

Title	人間の音声生成メカニズムに基づく音声合成方式に関する研究
Author(s)	平井, 啓之
Citation	
Issue Date	2008-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/4199
Rights	
Description	Supervisor: 党建武, 情報科学研究科, 博士

要旨

近年、コーパスベースの波形編集型音声合成技術の進歩により、高品質な音声の合成が可能となってきた。しかし、この方式では収録されていない声質の音声の合成は不可能である。よって、前後の音素の環境や声の高さが異なると音声の周波数特性が変化する為、合成する環境に合った全ての音素をデータベースに用意しておく必要がある。また、感情や個人性などの多様な音声の生成を行う場合は異なる発話様式毎に異なるデータベースを用意する必要がある。その結果、データベースのサイズの増大と言う欠点が生じることになる。今後、ますます複雑になる様々な機器と人間とのインターフェースとして音声合成を用いる為には、機器への組み込みが可能なコンパクトでかつ多様な音声合成することのできる音声合成方式の開発が不可欠であろう。このような方式の1つに人間の音声生成メカニズムに基づく音声合成方式がある。この方式は1)パラメータの補間特性が良く大幅なパラメータ圧縮を行うことが可能、2)人間の音声生成メカニズムを分析し解明した結果を直接モデルに反映することが容易である為に従来データベースに蓄積せざるを得なかった様々な特性を持つ音声を計算により生成することが可能、と言った特徴がある。そのため、これまでに幾つものモデルが提案されてきた。しかし、それらのモデルが実用化に至っているとは言い難い。その要因は、1)音声生成機構には未だ明らかにされていない生理機構が存在すること、2)音声生成機構はソース(音源)・フィルタ(声道形状)の結合としてモデル化できるがそれぞれを独立した要素としてモデル化が進められてきたこと、3)モデルの構築に用いる声道形状データの量と質が不十分であったこと、などが考えられる。そして、これらの要因は発話時の声道の観測が困難であった為にこれまで解決することができなかったのではないかと考えている。この問題に対し、近年 MRI の技術の進歩により様々な条件で発話時の声道の撮像が可能となってきた。よって我々は最新の MRI 技術を用いてこれまで抱えてきた問題を再考することで、この方式の実用化を目指すことにした。

本論文では、人間の音声生成メカニズムに基づく音声合成方式に関する我々の研究を論じる。ここで我々は、最新観測手段によりソースおよびフィルタの問題をそれぞれ再考し、両者の相互関係を取入れ2次元生理学的モデルを構築することにより人間のメカニズムに基づいた音声合成方式を試みる。さらに、本方式の実用化へのボトルネックとなる音質の問題を解決するため、3次元動画計測手法を用い声道断面積を正確に計測し声道形状モデルを高精度化することで高品質化を目指す。これらを行うため、本論文では、まずソースにおける F0 調節の生理機構について分析し、咽頭筋の活動以外の新たな F0 調節機構について提案する。つぎに、F0 調節と声道形状制御を1つのモデルに統合した発話器官の2次元モデルを構築し、提案した F0 調節機構により引き起こされる F0 調節と声道形状制御との相互作用の存在を明らかにする。最後に、この F0 調節機構を実装することが容易で、かつ高品質な音声の合成ができ、テキスト音声合成の実用化への応用に適した声道断面積モデルを用いた音声合成方式を提案する。

F0 下降の生理機構の解明では、複数の被験者が降下音階で持続発声を行った時の喉頭の正中矢状断面画像を MRI を用いて撮像し、喉頭周辺の器官の位置変化を調べることで F0 調節機構の分析を行った。その結果、輪状軟骨の後板が頸椎の自然湾曲に沿って下降する結果、声帯を短くする回転が生ずる現象を確認した。この生理機構によりこれまで問題とされてきた F0 下降に伴う喉頭下降の現象や外喉頭筋の活動の理由を説明することができる。また、この F0 調節機構を考慮すると、F0 調節はその力が舌骨を介し舌形状を変化させ、反対に調音動作は咽頭の位置を変化させ F0 が変化するという、ソースとフィルタ間の相互作用が生じる。よって、本研究ではつぎに、先に示した F0 調節機構を考慮し、舌と喉頭を含む全ての器官の動作を同時に計算することで、この相互作用を実現する発話器官の2次元計算モデルを提案する。本モデルでは、各筋の活動量を入力として舌・喉頭・顎などの発話器官に加わる全ての力が釣り合うときの各器官の位置が計算され、その位置より得られた声道形状および声帯長を用いて音声の生成が行われる。合成実験により、先に示した F0 調節機能が F0 調節と舌形状制御の相互作用を引き起こすこと、本モデルでその相互作用が実現できることが明らかとなった。また本モデルでは、全てのパラメータを特定の1名の話者の生理学的データと MRI 画像を用いて同話者に最適化することで、個人性を有する合成音声の再現を目指した。しかし、個人性の再現には至らなかった。この結果より、この合成方式の実用化には音質がボトルネックになると考えられる。よって最後に、フィルタについて、これまで行ってきた2次元断面上での声道形状の模倣を3次元に拡張することで、本方式を用いても高品質な音声の合成が可能であることを明らかにする。具体的には近年開発された3次元 MRI 動画データと実音声を用いて声道形状を推定することで高精度化を行った。声道形状のモデルには声道を任意の幅の円錐台の連続体で近似する声道断面積モデルを用いた。ただし、将来的に先に解明した F0 調節機構などの音声生成の生理機構を組み込むことができるように、声道断面積モデルは物理的な声道の器官の位置との対応が容易にとれることを考慮し構築した。幾つかの単語の合成実験の結果、声道の3次元動画データを用いることで、人間の音声生成メカニズムに基づく音声合成方式を用いても、高品質な音声の合成が可能であることが確認された。今後はこの声道断面積モデルに、先に示した F0 調節の生理機構を実装し、また、任意の音声を合成するために必要な音素を含む幾つかの単語の分析実験を追加することで、コンパクトで自然な音声合成できるテキスト音声合成システムを開発する予定である。