

博士（工 学） 横 山 徹

学 位 論 文 題 名

音声生成過程における 3 次元声道形状の  
モデル化に関する研究

学位論文内容の要旨

今日の高度情報化社会において、機械と人間の間の知的コミュニケーションの必要性が急速に高まっている中、音声は極めて自然かつ効率的なコミュニケーション手段として重要である。近年のデジタル信号処理技術の発達やコンピュータの性能の向上に伴い、音声合成や音声認識、音声符号化など、音声情報の工学的処理技術は既に実用化されており、多種多様なサービスが始まりつつある。しかし音声を媒体としたサービスの品質が要求されるにつれ、これまで明らかにされていない、音声が持つ非言語的情報、すなわち感情や個人性、あるいは自然性などがどのような過程で生じるか、についての本質的な理解が求められて来ている。

人間が音声を生成する際には、脳から声門や声道(舌、顎、口唇などで構成される)などの発声器官に運動神経指令が出され、各器官に対応する種々の筋肉が動いて空気振動としての音声が外部へ放射される。長い研究の歴史で積み重ねられた音声生成過程に関する知識を基礎として、近年の音声情報処理技術は発展してきているのだが、音声生成系で生じる物理現象の中には未だ解明されていない複雑なものが多い。例えば声門や声道内で生じる乱流は子音の生成のための音源となる重要な現象であるが、その特性はわかっておらず大きな問題となっている。従って音声生成過程の複雑な現象について検討し、モデル化することによって、合成音の品質向上や高能率符号化への寄与が可能であると考えられる。

人間の発声器官のうち、声道の 3 次元的形状は音声生成メカニズムの中でも最も重要な要素である。例えば声道内の狭めによる乱流や、音波の高次伝搬モード、声道の壁面からのエネルギー損失などを解析するためには、声道の 3 次元的な形状情報の取得が不可欠である。しかし現状において、声道の 3 次元形状を考慮した音声生成過程のモデル化をするためには、(1) 調音状態の声道の 3 次元形状情報を計測、抽出する手法が確立していないこと、(2) 声道の 3 次元的形状を 1 次元モデルに変換する手法についての検討が成されていないことが問題として挙げられる。

本論文は上記の問題点を克服すべく、人間の音声生成過程における物理的な現象についての考察と、それを適切に模擬した音声生成モデルの構築を目的として、断層画像からの声道形状データの構成法を提案し、得られた 3 次元声道形状のモデル化が声道内の音波伝搬特性に及ぼす影響を明らかにしたものである。本論文は全 7 章からなり、以下のように要約される。

第 1 章では、本論文の背景と目的が述べられている。音声生成過程レベルに立ち帰った

本研究の位置付けを示し、声道の3次元声道形状情報を考慮した音声生成系のモデル化という本論文の目的とそこに至る背景について述べ、各章の要旨を示した。

第2章では、本論文に関わる人間の音声生成メカニズムについて述べ、本研究の対象である調音器官の機能を示した。次に調音の方法という観点から、音声を母音と子音とに大別して、これらの生成の仕組みについて説明した。調音器官として重要な「声道」と「口唇」のモデル化について説明した。声道内を均一音響管の縦続接続モデルで、口唇を球形バッフルを持つ円形ピストン振動板によって近似するという本論文の位置づけを行い、これらのモデルによる音響現象の解析のための理論的説明を行った。

第3章では、MRIで得られた声道の矢状断面データをもとにその3次元形状を多面体メッシングによって構成する方法を提案している。この方法では各段階において全て自動計算によって処理するのではなく、視察によって判断するような対話処理を併用した。このことによって分岐を含むような複雑な声道形状であっても、尤もらしい形状を再構成することができることを示した。しかしこの手法では、マニュアル処理を含むことにより同一データに対して再現性が悪化する場合があること、処理時間が長いという問題を残している。

第4章では、第3章で得られた3次元声道形状から、重要な形状パラメータとして、正中断面形状、音波伝搬経路、声道断面積関数、および声道周長関数を推定する方法を提案した。そしてこれらのパラメータによって電気的等価回路の各素子値を決定し、得られる分布定数回路の伝達特性を計算した。その結果、従来報告されている伝達特性とほぼ近い特性が得られることにより、再構成した3次元声道形状の妥当性を確認した。またモデルの分岐形状のパラメータを変化させて、分岐が伝達特性に及ぼす影響を調べた。その結果、分岐の長さに対する感度が高く、その長さの正確な抽出が必要であることを示した。これらのことによって本論文で提案した3次元声道形状モデルの構築の有用性を示した。

第5章では、第3章で説明した処理効率の悪さを改善し、大量の断層画像を効率良く処理することを目的として、デジタル画像を用いて3次元声道形状を再構成するためのアルゴリズムを提案した。また本アルゴリズムを実現するために開発したGUIによる対話システムの概要を示した。データ処理を大幅に自動化することで効率の悪さを改善しており、得られた3次元声道形状を用いて推定された声道伝達特性は、一般的に知られている特性とほぼ同様な結果が得られた。これにより第5章で提案した3次元声道形状の再構成法の有効性を示した。

