

学位論文題名

A Study on Noise Robust Continuous Speech Recognition System Using Block Based Dynamic Range Adjustment

(ブロックに基づくダイナミックレンジの騒音頑健な連続音声認識システム)

学位論文内容の要旨

In recent years, great progress has been made in continuous speech recognition (CSR). Thus, state-of-the-art CSR systems can yield high recognition accuracy under clean condition. However, in practice, the noise can substantially diminish recognition rate.

Thus, much research has focused on improving CSR robustness. Work has been done on modifying the basic speech signal features. Nevertheless, conventional noise-robust CSR methods still suffer from various limitations in practical applications, e.g., the sources of variability introduced by corrupted speech signal. This motivates our study of CSR system for a more realistic model under both clean and noisy conditions. Based on Japanese newspaper article sentences (JNAS) database, our CSR system is built with the HMM Toolkit (HTK) and the test tool Julius that has been developed by Japanese researchers. For the characteristics of Japanese pronunciation, this thesis explores noise-robust and speaker-independent CSR.

In our research, we have evaluated the benefits and challenges of conventional CSR methods under clean and noisy environments. We have developed a noise-robust CSR system for evaluation of the proposed method. In the training stage, running spectrum analysis (RSA) and cepstral mean subtraction (CMS) are used to reduce the distortion embedded into training data set. The CMS is also used for the time invariant noise reduction of an observed speech waveform in recognition stage. The proposed method is used in the recognition stage as well. Then, a modified dynamic range adjustment (DRA) method is applied for CSR. The speech waveform is observed over an unlimited duration, because any continuous data are supposed into the CSR system. From the observed noisy continuous speech waveform, we select all short sentences. Each short sentence is divided into blocks with limits that maintain the relationship between the neighboring blocks. The proposed method effectively increases the similarity between clean and noisy speech features. As a result, the recognition rate has been improved by up to 15 percent in various noises with 10 dB SNR.

学位論文審査の要旨

主査	教授	宮 永 喜 一
副査	教授	野 島 俊 雄
副査	特任教授	小 柴 正 則
副査	教授	小 川 恭 孝

学 位 論 文 題 名

A Study on Noise Robust Continuous Speech Recognition System Using Block Based Dynamic Range Adjustment

(ブロックに基づくダイナミックレンジの騒音頑健な連続音声認識システム)

本論文は、連続音声認識に関する新しい方式の提案を行い、その方式実現と評価を行っている。音声認識は、現在、クラウドネットワークを利用した連続音声認識システムと、オートノーマス型のフレーズ認識システムに分かれている。認識率は、認識対象をかなり限定しているオートノーマス型のフレーズ・孤立単語認識システムが高い性能を維持しているが、認識できる対象は限定されており、より一般の応用には、クラウド型の音声認識システムが必要とされる。

オートノーマス型の音声認識システムが高い認識率を維持できる理由の一つに、ロバスト音声処理があげられる。ロバスト音声処理は、環境雑音が多くて、録音できる音声は、10dB 以下の SNR でも、必要な音声の特徴量を推定し、認識する手法である。この方式を連続音声に利用し、連続音声認識の性能を飛躍的に向上させることが本研究の目的である。

本論文は、下記のような構成となっている。

第 1 章は、音声認識の概要について述べている。

第 2 章は、連続音声認識システムの説明を行っており、音声分析から、音素の組合せによる音響認識、その結果から得られる音素記号列からの音声言語認識などに技術について説明している。

第 3 章は、オートノーマス型音声認識で利用されているロバスト音声認識手法について述べている。特に、本論文で拡張した、ランニングスペクトルフィルタリング (RSF) と、ダイナミックレンジアジャストメント (DRA) に関して詳しく説明している。

第 4 章は、本論文で提案している連続音声用のロバスト音声認識について説明している。連続音声と孤立音声フレーズとの大きな違いは、無音区間の存在である。そのため、本論文では、連続音声から、音響的に分割可能な音声ブロック (推定音声センテンス) というものを定義し、その音声ブロックを高速にかつ適切に求める方式を提案した。その後、音声ブロックについて、孤立音声に対して利用されていたランニングスペクトル分析法 (RSA) と DRA を導入し、結果として、従来の連続音声認識システムで実現されていた音声認識率を約 15%(10dB SNR) 向上させるたことを示した。

第 5 章は、トータルシステムの性能評価を行っている。

第 6 章は、上記の各章のまとめと、本研究の総括を行っている。

以上より、本論文では、雑音環境下における連続音声認識システムに関するシステムの設計・開発について詳細に検討し、特に新しいロバスト連続音声分析方式の提案・開発・評価を行った。これに

より新情報化社会に必要とされる新しい音声認識技術の実現を行った。

これを要するに、筆者は、雑音環境でも高精度な性能を有する新しい連続音声認識システムの提案・開発・評価を行った。これにより、音声認識技術に関する多くの有益な知見を得ており、情報科学・工学の分野に貢献するところ大なるものがある。

よって筆者は、北海道大学博士(工学)の学位を授与される資格あるものと認める。