

学位論文題名

有声音声発声時における声帯音源の特性に関する研究

学位論文内容の要旨

音声のうち、声帯振動が音源となっているものを有声音声という。有声音声において、生成された音声の特性を決定するものは、声帯の振動によって生じる有声音源の特性と、声帯から発生した音響エネルギーの伝搬経路である声道の特性の、両方であると考えられる。声道の特性によって音声の共振周波数(フォルマント)が決定し、フォルマント分布によって音の種類が決定されるため、音の種類を表す情報(言語情報と呼ぶことにする)は、声道の特性が担っていると考えられる。それに対して、音源の特性は、音声の個人性や感情などを表す情報(非言語情報と呼ぶことにする)を担っていると考えられる。

現在実用的に用いられている技術においては、言語情報を担う声道の特性が、非言語情報を担う音源の特性に優先して取り扱われる傾向がある。これは、声道の特性がフォルマント周波数という、比較的観測しやすいパラメータで取り扱うことができる一方、声帯の特性は、直接観測することが難しく、また有効なパラメータについての知見が十分に得られていないためと考えられる。このような状況において、有声音源の特性、あるいは声帯振動の特性について検討することは有用であると考えられる。声帯特性は直接観測することが難しいため、音声分析法を用いて声帯特性を推定する方法や、音声生成器官をモデル化した合成器を用いた検討が重要となる。

本論文は、以上のような背景から、人間の音声生成過程のうち、音源である声帯の特性について、音声分析、音声合成の両面から検討をおこなったものである。本論文は全 12 章からなり、以下のように要約される。

第 1 章においては、本論文の背景と目的が述べられている。本論文において注目している声帯特性の重要性を指摘し、実音声から声帯特性を推定する分析的な方法と、合成モデルを使って検討する方法の両方を用いることが、声帯特性についての検討に有効であることを述べるとともに、各章の要旨を示すことによって研究の方向性を示す。

第 2 章においては、音声生成器官の基本的な構造が述べられている。また、音声生成器官を近似するためのモデルとして、伝達関数モデルと音響管モデルを示し、特徴について述べる。

第 3 章においては、第 2 章において述べた伝達関数モデルと関係が深い、線形予測分析について述べる。また、少ないサンプル数の信号においても、線形予測分析の精度を確保するために有効である Fejér kernel について、その定義と適用方法を述べる。

第 4 章においては、第 2 章において述べた音響管モデルに基づいて、実際に声道アナログ音声合成器を構成するための基本的な方法について述べる。声門部、声道部、放射部の各部分が、物理的モデル、電気回路モデル、デジタルフィルタモデルの順にモデル化される様子を、順を追って解説する。

第 5 章においては、声門閉鎖情報を利用した声帯波形推定法について述べる。第 3 章で述べた方法を元に、音声の声道伝達関数をピッチごとに求めて逆フィルタリングを行う方法を提案する。実音声に対して提案法を適用し、良好な声帯波形推定が行われることを示すとともに、

分析次数や分析時における正規方程式の構成法が声帯波形推定に与える影響について検討する。

第 6 章においては、第 5 章において推定された声帯波形から、声帯波形モデルに基づいたモデルパラメータを推定する方法について述べる。また、ピッチごとにモデルパラメータを推定し、その時間変化について考察する。実験結果から、妥当と思われるモデルパラメータが推定できることを示す。

第 7 章においては、第 5 章において用いられた、EGG を用いた声門閉鎖区間の推定精度について検討を行う。また、推定された声帯波形から、周期的なピーク点を抽出して声門閉鎖区間の始点とした上で、再度分析を行う方法について、精度の検討を行う。以上から、EGG を用いた声門閉鎖区間の推定精度が悪化する場合があること、提案する閉鎖区間補正法が、閉鎖区間の推定精度を改善するのに有効であることを示す。

第 8 章においては、音源として声帯波形モデルを用いた低ビットレートボコーダについて述べる。符号化器の音源部分に声帯波形モデルを導入することによって、線形予測分析の残差波形をより精度良く近似出来ることを示す。また、主観評価による従来法との比較を示す。

