors de la réalisation de l'accent lexical anglais, un français va avoir tendance à garder ses propres schémas prosodiques, et à atténuer les caractéristiques de l'accent anglais. On peut ainsi prédire les principales déviations de l'apprenant. Le locuteur français aura tendance à allonger la dernière syllabe du mot, même lorsque celle-ci ne doit pas recevoir d'accent lexical. En admettant qu'il ait accentué la bonne syllabe, il est probable que les indices prosodiques ne seront pas suffisamment marqués : la syllabe à accentuer ne sera ni suffisamment intense ni suffisamment longue par rapport aux autres, les différences de hauteur seront trop faibles. La prononciation française sera également caractérisée par une absence de réduction et d'aspiration (occlusives sourdes sous l'accent).

En plus des problèmes d'accentuation, les apprenants français ont tendance à mal syllabifier les mots anglais, ce qui rend les comparaisons entre les réalisations anglaises et françaises plus complexes.

3.2. Evaluation automatique

L'effet bénéfique des retours visuels dans l'apprentissage des langues n'est plus à prouver [3]. Dans notre logiciel, les courbes mélodiques et énergétiques, ainsi que les durées de chaque segment et chaque syllabe sont représentées sur le spectrogramme.

La stylisation du contour mélodique est linéaire [7] et l'écart de hauteur entre la syllabe accentuée et les autres syllabes du mot est évalué en demi-tons.

Pour analyser correctement les retours visuels, il faut au préalable segmenter les différentes réalisations en syllabes et en phonèmes. L'étudiant prononce un mot ou un texte tiré du corpus (dans un premier temps, nous nous sommes limités au mot). La segmentation en phonèmes des réalisations des locuteurs anglais (natifs) est réalisée grâce à un alignement texte-parole développé au sein de l'équipe et utilisant des modèles de Markov [6]. En ce qui concerne les locuteurs non-natifs, l'alignement texte-parole nécessite un apprentissage spécifique afin d'adapter les modèles. Cet alignement est en cours de réalisation dans notre équipe. Pour la syllabation des mots isolés, nous utilisons un dictionnaire en ligne qui fournit le découpage en syllabes. Pour les phrases, des règles et des algorithmes de syllabation seront utilisés.

Une fois le signal segmenté en syllabes et en phonèmes, nous proposons une localisation de l'accent lexical. Signalons qu'il n'y a pas toujours véritablement de réalisation d'un accent lexical anglais chez l'apprenant français. Notre méthode vise alors à aider l'apprenant à se rapprocher au mieux des indices prosodiques du modèle.

Afin de détecter la position de l'accent lexical sur des mots isolés, nous avons choisi dans un premier temps de considérer comme syllabe accentuée la syllabe qui porte le pic de F0. Cet indice est efficace pour des mots isolés prononcés sans intonation particulière. Nous donnons les résultats de quelques tests sur la localisation de l'accent en section 4. Il faudra bien entendu utiliser par la suite des critères plus complexes, qui prennent en compte en particulier l'intensité.

Une évaluation plus complète est en cours de réalisation. Elle inclut des appréciations sur la réalisation des différents paramètres prosodiques, et exploite les connaissances sur la prosodie de L1 et L2. En particulier, on recherche si les différences de hauteur entre les syllabes sont suffisamment marquées, ou si la dernière

syllabe du mot n'est pas trop longue alors qu'elle n'est pas sous l'accent (erreur typique d'un locuteur français).

Nous proposons ensuite une comparaison visuelle des indices prosodiques (figure 1, deuxième spectrogramme). On indique sur la réalisation de l'apprenant la position de son accent lexical (rectangle vert) et la position de l'accent lexical de la référence (rectangle rouge). Si les deux rectangles sont superposés, le locuteur a placé l'accent sur la bonne syllabe. On montre également à l'apprenant les durées des syllabes et des voyelles du modèle (en haut du spectrogramme) et on indique si la syllabe accentuée se distingue suffisamment des autres syllabes (flèche et texte inséré). Cette évaluation est complétée par un fichier texte dans lequel sont interprétées les informations visuelles fournies.

Le phénomène de réduction, quand il va jusqu'à la suppression d'une syllabe, nous pose un problème spécifique. En effet, dans ce cas, le nombre de syllabes des réalisations des locuteurs anglais et français diffèrent (le locuteur français a tendance à prononcer toutes les syllabes). Ceci n'est pas considéré comme une erreur car la suppression de syllabes n'est pas obligatoire en anglais. Mais ce phénomène rend la comparaison difficile, voire impossible. C'est pourquoi la solution envisagée actuellement consiste à faire recommencer l'apprenant (après lui avoir indiqué qu'il avait prononcé une syllabe supplémentaire par rapport au modèle anglais) afin que la comparaison soit réalisable.

