

論文内容の要旨

博士論文題目 High-quality and Flexible Speech Synthesis with Segment Selection and Voice Conversion
(素片選択と声質変換を用いた高品質かつ柔軟な音声合成)

氏名 戸田 智基

(論文内容の要旨)

音声合成は機械から人間へ情報を伝達する上で重要な役割を担う。中でもテキスト音声合成 (TTS) は任意のテキストから音声を作成する技術であり、実用面における有効性の高さから研究開発が勢力的に行われている。現在の TTS の主流は大規模な音声データ処理に基づくコーパスベース方式と呼ばれるものであり、初期のルールベース方式と比較して劇的な進歩を遂げた。しかし、安定して自然性の高い音声を合成できるまでには至っておらず、様々な話者の音声を容易に合成するなどといった柔軟性も持ち合わせていない。

本論文では音声合成における二つの問題に取り組んでいる。一つはコーパスベース TTS における合成音声の自然性の改善であり、もう一つは話者性制御の改善である。前者に対しては、自然性に重要な影響を及ぼす要因として、音声コーパスからの最適な音声素片選択方法とそれに用いる選択尺度に着目し研究を行っている。後者に関しては、話者性制御に関しては、少量の音声データから多様な話者の音声合成を可能とする声質変換技術に着目して研究を行っている。

TTS における素片間の接続の不連続感を緩和するという観点から、音素単位とダイフォニ単位に基づく新たな素片選択法を提案し、知覚実験結果からも従来法より自然な音声を合成できることを実証している。とくに、音声コーパスから最適な素片を選択する尺度として、知覚実験結果から、RMS (Root Mean Square) コストが適していることを示した。

従来の統計的処理に基づく声質変換法である混合正規分布モデルに基づく変換法は、各話者の特徴量間の相関を用いることで連続的かつ精度の高い特徴量の変換を可能とするが、統計的処理により変換スペクトルは過剰に平滑化されたものとなり、変換音声の品質劣化が生じる。この過剰な平滑化を防ぐために、周波数軸伸縮によるスペクトル変換処理を導入した新たな声質変換法を提案している。評価実験結果から、提案法は従来法と比較し、同等の話者性変換精度を維持しながら変換音声の品質を改善できることを示している。

氏名	戸田 智基
----	-------

(論文審査結果の要旨)

音声規則合成 (TTS) は機械から人間へ情報を伝達する上で重要な役割を担い、実用面における有効性の高さから研究開発が勢力的に行われている。現在の TTS の主流は大規模な音声データ処理に基づくコーパスベース方式と呼ばれるものであり、初期のルールベース方式と比較して劇的な進歩を遂げたが、安定して自然性の高い音声を合成できるまでには至っておらず、様々な話者の音声を容易に合成するなどといった柔軟性も持ち合わせていない。これらの問題に対して研究開発を行い、ATR での TTS の開発に研修研究員として参加し、音声規則合成システム「XIMERA」を完成させた。この TTS システムには、戸田君の研究成果が多く含まれている。とくに、TTS における素片間の接続の不連続感を緩和するという観点から、音素単位とダイフォニ単位に基づく新たな素片選択法を提案、音声コーパスから最適な素片を選択する尺度として、知覚実験結果から RMS (Root Mean Square) コストが適していることを示したことは評価できる。

従来の統計的処理に基づく声質変換法は、変換スペクトルが過剰に平滑化されたものとなり、変換音声の品質劣化が生じていました。この過剰な平滑化を防ぐために、周波数軸伸縮によるスペクトル変換処理を導入した新たな声質変換法を提案しています。この研究成果は、多様な声質の合成音を高精度に生成する研究として評価できます。この論文で、この3月に電気通信普及財団から「第18回 テレコムシステム技術学生賞」を受賞します。

戸田智基君は、音情報処理学講座の博士課程2年の学生ですが、音声の声質の変換および音声規則合成システムの研究で、大きな成果をあげ、国内はもとより、海外からも優秀な研究者として認められています。戸田君の研究活動の高さは、4編の学術論文、12編の国際会議論文として発表していることから推測できます。D2のときに学振の博士に応募して採用され、4月からは学振のPDとして研究活動を続けます。

以上述べたように、本論文の音声規則合成の分野への貢献は、極めて大きい。平成15年1月16日に開催した公聴会の結果も参考にして、本博士論文の審査を行い、本論文は博士論文(工学)の学位論文として十分な価値があるものと判断した。