第 9 章においては、高次統計量を用いた分析方法について述べるとともに、音声波形に適用する検討を行う。実音声に対して安定した分析を行うために、MA 係数を推定するために用いられる GM 方程式を、安定して解くための方法を提案する。実験によって、提案法が AR 分析では補正できないスペクトル傾きを推定できることを示す。最後に、高次統計量を用いた分析法が持つ不安定性に関する問題を提起する。

第 10 章においては、有声母音の発声時に、声門下の特性が音声の特性に与える影響についての検討を、声道アナログ音声合成器を用いて行う。声門下のモデルを組み込んだ声道アナログ音声合成器を用いて合成された音声を分析することによって、声門下の特性が音声の周波数特性に与える影響について定性的な議論を行う。また、合成器の時間応答波形から、ある時間時点における周波数特性を精度良く推定する方法を導入して、より精度の高い検討を行う。以上の検討から、声門下と声道が直結される声門開放区間において、特に 2 kHz 以下の周波数で、音声の周波数特性が声門下インピーダンスの影響を受けることを示す。

第 11 章においては、フォルマント周波数の移動に着目して、声門開閉によるフォルマントの変化について検討を行い、その原因を明らかにする。単純化した音声生成モデルを用いて、声門の開閉によるフォルマントの変化や、声門下インピーダンスの音響質量成分と音響コンプライアンス成分が、周波数特性に与える影響について考察を行うとともに、実験によって考察の妥当性を示す。また、単純化した音声生成モデルを用いた検討から得られた知見を用いて、実際の声門下インピーダンスの影響について考察する。以上から、声門における境界条件の変化によって、声門開放区間のフォルマント周波数が、声門閉鎖区間のそれより上昇することを示す。また、声門下インピーダンスの虚部の影響によって、声門開放区間のフォルマント周波数が、声門下を考慮しない場合と比較して高域側および低域側の両方に移動し得ることを示す。

第 12 章では結論として、各章で得られた成果を列挙するとともに、今後の研究の方向性などについて述べる。

本論文の主要な成果は、音声の声門閉鎖区間を選択的に分析する方法と声道伝達関数の時変性を考慮した逆フィルタ法によって、音声波形から声帯波形を高精度に推定する方法を提案し、推定された声帯波形の性質について一定の知見を得たこと、声帯波形の推定精度を改善させるために、高次統計量を用いた音声分析方法について検討を行い、推定のロバスト性を改善する方法を提案するとともに、なお残る安定性に関する問題点を明らかにしたこと、および声道アナログ音声合成器を用いた検討によって、声門下の特性や声門の開閉がフォルマントに与える影響を明確にしたことである。

学位論文審査の要旨

主 査 教 授 山 本 強

副 査 教 授 永 井 信 夫

副 査 教 授 北 島 秀 夫

副 査 教 授 三 木 信 弘 (公立はこだて未来大学)

学 位 論 文 題 名

有声音声発声時における声帯音源の特性に関する研究

本論文は、人間の音声生成過程のうち、音源である声帯の特性について、音声分析、音声合成の両面から検討を行ったものである。

人間の音声生成過程についての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。しかしながら、音声情報処理においては、言語情報を担う声道の特性が、非言語情報を担う音源の特性に優先して取り扱われる傾向がある。これは、声道の特性がフォルマント周波数という、比較的観測しやすいパラメータで取り扱うことができる一方、声帯の特性は、直接観測することが難しく、また有効なパラメータについての知見が十分に得られていないためと考えられる。このような背景から、本論文では、音声生成器官のうち、有声音源となる声帯の特性に着目して、音声分析と逆フィルタ法を用いた声帯特性推定や、声道アナログ音声合成器を用いた検討を行った。

第1章においては、本論文の背景と目的を述べた。

第2章においては、音声生成器官の基本的な構造を述べるとともに、音声生成器官を近似するためのモデルとして、音声分析を行う際に用いる伝達関数モデルと、声道アナログ音声合成器を構築する際に用いる音響管モデルを示した。