3.3. Correction automatique des déviations

Dans une version antérieure de notre logiciel, les modifications du signal de parole étaient manuelles, donc réalisées par l'utilisateur. Nous avons mis au point un retour auditif automatique qui remplace les indices prosodiques de l'apprenant par ceux de la référence tout en gardant son registre et son timbre de voix.

Grâce aux techniques de modifications du débit et de la fréquence fondamentale (F0) de la parole développées au sein de l'équipe, nous avons la possibilité de « copier » automatiquement les caractéristiques prosodiques d'un modèle. Dans un premier temps, on aligne les durées relatives des phonèmes de l'apprenant sur les durées relatives des phonèmes du modèle. Dans un second temps, on recalcule le contour mélodique de l'apprenant par une interpolation linéaire du contour du modèle. L'alignement des durées permet d'éviter les effets de bord et en particulier les problèmes de voisement/non-voisement. Ces opérations doivent sauvegarder le timbre et la hauteur de voix de l'apprenant. Pour ce faire, on procède à un recalage du contour mélodique de l'apprenant en jouant sur les moyennes de F0 de l'apprenant et de la référence. Un exemple est donné sur la figure 1 (troisième spectrogramme).

Nous envisageons également une autre stratégie de correction : le renforcement automatique des déviations les plus importantes commises par l'apprenant. En exagérant ces déviations, on espère que l'apprenant prendra conscience de ce qu'il ne faut pas faire.

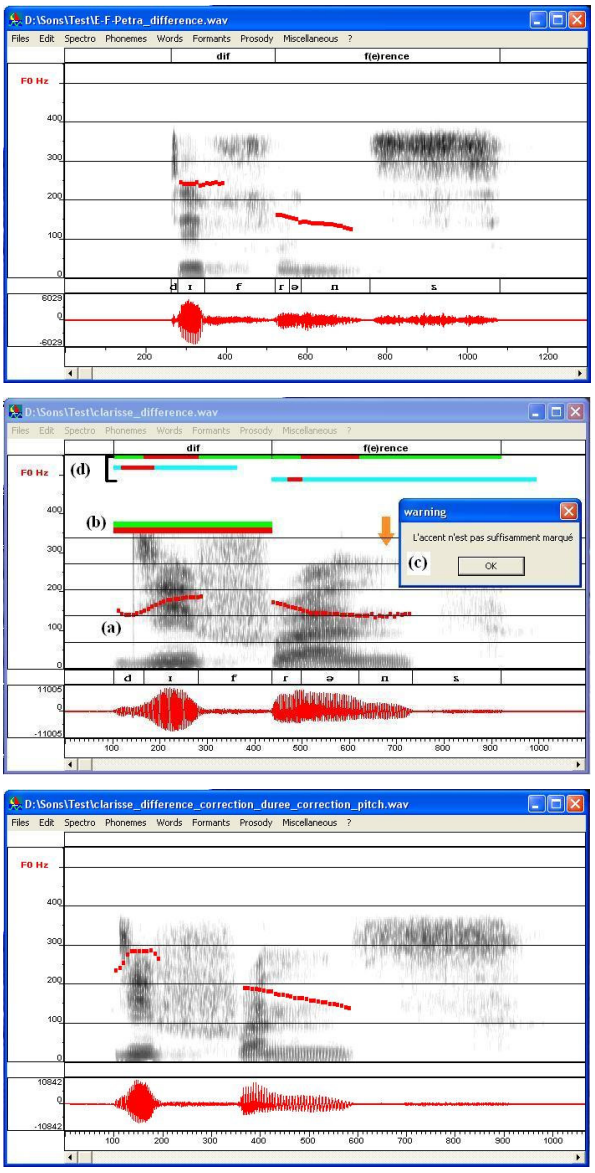

Figure 1 : *Comparaison de la prononciation du mot « difference » : en haut la référence, au centre l'apprenant français et en bas la correction automatique proposées à partir de la voix de l'apprenant. Au centre, la courbe (a) représente le contour mélodique original superposé sur le spectrogramme ; les barres horizontales (b) indiquent les positions des accents ; un petit texte (c) affine l'évaluation, et les durées relatives (d) sont mises à l'échelle de l'apprenant (en haut et en vert pour l'apprenant, en bas et en bleu pour la référence).*

3.4. Un exemple

Nous analysons ici le mot « difference » (« différence » en français) et dans le cas étudié, la référence (une locutrice anglaise) n'a pas prononcé la deuxième voyelle (« ə »).

L'interface Winsnoori nous offre la possibilité d'étiqueter le mot en syllabes et en phonèmes. La transcription

phonétique en API est la suivante : [‘d ɪ f r ə n s]. La comparaison effectuée est montrée sur la figure 1.

