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

Abstract

Correct pronunciation of spoken language requires the appropriate modulation of acoustic characteristics of speech to convey linguistic information at a suprasegmental level. Such prosodic modulation is a key aspect of spoken language and is an important component of foreign language learning, for purposes of both comprehension and intelligibility. Computer aided pronunciation teaching involves automatic analysis of the speech of a non-native talker in order to provide a diagnosis of the learner's performance in comparison with the speech of a native talker. This thesis describes research undertaken to automatically analyse the prosodic aspects of speech for computer aided pronunciation teaching.

It is necessary to describe the suprasegmental composition of a learner's speech in order to characterise significant deviations from a native-like prosody, and to offer some kind of corrective diagnosis. Phonological theories of prosody aim to describe the suprasegmental composition of speech for a specific language. It is argued here that the suprasegmental composition of the speech of a non-native talker can be highly influenced by mother-tongue interference thereby rendering a language-specific phonological representation of prosody inappropriate. Moreover, languages vary in the way acoustic characteristics of speech are modified to manifest prosodic aspects of speech and the only secure means available to describe prosody for foreign language teaching therefore lies in an acoustic-phonetic representation. The automatic prosodic analysis of speech presented in this thesis aims to provide such an acoustic-phonetic representation.

The prosodic aspects of speech are described in a syllabic domain which is synchronised with a phonetic segmentation. An algorithm is presented which groups acoustic-phonetic segments into syllabic units. The acoustic-phonetic syllabification is shown to correlate with phonological syllabification. The fundamental frequency ($F\emptyset$) of speech, the duration and energy of phonetic units and the vowel quality of syllable nuclei play important roles in characterising the prosodic features of stress, rhythm, and intonation. The determination of $F\emptyset$ is required as an initial process in the automatic prosodic analysis of speech. The problems of determining $F\emptyset$ in a way which minimises errors in prosodic analysis are addressed, since the $F\emptyset$ contour of an utterance is affected by segmental content, micro-perturbations, the talker's anatomy and physiology together with errors involved in its determination from the speech waveform. Methods of speaker normalisation and piecewise stylisation of $F\emptyset$ contours are described and a method to process the $F\emptyset$ contour in order to locate and characterise pitch accents and thus provide an acoustic-phonetic description of intonation, is highlighted. Measurements of duration, energy and vowel quality are investigated with respect to their correlation with sentential stress. The process of analysing syllable prominence is complicated by the interaction of these acoustic features in the manifestation of stress and by the fact that they are also influenced by factors other than stress. The duration, energy and vowel quality of phonetic units vary due to acoustic-phonetic context, syllable length and syllable prominence levels. The research described in this thesis aims to normalise acoustic features for non-prosodic aspects of speech and to combine the processed acoustic features to form a prosodic description of speech. The combination of the acoustic features assumes that stress is predominantly marked by variations in duration, energy and vowel quality, and that pitch accents are marked by the melody of fundamental frequency. *FØ* can be used as a secondary cue to the location of prominent syllables because pitch accents are observed to fall on prominent syllables.

The resultant, automatically determined, prosodic description is shown to be useful in comparing the prosodic aspects of the speech of a non-native learner of English with the speech of a native English talker.