

令和 3 年 6 月 21 日現在

機関番号：33919

研究種目：基盤研究(C) (一般)

研究期間：2017～2020

課題番号：17K00253

研究課題名(和文)3次元声道形状と声帯音源の高精度抽出が可能な高品質音声分析変換合成方式の開発

研究課題名(英文)Development of high-quality speech analysis-synthesis systems with ability to extract 3D vocal tract shape and vocal cord vibration signal precisely

研究代表者

坂野 秀樹 (Banno, Hideki)

名城大学・理工学部・准教授

研究者番号：20335003

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：音声信号から声道形状に相当するものを求める方法として、PARCOR係数から声道断面積関数を算出する方法がある。しかし、この声道断面積関数では、本来とは異なった形状が得られることがある上、複雑な形状は表現されない。これを踏まえ、本研究では音声信号から高精度に3次元声道形状を推定する方法について検討してきた。PARCOR係数から算出されるものは1次元の形状データに相当するため、まず3Dプリンタによってその声道模型を作成した上で、この声道模型を用いて音響特性を測定した結果と、形状データから音響特性を推定するFDTD法などによる結果との比較・分析を行った。そしてこの結果に基づき、提案法の改良を行った。

研究成果の学術的意義や社会的意義

音声信号から発声器官のパラメータを推定する研究や口唇の画像を生成する研究、3次元声道形状から音声を合成する研究は存在しているが、音声信号から3次元声道形状を推定し、さらにそれを用いて高品質に音声を合成する研究は世界的にも類を見ず、極めて独創的な研究である。今回の研究では、詳細な3次元声道形状を推定する部分の実現はできなかったが、今後、言語教育における発声の可視化への応用や、声質変換などの応用における新しい音声補間の方法の開発にもつながるなど、極めて意義深い研究であると考えている。

研究成果の概要(英文)：One of the methods to estimate vocal tract shape information from speech signal is the PARCOR analysis-based method which converts the PARCOR coefficients of speech signal into the vocal tract area function. However, the estimated vocal tract area function sometimes is incorrect and does not always represent complicated shape. Accordingly, we started the study on the method estimating 3-D vocal tract shape precisely from speech signal. Firstly, physical 1-D vocal tract models which correspond to the estimated vocal tract area function were created by 3-D printing, then the acoustic characteristics of the models were measured. Secondly, the characteristics were compared with simulation results generated by acoustic simulation methods such as the FDTD method which can generate simulated acoustic characteristics from shape information. Lastly, based on these comparisons, we improved our method.

研究分野：音情報処理

キーワード：3次元声道形状 声道断面積関数 PARCOR分析 フォルマント FDTD法

## 1. 研究開始当初の背景

音声は、声帯の振動で発生する**声帯音源**が**声道**を通った後、口唇で放射されて生成される。この声道は、3次元的に複雑な形状を持ち、その形状によって話者性の大部分が決定付けられるほか、舌や下顎の運動によって形状が変化し、その形状に応じた音素の生成をさせるものとなっている。工学的には、声帯音源信号が、声道及び口唇の放射による音色変化を実現する線形の**声道フィルタ**を通り、音声になるという**ソースフィルタモデル**に基づき、声道フィルタに相当する情報を抽出する処理が通常行われる。声道フィルタ情報を抽出する手法としては、線形予測係数を抽出する線形予測分析、線形予測係数との相互変換が可能な **PARCOR 係数**を抽出する PARCOR 分析などがある。これらの手法は声道形状そのものの抽出を目的とするものではないが、PARCOR 係数は、声道形状の情報をある程度近似する**声道断面積関数**に変換ができることが知られている。物理的な声道断面積関数は、3次元声道形状を等間隔に輪切りにした際の断面積の系列である。これに対し、PARCOR 係数によって求められる声道断面積関数は、断面積のみ異なる円筒状の音響管を縦続接続して声道形状をモデル化した**ケリーの声道モデル**に基づくもので、この音響管の断面積の系列に対応している。ただし、この声道断面積関数を算出する際には、声帯音源と口唇の放射の特性を予め取り除いておく必要があり、その方法として、中島らの提案した**適応逆フィルタ法**<sup>[1]</sup>などが知られている。我々は、この適応逆フィルタ法を高いサンプリング周波数でも動作するように改良した上で、高品質音声分析変換合成法である **STRAIGHT** と組み合わせ、高品質な声質制御が可能であることを示した<sup>[2]</sup>。また、英語教育などへの応用を目的とし、入力音声信号から、例えば英語音素の /r/ と /l/ の違いを可視化する試みにおいても一定の成果が得られていた。

