

審査の結果の要旨

論文提出者氏名 齊藤 隆

本論文は「新たな発話に対する柔軟性を備えたコーパスに基づく音声合成システムに関する研究」と題し、波形接続型コーパスベース音声合成において、音声コーパス作成のコストの低減、新しいスタイルの音声の再現性の向上、を実現したもので、全6章からなる。

第1章は「序論」であって、音声合成の内容とその歴史・現状を概観した上で、現在のコーパスベース音声合成では、新しい音声に対して、合成システム構築に至るコストが高いという問題があることを指摘し、本論文の意義、目的を述べている。その上で、研究の方向を、韻律変形を伴う波形接続型コーパスベース音声合成に定めている。また章の構成が示されている。

第2章は「本研究のベースとなる音声合成システム」と題し、まず、本論文で提案する音声合成システムの特徴が、多種多様なコーパスへの高い適応性を有する点にあるとし、その具体的な内容として、話者ごとの韻律的特徴を有する波形素片（ボイスフォント）の作成、コーパスの作成を容易に行えるツールの開発、適用分野の録音音声とのシームレス音声合成による高品質音声合成の実現を挙げている。次に、システムの概略を述べた上で、システムでは、ピッチ同期波形重畳法である TD-PSOLA を採用しており、そこで問題となるピッチマークについて、ウェーブレット変換を用いたマーキング法を開発したとしている。96%の自動抽出精度を達成している。さらに、合成単位とその選択について詳細な検討を行い、音素環境適合度と韻律変形歪による単位選択で良好な結果が得られることを確認している。

第3章は「発話への柔軟性をもつイントネーションの生成」と題し、基本周波数（F0）形状素片を用いる手法を提案している。この手法は、素片を用いることで、オリジナルデータの韻律の特徴をよく再現できる点にある。素片の選択・接続手法について述べた後、アニメーションを含む種々の発話スタイルに対して、予測精度の評価実験を行い、0.21oct.の誤差で予測できたとしている。

第4章は「音声合成のための頑健性を持ったセグメンテーション」と題し、まず、セグメント信頼度を指標とした音素境界の検出について述べている。繰り返し分割による最適化を行うことで、DP 照合によって 10.9ms の境界誤差と、HMM 強制アライメントと同等な性能を得ている。音素モデル構築などのコストがかからないという特徴がある。次に、音素区分の情報に加え、呼気段落、アクセント句、ピッチマークなどの発話の区分情報を同時に得、F0 形状素片を選択する手法について述べている。

第5章は「話者データベース構築のツール化」と題し、ボイスフォント作成を支援するボイスフォントビルダーを構築した結果について述べている。分析結果の表示、編集に工夫を凝らしたもので、音声の専門家でなくても同等品質の音声合成用ボイスフォントを構築可能なことを示した。また、作成時間も、従来と比べ3分の1程度に開発時間を短縮し得ることを示した。

第6章は、「結論」であって、各章の概要を述べた上で、開発した音声合成システムの特徴が、小さな負担で新しいスタイルの発話の特徴を保持した音声を合成し得る点にあるとし、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、TD-PSOLA に基づく波形接続型コーパスベース音声合成において、新しい音声から、専門知識無しに、従来よりも大幅に少ない時間で、合成用音声コーパスを構築する手法を提案・開発するとともに、ピッチ同期波形重畳法で問題となるピッチマーク検出を高精度に行う手法を開発したものである。韻律について F0 形状素片を準備するなどの特徴があり、それにより、音声コーパスの発話スタイルの再現性に富んだ音声合成を可能としている。従来の波形接続型音声合成で問題となっていた音声コーパス作成のコストを低減し、新しいスタイルの音声の再現性を向上させたもので、音声合成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。