

審査の結果の要旨

氏名 小橋川 哲

本論文は「Study on Efficient Prior Control for Realizing Practical Systems of Speech Recognition（音声認識の実用化のための効率的な事前制御技術に関する研究）」と題し、音声認識の実用化のために、利用場面に応じて活用できる前提条件に基づく事前制御技術を導入する事で、背景雑音や話し言葉音声に頑健な技術を開発したもので、全7章からなり、英文で書かれている。

第1章は「Introduction」であって、音声認識の現状について述べた後、その実用性向上のために、本論文では、タブレット端末利用時における耐雑音音声認識、コンタクトセンタや議会における話し言葉音声認識、の2つの課題を対象とするとしている。また章の構成が示されている。

第2章は「Proposed model adaptation and normalization using pre-observed noise」と題し、タブレット端末利用の音声認識で問題となるS/N比の変動および空間伝達特性の影響に対応するため、事前観測雑音を用いたモデル適応と正規化技術を開発している。定常的な背景雑音は事前に観測可能という前提を置き、事前観測した加法性雑音のみを用いて、スペクトルサブトラクション(Spectral Subtraction)による加法性雑音抑圧手法、モデル合成法(Parallel Model Combination)による加法性雑音適応手法、ケプストラム正規化(Cepstral Mean Normalization)による乗法性雑音正規化手法を融合させ、S/Nの変動や空間伝達特性である乗法性雑音(歪み)に強い音声認識方式を開発している。

第3章は「Fast unsupervised adaptation based on efficient statistics accumulation using frame independent confidence within monophone statics」と題し、コンタクトセンタでの利用を念頭に、フレーム独立信頼度を用いた事前高速教師なし適応技術を開発している。コンタクトセンタでは、限られた処理時間で認識精度の向上を達成することが求められるが、これを達成するため、音響モデル中の限られたGMM(Gaussian Mixture Model)を用いて事前処理を行い、その結果を利用して、最尤線形回帰法(MLLR: Maximum Likelihood Linear Regression)と性別選択を融合することで、認識率を向上させる技術を開発している。コンタクトセンタ利用ではリアルタイム処理は必須出ないという前提を活用した処理となっている。

第4章は「Efficient data selection for speech recognition based on prior confidence estimation using speech and monophone models」と題し、事前信頼度により、コンタクトセンタでのテキストマイニングに適した音声データのみを選択して認識対象とする手法を開発している。認識信頼度が低い音声データはテキストマイニング処理では棄却されるが、音声認識プロセスでは、そのようなデータの処理に過剰な計算コストが割かれている。これに対処するために、開発した手法では、音響モデル中の限られたGMMを用いて高速に信頼度を推定し、精度が高い音声のみを音声認識対象とする。

第5章は「Efficient beam width control to estimate excessive recognition time based on score range estimation」と題し、コンタクトセンタ向けの技術として、認識精度が低く後段で棄却すべき低音質の音声データに対する音声認識処理の計算コストを削減する手法を開発している。これは、事前ビーム幅制御による認識処理時間安定化と呼ばれる技術で、対象音声データに対し、音響モデル中の限られたGMMを用いて予め高速に走査を行い、得られた事前スコアから、音声認識処理中の探索空間の広がりやを制御する手法である。大量音声データの中に含まれる低精度・低品質音声には計算量が無駄にかけなくて良いという前提に基づき、音声認識処理時間を安定化させる技術となっている。

第6章は「Fast acoustic pre-processing against recording environment and speaker changes for parliamentary meeting transcription」と題し、議会録作成支援音声認識システムの実用化に向け、議会の收音環境および話者・環境の変動に対応するための、高速事前音響処理を開発している。これは、対象の議会音声区分・蓄積されて送られてくるという前提を活かしたもので、チャンネル選択、話者インデクシング、特徴量正規化、教師なし適応といった複数の事前音響処理を高速に行い、限られた計算時間の中で、変動する音声に対して高い音声認識精度を実現する方式である。

第7章は「Conclusion」であって、各章の概要を述べた上で、今後の課題・展望を整理している。

以上を要するに、本論文は、タブレット端末やコンタクトセンタなどの利用場面に即した導入可能な事前制御技術を開発し、従来の音声認識方式では得られない性能と効果を達成したものである。背景雑音や話し言葉音声に頑健な実用的な音声認識手法を開発したものであって、音声認識の発展に大きく寄与したものであり、電子情報学に貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。