## 2. 研究の目的

我々は、音声信号から声道形状に相当するものを求める方法の一つである、PARCOR 係数から声道断面積関数を算出する方法を改良し、高品質音声分析合成システムの構築を進めている。しかしながら、この声道断面積関数では、本来の声道形状とは異なったものが得られることがある上、声道内に分岐が存在する場合など、そもそも複雑な声道形状の情報は表現されない。そこで、本研究では、音声信号から高精度に3次元声道形状を推定する方法について検討する。また、入力音声と推定される3次元声道形状の情報を利用し、声帯音源についても高精度に推定する。そして、これらの推定された情報に対して必要に応じて変形を加えた後、高品質に音声を合成することが可能な手法の開発を行う。

## 3. 研究の方法

目的を達成すべく、主に、(1)PARCOR 係数による声道断面積関数の精度向上、(2)3D プリントによる声道模型を用いた検証、(3)音声信号からの3次元声道形状及び声帯音源の推定と音声の合成の3点について重点的に取り組んだ。

## 4. 研究成果

### (1) PARCOR 係数による声道断面積関数の精度向上

最初に、PARCOR 係数による声道断面積関数の精度向上に取り組んだ。精度を評価するためには、まず、正解として利用する声道断面積関数が必要であり、以下の2種類を用意した。(a)3次元声道形状データから直接算出される声道断面積関数。これは3次元声道形状の本来の声道断面積関数に相当するため、物理的な形状としての正解と言って良い。ここでは、すでに研究室で作成済みであった3次元声道形状データを可視化するツールを改良し、中心線を自動算出するようにした上で、必要に応じて修正を行えるよう改良を行った。(b)3次元声道形状データを用いた上で、形状データから伝達特定をシミュレートできる FDTD 法によって算出された伝達特性から PARCOR 係数を求めた後、計算される声道断面積関数。音声の合成を行うことを考えた場合には、この声道断面積関数が最良ということになる。(a)と(b)の形状の比較から、ケリーの声道モデルによる声道断面積関数がどの程度本来の声道断面積関数に近いのかを評価した。その結果、ある程度の形状の一致はあるものの、異なっている部分も多いことが明らかとなった。

次に、正解の声帯音源のデータとなる様々な周波数特性を持つ人工的な声帯音源信号を作成した。ここでは、声帯音源信号としてインパルス列、三角波、鋸歯状波などを用いた。FDTD 法で算出した声道フィルタに対し、この声帯音源信号を入力することで音声合成できるため、この合成音に対し、適応逆フィルタ法を用いて、PARCOR 係数による声道断面積関数と残差信号を算出する実験を行った。そして、得られる声道断面積関数が正解にどの程度近いのか、残差信号のスペクトルがどの程度入力した声帯音源信号のスペクトルに近いかを評価した。その結果、適応逆フィルタ法による方法である程度の抽出は可能であるものの、精度は十分であるとは言えないことが明らかとなった。この原因の一つは、ケリーの声道モデルによる声道断面積関数が仮定する1次元的な音波の伝播が、実際の3次元的な音波の伝播と異なっていることである可能性があったため、3D プリントによって1次元声道形状に基づく声道模型を造形して検証を進める

