Word Boundary Detection and Speech Recognition of Noisy Speech by Means of Iterative Noise Cancellation Techniques

Min-In Chung, W. M. Kushner and J. N. Damoulakis

Gould Research Center
40 Gould Center
Rolling Meadows, IL 60008

Abstract

This paper describes the approach for enhancing the performance of a speaker-dependent, discrete word recognition system in a noisy environment by means of cepstral subtraction techniques applied iteratively. A series of experiments have shown that these iterative methods provide enhanced performance in the word boundary detector which results in a significant improvement in speech recognition accuracy.

Description of Effort

This paper describes the approach for enhancing the performance of a speaker-dependent, discrete word recognition system in a noisy environment by means of spectral or cepstral subtraction techniques applied iteratively. A series of experiments have shown that these iterative methods provide enhanced performance in the word boundary detector which results in a significant improvement in speech recognition accuracy.

Several experiments have been performed in various broadband noise environments simulated in a reverberent room. The recordings of noisy speech have been obtained by acoustically mixing noise with the original noise-free speech signal. A 36 word vocabulary consisting of the digits 0 through 9 and other command words made up the template library. The resultant signal-to-noise ratio (S/N) of the input noisy speech ranges from 5 to 30 dB with a bandwidth of 5 kHz.

The iterative noise reduction methods consist of repetitive applications of the spectral or cepstral subtraction techniques. During each iteration the noise reduction algorithm updates and computes a noise threshold function representing an estimation of the components of the broadband noise contained in the input signal. A scaled version of the noise threshold function is subtracted at each iteration from a corresponding function containing the noisy input signal. The subtraction provides a new signal containing less noise power than the previous signal. Experiments have shown that such iterative noise reduction methods can significantly improve the S/N of the input signal without significantly distorting the speech. Word boundary detection takes place on the resultant improved signal.

Following the noise reduction operations, the processed speech is passed into a conventional discrete word recognition system using linear predictive coding and employing dynamic programming with the Itakura distance measure for pattern matching. A simple, multiple-threshold energy-based word boundary detector is used to locate the starting and ending points of the words to be recognized. Two different approaches are used to evaluate the performance of the discrete speech recognition system. In the first method, the words are corrupted by broadband noise to produce 30, 20, 10 and 5 dB S/N. Subsequently, they are compared with a template library generated in a 60 dB environment. The noise corrupt input words are compared with the template library before and after they are processed by the iterative noise reduction methods. In the second method, the experiment is repeated except that the template library words are also processed by the same noise reduction methods.

The experiments have shown that the performance of the integrated speech recognition and iterative noise cancellation system is significantly better than the system using no noise cancellation techniques and that the templates should be processed in the same manner as the input words. A real-time design architecture with proper real-time execution timing requirements is presented.

CH2118-8/85/0000-1838 $1.00