AUTOMATIC PROSODIC ANALYSIS FOR COMPUTER AIDED PRONUNCIATION TEACHING

Paul Christopher Bagshaw

A thesis submitted for the degree of
Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

1994

Résumé

Analyse automatique de la prosodie pour l'enseignement de la prononciation à l'aide d'ordinateurs

La prononciation correcte du langage oral nécessite la modulation appropriée des caractéristiques acoustiques de la parole afin de communiquer l'information linguistique à un niveau suprasegmental. Une telle modulation prosodique est un aspect-clé du langage oral et une composante importante de l'apprentissage d'une langue étrangère, pour des buts tant de compréhension que d'intelligibilité. L'enseignement de la prononciation à l'aide d'ordinateurs nécessite l'analyse automatique de la parole d'une personne dont la langue en question n'est pas la langue maternelle afin de procurer un diagnostic de la performance de l'élève en comparaison avec la parole d'un locuteur natif. Cette thèse décrit la recherche entreprise afin d'analyser automatiquement les aspects prosodiques de la parole pour l'enseignement de la prononciation à l'aide d'ordinateurs.

Il est nécessaire de décrire la composition suprasegmentale de la parole de l'élève afin de caractériser des déviations significatives de la prosodie native, et afin d'offrir un certain diagnostic de correction. Les théories phonologiques de la prosodie ont pour but de décrire la composition suprasegmentale de la parole pour une langue spécifique. L'argument considéré dans ce document est que la composition suprasegmentale de la parole d'un locuteur non-natif peut être grandement influencée par l'interférence de sa langue maternelle propre, ceci rendant alors inappropriée une représentation phonologique de la prosodie spécifique au langage. De plus, les langues varient au niveau de la manière dont les caractéristiques acoustiques de la parole sont modifiées pour manifester les aspects prosodiques de la parole, et le seul moyen sûr disponible pour décrire la prosodie pour l'enseignement d'une langue étrangère se situe donc au niveau d'une représentation acoustico-phonétique. L'analyse prosodique automatique de la parole présentée dans cette thèse a pour but de procurer une telle représentation acoustico-phonétique.

Les aspects prosodiques de la parole sont décrits dans un domaine syllabique qui est synchronisé avec une segmentation phonétique. On présente un algorithme regroupant les segments acoustico-phonétiques en unités syllabiques. Il est démontré que la syllabification acoustico-phonétique est en corrélation avec la syllabification phonologique. La fréquence fondamentale ($F\O$) de la parole, la durée et l'énergie des unités phonétiques et la qualité vocalique des noyaux syllabiques jouent un rôle important dans la caractérisation des traits prosodiques d'accent, rythme et intonation. Il est nécessaire de déterminer $F\O$ initialement pour l'analyse prosodique automatique de la parole. On considère les problèmes posés par la nécessité de déterminer $F\O$ de manière à minimiser les erreurs d'analyse prosodique, étant donné que le contour de $F\emptyset$ d'un énoncé est affecté par le contenu segmental, les micro-pertubations, l'anatomie et la physiologie du locuteur ainsi que les erreurs entraînées par sa détermination basée sur le signal de la parole. On décrit des méthodes de normalisation du locuteur et de stylisation pièce par pièce des contours de $F\emptyset$ et on fait apparaître une méthode permettant de traiter les contours $F\emptyset$ afin de localiser et caractériser les accents intonatifs et par là de procurer une description acoustico-phonétique de l'intonation. On examine des mesures de durée, énergie, et qualité vocalique du point de vue de leur corrélation avec l'accent de phrase. Le procédé d'analyse de la proéminence de syllabe est compliqué par l'interaction de ces traits acoustiques dans la manifestation de l'accent et par le fait qu'elles sont aussi influencées par des facteurs autres que l'accent. La durée, l'énergie et la qualité vocalique des unités phonétiques varient en fonction du contexte acoustico-phonétique, de la longueur des syllabes, et des niveaux de proéminence des syllabes. La recherche décrite dans cette thèse a pour but de normaliser les traits acoustiques pour les aspects non-prosodiques de la parole, et de combiner les traits acoustiques traités afin de former une description prosodique de la parole. La combinaison de traits acoustiques suppose que l'accent est caractérisé principalement par des variations de durée, énergie et qualité vocalique, et que les accents intonatifs sont caractérisés par la mélodie de la fréquence fondamentale. $F\emptyset$ peut être utilisé comme indice secondaire quant à la localisation des syllabes proéminentes étant donné qu'on observe une chute des accents intonatifs sur les syllabes proéminentes.

On démontre que la description prosodique résultante, déterminée automatiquement, est utile à la comparaison des aspects prosodiques de la parole d'une personne dont l'anglais n'est pas la langue maternelle avec la parole d'une personne dont l'anglais est la langue maternelle.

Translated by Nathalie Vergeynst