こととした。

### (2) 3D プリンタによる声道模型を用いた検証

PARCOR 係数による声道断面積関数の精度向上と並行して、3D プリンタを用いて声道模型を作成し、その音響特性を調べることによる検証も行った。

最初に、3D プリンタで 1 次元の声道模型を造形した上でその音響特性を測定し、この誤差を小さくすべく検討を進めた。まず、音響特性の測定において、測定方法による音響特性の違いについて検証を行った。ここでは、声帯側から音源を入力した場合と、口唇側から音源を入力した場合の比較を行い、口唇側から音源を入力した場合の方が安定して測定できることが明らかとなった。次に、3D プリンタ造形時の精度に相当する積層間隔を変えて声道模型を造形した結果、積層間隔による音響特性の違いはわずかであり、第 4 フォルマントに相当するスペクトルのピークなど、2kHz 以上の高域に違いが出るのみであることが分かった。一方、積層間隔に相当する空間離散間隔が異なる声道形状の情報から音響特性を FDTD 法によって模擬したところ、同様に 2kHz 以上の高域に違いが見られたが、特に第 4 フォルマントに相当するスペクトルのピークに大きな違いが見られ、3D プリンタによる声道模型の場合と異なることが分かった。また、FDTD 法による音響特性が声道模型の音響特性とどのように異なるか調査したところ、第 4 フォルマント以降のピークに大きく違いが見られることが分かった。この問題について検証を進めたところ、その多くは 1 次元 FDTD 法によるシミュレーションでは 3 次元情報の表現が不十分であったことに起因するものであり、3 次元 FDTD 法を用いることによりかなり解消されることが確認された。さらに、3D プリンタによる声道模型は、造形時に膨張して内径が 0.5mm 程度小さくなっており、これによる影響で第 2 フォルマントにおいてもずれが見られることが分かった。

また、音声信号から PARCOR 分析を介して声道形状を推定する手法の改良方法をいくつか検討したが、精度の改善につながらないものがほとんどであったため、その原因を調査する実験も行った。ここでは、声道形状が直角に変化する部分も含む場合とそうでない場合について検証を行っている。両者の伝達関数におけるフォルマント周波数の比較により、声道形状が直角に変化する部分を含む場合には、1 次元のシミュレーション手法で第 4・第 5 フォルマント周波数で特に大きな誤差が生じることが明らかとなった。一方で、声道形状が滑らかに変化する場合には、誤差が手法によらず比較的小さいことが分かった。音圧分布シミュレーションの結果より、前者の声道形状において直角に変化する部分で発生する球面波の影響であることが示唆された。実際の人間の声道形状では直角に変化することはほとんどないため、PARCOR 分析を用いた声道形状推定において精度向上が見られなかった原因の一つは、評価に用いていた声道形状が、直角に変化する部分を含むことであった可能性が高い。これを踏まえ、再実験を行ったところ、推定精度の向上が確認された。また、フォルマント周波数の誤差については、1 次元のシミュレーション手法では、高次のフォルマント周波数が高い周波数に推定される方向に誤差が発生することが判明した。高い周波数への誤差は、周波数軸の伸長で近似できるため、伝達関数の周波数軸を伸長する処理を行うことによって推定精度を改善することができる可能性がある。この部分の研究については、当初の予定よりも大きな成果が見られたと考えており、現在、原著論文を投稿中の段階である。

### (3) 音声信号からの 3 次元声道形状及び声帯音源の推定と音声の合成

音声信号から 3 次元声道形状を抽出して可視化するとともに、声帯音源情報を抽出して 3 次元声道形状情報と声帯音源情報から音声を合成するシステムを構築した。現状では、一連の処理がリアルタイム動作するものになっているものの、当初の目標であった高精度な 3 次元声道形状の抽出はできておらず、さらなる検討を続けていく必要がある。

### 参考文献

- [1] デコンボリューションによる声道形の推定と適応型音声分析システム, 中島 他, 日本音響学会誌, Vol.34, No.3, pp.157-166, 1978.
- [2] High quality voice manipulation method based on the vocal tract area function obtained from sub-band LSP of straight spectrum, Arakawa et al., Proc. ICASSP, pp.4834-4837, 2010.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 後藤 康泰, 坂野 秀樹, 旭 健作
2. 発表標題 3Dプリンタによる声道モデルの音響特性の計測実験に対するFDTD法を用いた検証
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 後藤 康泰, 坂野 秀樹, 旭 健作
2. 発表標題 3Dプリンタによる声道モデルの音響特性の計測実験に対する1次元FDTD法を用いた評価
3. 学会等名 令和元年度 電気・電子・情報関係学会 東海支部連合大会
4. 発表年 2019年

1. 発表者名 後藤康泰, 坂野秀樹
2. 発表標題 声道モデルを3Dプリンタで作成する際の精度と音響特性の評価
3. 学会等名 平成30年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2018年

1. 発表者名 後藤康泰, 坂野秀樹
2. 発表標題 3Dプリンタを用いて作成した声道モデルの精度と音響特性の評価
3. 学会等名 日本音響学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 奥野真基, 坂野秀樹
2. 発表標題 FDTD法を用いた声道モデルの音響解析におけるGPGPUを用いた最適化の検討
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 後藤 康泰, 坂野 秀樹
2. 発表標題 FDTD 法による声道伝達関数と3D プリンタによる声道モデルの音響特性との比較
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------