第6章では、3次元声道形状を1次元モデルで近似する際の不確定要素として、音波伝搬経路の推定、音響管のセクション数、口唇の突き出し形状が声道伝達特性に及ぼす影響について検討した。局所的に音波伝搬経路を変更する実験を行った結果、音波伝搬経路の推定に対する声道伝達特性の感度は、経路が変動した部位に応じて程度差はあるが、平均的には小さいことを明らかにしている。また実験により1次元声道モデルの音響管のセクション数は、30以上にすべきであることを明らかにしている。さらに口唇部の突き出しを考慮して1次元モデルに組み込むことによって口唇端での境界条件の重要を示している。

最後に本論文で、3次元声道形状データを構築するためのアルゴリズムを提案し、それを実現するシステムを開発したこと、そして声道形状を近似する際のモデルの構造やモデル化の音響的影響など、高度な音声生成モデルの構築のための基礎的なデータを得たことが主要な成果である。

# 学位論文審査の要旨

主　査　教　授　小　川　吉　彦

副　査　教　授　宮　永　喜　一

副　査　教　授　永　井　信　夫

副　査　助教授　三　木　信　弘

## 学　位　論　文　題　名

### 音声生成過程における3次元声道形状の モデル化に関する研究

本論文は、断層画像から声道形状データを構築する手法を提案し、それを適切にモデル化して得られた3次元形状構造が、声道内の音波伝播特性におよぼす影響を明らかにした成果をまとめたものである。

人間の音声生成過程に関する研究は長年にわたり、これに基づいて近年の音声情報処理技術は発展してきた。しかし、音声生成系で生じる物理現象の中には、未だ解明されていない複雑な部分も多い。これらには、複雑性のために未だに解析ができないものが多い。そこで、これらの複雑系を適切なモデル化によって解析可能にすれば、求められる合成音の品質向上や高能率符号化へ大きな寄与をすることができる。本論文では、このような複雑系の一つに声道の3次元形状があり、しかもこれが声道での音声生成メカニズムの中でもっとも重要な要素の一つであることに着目した。そして、調音状態での声道の3次元形状情報を計測して、その構造を決定した。さらに、それを適切にモデル化して音声伝播特性に与える影響を解明したものである。

第1章では、本論文の背景と目的を述べた。

第2章では、人間の音声生成のメカニズムを述べ、調音器官の機能を明らかにした。特に、母音と子音に大別して生成の仕組みを述べ、調音器官のモデル化に重要な部分の位置付けを行った。

第3章では、磁気共鳴映像法(MRI)で得られた声道の矢状断面データから、声道の3次元形状を多面体メッシングで構築する手法を提案した。ただし、この手法では、全自动解析を行う代わりに、一部で視察判断を行う対話型処理を採用した。このため

に生じる再現性の劣化や処理時間の長さなどの問題は、第 5 章で改善した。

第 4 章では、第 3 章で得られた 3 次元声道形状から、正中断面形状、音波伝播経路、声道断面積関数および声道周長関数の音波伝播特性解析に重要な形状パラメータを推定する手法を提案した。また、これらのパラメータから決定した分布定数回路による伝達特性を計算し、従来報告されている伝達特性にほぼ近い特性が得られたことで、構築した 3 次元声道形状の妥当性を確認した。さらにモデルの分岐形状のパラメータを可変にし、分岐の長さに対する感度が高く、その長さの正確な決定が必要なことから提案した 3 次元形状モデルの必要性を強調した。

第 5 章では、第 3 章の処理効率の悪さを改善して、大量の断層画像の高速処理を目的とした 3 次元声道形状構築のアルゴリズムを提案した。また、本アルゴリズムを実現するために GUI による対話システムを開発した。得られた 3 次元形状をモデル化して解析した声道伝達特性は、一般的に知られる特性と比べて妥当な値を示した。

第 6 章は、3 次元声道形状を 1 次元モデルで近似するときに現れる不確定要素としての、音波伝播経路の推定、音響管のセクション数、口唇の突き出し形状が声道伝達特性におよぼす影響について考察した。この結果、音波伝播経路については多少のずれがあつても問題ないことが判明した。セクション数については 30 以上にすべきと結論した。さらに口唇端での境界条件の重要性を示した。

第 8 章では、本論文の結論を述べ、論文全体の成果を要約した。

これを要するに、著者は断層画像から 3 次元声道形状を構築するためのアルゴリズムを提案してその実現システムを開発し、それを適切にモデル化することで声道形状が与える音響的影響を明らかにして、高性能な音声生成モデルを構築するための基礎的データを得たことを述べたものであり、画像処理工学および音声情報処理工学の分野に寄与するところ大である。

よって著者は、北海道大学博士（工学）の学位を授与される資格あるものと認める。