Automatic Speech Punctuation Using Prosodic Cues

Tal Levy, Vered Silber-Varod, Ami Moyal
ACLP – Afeka Center for Language Processing, Afeka Tel Aviv Academic College of Engineering
tall1@mail.afeka.ac.il;{veredsv;amim}@afeka.ac.il

Abstract

The paper introduces and integrates prosodic elements into an Automatic Speech Recognition (ASR) engine in order to generate punctuation for the output text. Prosody is defined as the rhythm (timing) and intonational (f_0) aspect of a language and its characteristic signals are ignored in today's ASR engines. Still, research shows that pauses, changes in pitch range and amplitude, global pitch declination, melody and boundary tone distribution, as well as, speech rate variation are all prosodic elements that can be used to incorporate punctuation into ASR and make the output clearer and more readable.

This paper describes a meta-textualized approach for segmenting speech while inserting punctuation marks in parallel to the word recognition phase. The insertion points of the punctuation are determined using acoustic models for the prosodic features of three types of punctuation: period, comma and question mark. Rather than filtering gaps that have features that are consistent with silence and then analyzing other prosodic features, this method searches for segments containing specific prosodic parameters that correspond to speaker-intended punctuation and ranks them according to probabilistic graphs. A Neural Network (NN) was used to estimate the weights assigned to each prosodic feature that corresponds to a particular punctuation mark, and then to connect the separate features to a more accurate combined recognizer. Evaluation was performed using chapters of a read book and compared to the written version of the book. An analysis of the various prosodic features and their contribution to the speech punctuation will be presented.