Robust Speech Recognition and its Robot Implementation

by

Professor Yoshikazu MIYANAGA

Hokkaido University

Sapporo, Japan

Abstract:

Currently, there are many equipment who has the human interface of automatic speech recognition (ASR). If you would use ASR in your small terminal outside, your ASR must be strong against various noises. In this topic, some of noise robust speech recognition are introduced. In addition, a robot application including the ASR system is also introduced.

This topic first introduces a noise robust speech recognition system with advanced speech analysis techniques named MSC (Modulation Spectrum Control)/DRA (Dynamic Range Adjustment). The dynamic range of speech features obtained from noisy speech is usually smaller than that from the same speech without noise since some speech features are hidden in noise. This difference may cause recognition errors. Therefore the adjustment of dynamic range under various noises can realize the accurate extraction of speech features. DRA and MSC focus on the speech feature adjustment. DRA normalizes dynamic ranges and MSC eliminates the noise corruption of speech feature parameters. The experiments on isolated word recognition were carried out using 40 male and female speakers for training, and 5 male and female speakers for recognition. The result of recognition rate improving from 17% to 64% versus running car noise at -10dB SNR is shown as an example. In addition to these new algorithms, we have developed a new LSI suitable for a robot system.

This topic presents a scalable architecture for realizing a real-time speech recognizer based on a word HMM (Hidden Markov Model). HMM-based recognition algorithms are classified into two acoustic models, i.e., phone-level model and word-level model. The phone-level HMM has been widely used in current speech recognition systems which permit large-sized vocabularies. Whereas the word-level HMM has been constrained to small-sized vocabularies because of high computation cost, it can realize excellent recognition performance. In order to overcome the shortage, we adopt the scalable

architecture focused on the word HMM structure. The proposed architecture can flexibly improve recognition performance and extend word vocabularies. The computation time is hardly increasing.

In order to demonstrate practical solutions, we have designed and evaluated a total system recognizer including speech analysis and noise robustness on a 0.18μm CMOS standard cell library. The recognition time is 35.7μs/word at 128MHz operating frequency. The recognizer can achieve over middle-sized vocabularies in real-time response. This chip has been implemented into several robots. This topic shows its performance.