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

Zusammenfassung

Automatische prosodische Analyse für Computer-unterstützten Ausspracheunterricht

Die angemessene Modulation der akustischen Merkmale der Sprache ist notwendig für die korrekte Aussprache gesprochener Sprache, damit die linguistische Information auf suprasegmentaler Ebene vermittelt werden kann. Eine solche prosodische Modulation ist Schlüsselaspekt der gesprochenen Sprache und ein wichtiger Bestandteil des Lernens von Fremdsprachen, sowohl für das Verstehen als auch für das Verstandenwerden. Computerunterstützter Ausspracheunterricht beinhaltet automatische Sprachanalyse eines Nicht-Muttersprachlers damit eine Diagnose der Leistung des Schülers im Vergleich zu der eines Muttersprachlers gemacht werden kann.

Es ist notwendig, die suprasegmentale Zusammentstellung der Sprache eines Schülers zu beschreiben, um die bedeutenden Abweichungen von der muttersprachlichen Prosodie zu charakterisieren, und um eine Art korrektive Diagnose zu bieten. Phonologische Theorien der Prosodie beabsichtigen die suprasegmentale Zusammenstellung einer bestimmten Sprache zu beschreiben. Es wird hier argumentiert, daß die suprasegmentale Zusammenstellung der Sprache eines Nicht-Muttersprachlers deutlich durch die eigene Muttersprache beeinflußt werden kann und damit eine sprach-spezifische phonologische Representation unangemessen macht. Außerdem varieren Sprachen in der Art in der die akustischen Charakteristiken modifiziert werden, um die prosodischen Aspekte der Sprache zu manifestieren und die einzig sichere Art, um Prosodie für Fremdsprachenunterricht zu beschreiben, ist deshalb die akustisch-phonetische Representation. Die automatische prosodische Sprachanalyse, die in dieser Dissertation vorgestellt wird, versucht eine solch akustisch-phonetische Representation zu bieten.

Die prosodischen Aspekte der Sprache werden in einer Silbendomäne beschrieben, welche mit einer phonetischen Segmentation synchronisiert ist. Ein Algorithmus wird vorgestellt, der akustisch-phonetische Segmente in Silbeneinheiten gruppiert. Es wird gezeigt, daß die akustisch-phonetische Versilbung mit der phonologischen Versilbung korreliert. Die Grundfrequenz (*FØ*) von Sprache, die Dauer und Energie phonetischer Einheiten und die Vokalqualität vom Silbennuklei spielen eine wichtige Rolle in der Charakterisierung der prosodischen Merkmale Rhythmus, Intonation und Betonung. Die Feststellung von *FØ* ist notwendig als anfänglicher Prozeß in der automatischen prosodischen Sprachanalyse. Die Probleme werden addressiert, die auftreten, wenn man versucht *FØ* so festzustellen, daß minimale Fehler in der prosodischen Analyse auftreten, da die *FØ*-Kontur einer Äußerung beeinflußt wird durch den segmentalen Inhalt, Mikrostörungen, Anatomie und Physiologie des Sprechers und die Fehler, die zusammenhängen mit

der Determination der Sprachwellenform. Methoden der Sprechernormalisierung und die teilweise Stilisation von *FØ* Kontouren werden beschrieben und eine Methode wird hervorgehoben, um *FØ* Kontouren zu verarbeiten damit die Tonhöhenakzente lokalisiert und charakterisiert werden können und dadurch eine akustisch-phonetische Beschreibunger der Intonation geboten wird. Messungen von Dauer, Energie und Vokalqualität werden untersucht hinsichtlich ihrer Korrelation zur Satzbetonung. Der Prozess der Analyse von Silbenprominenz wird verkompliziert durch die Interaktion dieser akustischen Merkmale in der Manifestation von Betonung und dadurch, daß sie ebenfalls durch Faktoren außer Betonung beeinflußt werden. Die Dauer, Energie und Vokalqualität der phonetischen Einheiten varieren aufgrund des akustisch-phonetischen Kontextes, Silbenlänge und Silbenprominenzhöhen. Die Forschung, die in dieser Arbeit beschrieben wird, versucht die akustischen Merkmale zu normalisieren für non-prosodische Aspekte der Sprache, und die verarbeiteten akustischen Merkmale zu kombinieren, um eine prosodische Beschreibung der Sprache zu bilden. Die Kombination von akustischen Merkmalen nimmt an, daß Betonung überwiegend durch Variationen in Dauer, Energie und Vokalqualität markiert wird, und daß Tonhöhenakzente durch die Melodie der Grundfrequenz gekennzeichnet werden. *FØ* kann als sekundäres Merkmal für die Lokalisierung von prominenten Silben dienen, da festgestellt werden kann, daß die Tonhöhenakzente auf prominente Silben fallen.

Es wird gezeigt, daß die daraus resultierende, automatisch determinierte, prosodische Beschreibung nützlich ist, für den Vergleich von prosodischen Aspekten der Sprache eines nicht-muttersprachlichen Englischschülers mit der Sprache eines englischen Muttersprachlers.

Translated by Miriam Eckert