論 文 の 内 容 の 要 旨

論文題目　　テキスト音声合成のための基本周波数パターン生成過程モデルに基づく柔軟な韻律制御
氏　　名　　越智 景子

話し言葉である音声言語は，人間と機械間のインターフェースにおける入出力手段として利用が広まりつつあり，高度に発達した情報機器の操作が，人間同士の意思伝達のようにより自然に行えるようになることを目指した研究が各分野で行われている．

テキスト音声合成とは，任意のテキストを音声信号に変換する技術である．音声出力を利用したインターフェースにおいては，発話内容を適切にユーザに伝達するためには，合成した音声の音質のみならず韻律的特徴の適切な制御が重要である．音声の韻律的特徴とは，アクセントやイントネーション，発話速度などである．これは音声の個々の音韻に対応する分節的特徴よりも広い範囲に現れる特徴であるため，超分節的特徴とも呼ばれ，文字言語にも含まれる語義・統語・意味などの言語情報の伝達は勿論，意図・態度・感情といったパラ言語情報・非言語情報の伝達にも重要な役割を果たしている．近年の研究により，長時間の音声資料の利用から音声合成音質には飛躍的な改善が見られているが，韻律の制御は依然大きな課題となっている．とくに，韻律的特徴の制御は多様な発話様式での音声合成といったより柔軟な合成技術を目指すうえでは，不可欠である．

従来は朗読調で均一な話し方による音声合成が一般的であったが，状況に応じて伝えるべき情報の強調や，発話様式を変化させた音声出力を実現することにより，ユーザにより円滑に情報の伝達が可能になると考えられる．しかし，既存のコーパスベース音声合成システムにおいて強調といった合成音声の発話様式の多様化を行うには，新たに同一話者で当該の様式での発話を収録しなおす必要があるという問題がある．これは多様で柔軟な発話様式の合成を行ううえでコーパスの膨大化を招く結果となる．

そこで，本研究では，アクセントや統語構造といった言語情報と明瞭な対応関係を持つ基本周波数パターン生成過程モデル（F0 モデル）使用することにより，比較的少量のコーパスで多様な韻律制御を行う手法を提案する．

音声の韻律的特徴を表す主要な物理量は，基本周波数の時間変化パターン(F0 パターン)，単音の持続時間の長さである音韻継続長，息継ぎである休止，声の大きさに対応するパワーが挙げられる．ピッチ・アクセント言語である日本語ではとくに F0 パターンが重要な役割を果たしているといえる．

F0 モデルとは，F0 パターンをフレーズ成分とアクセント成分の重畳として表現するもので，それぞれの成分が発話内容のテキストの言語情報と明確な対応を持っている．また，各成分はフレーズ制御機構，アクセント制御機構に対するフレーズ指令，アクセント指令という離散的な信号列の入力結果としてモデル化されるため，指令の生起時刻と大きさという少ないパラメータで F0 パターンの曲線を近似できるという利点がある．

F0 モデルの利点を活かし，我々はテキストの入力から韻律的特徴量を制御する手法を開発した．すなわち，休止の位置および長さ，音韻継続長，F0 モデルの指令のパラメータを各韻律的特徴間の密接な関係を考慮して推定することによって高品質な韻律を生成するものである．

さらに，柔軟な音声合成の実現を目指して強調のための焦点を付与した音声を対象とし，既に開発した F0 モデルの枠組みによる韻律制御手法を応用した合成手法の開発を行った．発話の焦点は話者が特に強調したい部分に置かれるもので，音声合成において伝わりやすい音声出力を目指すうえでその制御は重要な課題であるといえる．
そこで，韻律のコーパスを用いた手法により任意のテキストから指定した位置に焦点を置いた音声の合成を行う手法を開発した．焦点をある箇所に置くことによって生じる韻律的特徴の変化に着目し，既存の焦点を想定しない音声合成に焦点制御機能を付加するものである．特定の文節に焦点を当てて読み上げた音声とどこにも焦点を指定せずに読み上げた音声を収録し，F0 パターン F0 モデルを用いて定量的に記述し，それをもとに，比較的少量のコーパスから任意のテキストと強調する文節の指定した音声を合成するシステムを開発した．

提案手法では，焦点のある発声とない発声について F0 モデルの指令の差分について機械学習を行う．差分により焦点付与を想定しない韻律制御（ベースライン手法）で生成した指令を修正することによって，焦点付与を実現する．それにより少量のコーパスで学習が可能であり，必ずしもベースラインと同じ話者の音声を用意する必要がない．さらに，この焦点制御における F0 モデルの差分の推定値を用いて強調の程度を補間する手法を開発した．