第3章においては、第5章において用いられる、線形予測分析と Fejér kernel について、基礎的な説明を行った。

第4章においては、第10章および第11章で用いられる声道アナログ音声合成器について、その構成法を述べた。声門部、声道部、放射部の各部分が、物理的モデル、電気回路モデル、デジタルフィルタモデルの順にモデル化される様子を、順を追って解説した。

第5章においては、第3章で述べた分析方法を用いた、声門閉鎖情報を利用した声帯波形推定法について述べた。実音声に対して提案法を適用することによって、良好な声帯波形推定が行われることを示した。また、提案法を適用する際の分析次数や、分析時における正規方程式の構成法が声帯波形推定に与える影響について検討を行った。

第6章においては、第5章において推定された声帯波形から、声帯波形モデルに基づい

たモデルパラメータを推定する方法について述べた。実際にピッチごとのモデルパラメータを推定し、声帯波形パラメータの時間変化を示した。

第7章においては、第5章において用いられた、EGGを用いた声門閉鎖区間の推定精度について検討を行った。また、EGGを用いた声門閉鎖区間の推定精度を改善する方法を提案し、実験によって精度が改善されることを示した。

第8章においては、音源として声帯波形モデルを用いた低ビットレートボコーダについて述べた。符号化器の音源部分に声帯波形モデルを導入することによって、線形予測分析の残差波形をより精度良く近似出来ることを示した。

第9章においては、高次統計量を用いた分析方法について述べるとともに、音声波形に適用する検討を行った。最初に、高次統計量についての基本的な性質について述べた。次に、疑似音声信号を用いて、高次統計量を用いた分析方法が持つこれらの性質について確認を行った。実音声に対して安定した分析を行うために、MA係数を推定するために用いられるGM方程式を、安定して解くための方法を提案した。提案法によって、分析のロバスト性が向上することを確認した。提案した方法を実音声の子音部に適用して実験を行い、AR分析のみでは補正できなかったスペクトル傾きを、提案法によって推定できることを示した。

第10章においては、有声母音の発声時に、声門下の特性が音声の特性に与える影響についての検討を、声道アナログ音声合成器を用いて行った。声門下のモデルを組み込んだ音声合成気を用いて合成された音声を分析することによって、声門下の特性が音声の周波数特性に与える影響について定性的な議論を行った。また、合成器の時間応答波形から、ある時間時点における周波数特性を精度良く推定する方法を導入して、より精度の高い検討を行った。以上の検討から、声門開放区間において、特に2 kHz以下の周波数で、音声の周波数特性が声門下インピーダンスの影響を受けることを示した。

第11章においては、フォルマント周波数の移動に着目して、声門開閉によるフォルマントの変化について検討を行い、その原因を明らかにした。単純化した音声生成モデルを用いて、声門の開閉によるフォルマントの変化や、声門下インピーダンスの音響質量成分と音響コンプライアンス成分が、周波数特性に与える影響について考察を行うとともに、実験によって妥当性を示した。また、単純化した音声生成モデルを用いた検討から得られた知見を用いて、実際の声門下インピーダンスの影響について考察する。以上から本章では、声門における境界条件の変化によって、声門開放区間のフォルマント周波数が、声門閉鎖区間のそれより上昇することを示した。また、声門下インピーダンスの虚部の影響によって、声門開放区間のフォルマント周波数が、声門下を考慮しない場合と比較して高域側および低域側の両方に移動し得ることを示した。

第12章では結論として、各章で得られた成果を列挙するとともに、今後の研究の方向性などについて述べた。

これを要するに、著者は、音声の声門閉鎖区間を選択的に分析する方法と声道伝達関数の時変性を考慮した逆フィルタ法によって、音声波形から声帯波形を高精度に推定する方法を提案し、推定された声帯波形の性質について一定の知見を得た。また、声帯波形の推

定精度を改善させるために、高次統計量を用いた音声分析方法について検討を行い、推定のロバスト性を改善する方法を提案するとともに、なお残る安定性に関する問題点を明らかにした。さらに、声道アナログ音声合成器を用いた検討によって、声門下の特性が音声の特性に与える影響を明確にした。そのため、本論文は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。