Dans cet exemple, l’accent est bien positionné mais on voit nettement que la différence de hauteur entre les deux syllabes est forte chez la locutrice anglaise, alors qu’elle est très faible chez l’apprenant (une locutrice française). On remarque également d’importants écarts dans les durées relatives des voyelles chez les deux locutrices. Ceci montre bien que l’apprenant français a conservé les caractéristiques prosodiques de sa langue maternelle (allongement de la dernière syllabe du mot).

4. PREMIÈRE ÉVALUATION DE LA MÉTHODE

4.1. Identification de l’accent lexical

Afin de valider notre méthode (« grossière » pour l’instant) de localisation de l'accent lexical, nous avons effectué une expérience préliminaire avec les mots isolés transparents extraits de notre corpus et prononcés par une locutrice anglaise. Nous comparons en fait la position de l'accent donnée par un dictionnaire de référence (Le ROBERT & COLLINS) avec celle donnée par notre méthode. La détection est correcte pour 40 mots sur 44. Pour deux des quatre mots restants, la locutrice a accentué le mot de façon différente de celle proposée par notre dictionnaire, mais l’accent qu’elle a réalisé a été indubitablement bien localisé par notre méthode. Signalons du reste que l’accentuation choisie par la locutrice est proposée par d’autres dictionnaires. Une détection erronée provient d’un problème de frontière syllabique, et une autre d’un conflit entre indices prosodiques (l’intensité aurait été pour ce cas un meilleur indice que la F0). Ce type de problèmes sera probablement éliminé par la suite avec des critères de localisation plus complexes.

4.2. Réalisation d’une locutrice française

Les mots étudiés ci-dessus ont été prononcés par une locutrice française. Leur analyse confirme les problèmes listés précédemment (section 3.1). Les phénomènes de réduction extrême (disparition d’une syllabe), observés chez la locutrice anglaise, mais jamais constatés pour la locutrice française, ainsi que des problèmes de syllabation incorrecte, se sont avérés très fréquents.

5. CONCLUSION ET PERSPECTIVES

Dans ces travaux, nous avons proposé une évaluation automatique des déviations dans le cadre de l’apprentissage de la prosodie anglaise par des apprenants français qui exploite à la fois des outils d’analyse et de traitements du signal, et des connaissances sur la prosodie de L1 et L2.

Un travail important doit être mené pour affiner la localisation de l’accent (notamment par une prise en compte de l’énergie, et une meilleure définition des fenêtres de recherche du pic de F0). Nous devons également compléter notre évaluation, en prenant en compte plusieurs critères d’appréciation (décrits en section 3.2). Une étude sur les seuils de déclenchement des différents retours doit être conduite afin de déterminer en particulier à partir de quand une réalisation s’éloigne de la cible (par exemple quand peut on dire que la différence de hauteur entre la syllabe accentuée et les autres syllabes est significative ?). Enfin, nous projetons d’analyser des phrases simples pour lesquelles nous testerons des modèles prosodiques.

BIBLIOGRAPHIE

[1] P. C. Bagshaw. Automatic prosodic analysis for computer-aided pronunciation teaching. *PhD thesis.* University of Edinburgh, 1994.

[2] A. Bonneau, M. Camus, Y. Laprie, and V. Colotte. A computer-assisted learning of English prosody for French students. In *Proceedings of InSTIL/ICALL.* Venice 17-19 June, 2004.

[3] K. De Bot. Visual feedback on intonation I: Effectiveness and induced practice behaviour. In *Language and Speech*, volume 26, 4, 331-350, 1983.

[4] V. Colotte and Y. Laprie. Higher pitch marking precision for TD-PSOLA. In *Proceedings of XI European Signal Processing Conference (EUSIPCO)*. Toulouse, 2002.

[5] R. Delmonte. A prosodic module for self-learning activities. In *Speech Prosody.* Aix-en-Provence, 2002.

[6] D. Fohr, J.F. Mari, J.P. Haton. Utilisation de modeles de markov pour l'etiquetage automatique et la reconnaissance de BREF80. In *Journées d’études sur la parole (JEP)*, Avignon, 1996.

[7] J. ’t Hart. F0 stylization in speech: straight lines versus parabolas. In *Journal of the Acoustical Society of America*, 6, 3368-3370, 1991.

[8] J. Komissarchik and E. Komissarchik BetterAccent Tutor – Analysis and Visualization of Speech Prosody. In *Proceedings of InSTIL*. Dundee, Scotland, August 2000.

[9] Y. Laprie. Snoori, a software for speech sciences. *MATISSE*, 1999.

[10] P. Martin. WinPitch LTL II, a Multimodal Pronunciation Software. In *Proceedings of InSTIL/ICALL.* Venice 17-19 June, 2004.