様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 ２９ 日現在

機関番号：８２６２６
研究種目：基盤研究(C)（一般）
研究期間：2017～2019
課題番号：１７Ｋ００２５８

研究課題名（和文）喉頭全摘出者の代替発声を対象とした声質改善装置の研究開発

研究課題名（英文）Research and development of a voice quality enhancing method for alternative vocalizations by a laryngectomized person

研究代表者
佐宗　晃（Sasou, Akira）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・研究グループ長

研究者番号：５０３１８１６９

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）：本研究課題では、声帯振動による周期的な音源で発声する健常者音声だけでなく、食道入り口部の振動による非定常・非周期的な音源で発声する食道発声音声のような特殊な音声を含む、より広範囲な音声を対象とする音声分析法の実現を目的として、音源を表すHMMの最適なトポロジをボトムアップに自動生成すると共に、そのモデルパラメータも同時推定し、声道特性と音源をより高精度に分離して抽出できる新しい音声分析法を構築した。そして、食道発声音声の声質劣化要因である音源を、喉頭全摘出前に収録した音声から推定した声帯音源などに入れ替えて再合成することで、もとの声質に近い発声を可能にする声質改善装置の実現可能性を検討した。

研究成果の学術的意義や社会的意義
音声は最も重要なコミュニケーション手段であり、高齢者のみならず人が充実した社会生活を送るために重要な要素である。しかし、喉頭がんの進行などにより喉頭全摘出手術を余儀なくされ、自分の声を失う高齢者は少なくない。本研究課題で実現を目指す声質改善装置は、不幸にして自分の声を失った人が、手術前の自分の声に近い発声を取り戻し、再度、自信をもって声による人とのコミュニケーションをとれるようにすることで、積極的な社会参加を促すことを目的とする。また、声帯音源で発声する健常者音声の分析手法は従来多く提案されているが、食道発声音声のような音声を対象としたものは決して多くない。

研究成果の概要（英文）：The main purpose of this research project is to realize a speech analysis method applicable to various types of speech including not only a normal voice uttered by a periodic excitation source due to vocal cord vibration but also special voice such as an esophageal voice uttered by a non-stationary, non-periodic excitation source due to vibration at the upper part of the esophagus. We constructed a new voice analysis method that can automatically generate the optimal topology of the HMM that represents the excitation source from the observed voice signal, and can also simultaneously estimate the model parameters. Based on the proposed speech analysis method, we then investigated the feasibility of voice quality enhancing method to replace the excitation source, which is the voice quality deterioration factor of the esophageal voice, with a vocal cord excitation source estimated from the normal voice recorded before total laryngectomy and resynthesize it.

研究分野：音声・音響信号処理、パターン認識

キーワード：音声分析　食道発声音声　声質改善　AR-HMM

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

声帯振動による音源が声道を通って空気中に放出される音声から、その重要な構成要素である声道特性と音源を分離して抽出する音声分析法は古くから研究されている。しかしながら音源は体外から直接観測できないため、Glottal Inverse Filtering（GIF）などの音声分析法やその精度評価法などについていまだに検討が続けられている。研究代表者は、声道フィルタを Auto-Regressive(AR)フィルタで、声帯振動による音源を Hidden Markov Model(HMM)で、表現した AR-HMM を音声の音響分析に適用し、従来の音声分析法では基本周波数の高い音声から声道特性の推定が困難となる問題を、大幅に改善できることを示した。声帯振動による音源波形は周期的となるため、AR-HMM の提案当初は、HMM の状態間を周回するように各状態を接続したリング状トポロジを仮定し、AR-HMM パラメータを推定していた。しかし、もしトポロジも観測音声に適合するように自動生成できれば、話者の個人性や感情、または声帯疾患を伴った病的音声や声帯振動とは全く異なる音源で発声する食道発声音声などの音響的特徴や声質に関する有用な情報が、より的確に抽出可能となるかもしれないという発想に至った。通常の発声における声帯振動の音源に関しては古くから研究され、いくつかの音源波形モデルが提案されており、それらに基づいた Analysis-by-Synthesis により声道モデルと音源波形モデルのパラメータを高精度に推定する分析法や、声帯振動による周期的駆動の特性を利用して滑らかな声道特性を推定する分析法など、従来、数多くの音声分析法が提案されている。一方、食道発声音声は、食道に飲み込んだ空気を逆流させ、食道入り口部を振動させることで声を出す発声法である。この音源は非定常的かつ非周期的であり、これまでに十分な研究の蓄積も無く、声帯音源のような波形モデルを構築するのは非常に難しい。このため、従来の音声分析法では食道発声音声のような特殊な音声から高精度にその声道と音源の特徴を抽出するのは困難であった。

音声は最も重要なコミュニケーション手段であり、高齢者のみならず人が充実した社会生活を送るために欠かせない要素である。しかし、喉頭がんの進行により喉頭全摘手術を余儀なくされ、自分の声を失う高齢者は少なくない。不幸にして喉頭全摘出となった場合、電気式喉頭や、ゲップを音源とする食道発声法などの代替発声法が利用されるが、習得が困難で、明瞭性や自然性が大幅に劣化するなどの問題点がある。これまでにも、ディジタル信号処理による基本周波数変換などの声質改善機能を備えた携帯型発声補助装置の実現の試みなどがあるが、途中で声質改善機能を持たないアナログ拡声装置の開発に見直されたという経緯をもつ。ディジタル信号処理方式の優位性が必ずしも示せなかった原因の１つは、当時採用していた音声分析法が健常者音声の生成モデルに基づいて構築されており、食道発声音声に対しては十分な分析精度が得られなかったためと考えられる。近年、ソースとターゲットの音声データを大量に用意し、その音響特徴量の結合確率密度関数を学習することで、声質変換を実現する手法などが開発されている。しかし、これを食道発声音声の声質改善に適用する場合、食道発声のサンプル音声を大量に用意する必要があり、声質改善装置を最も必要とする食道発声法初心者への負担が大きいという問題点がある。

２．研究の目的

本研究では、食道入り口部の振動による非定常・非周期的な音源で発声する食道発声音声のような特殊な音声に対して、音源を表す HMM の最適なトポロジをボトムアップに自動生成すると共に、そのモデルパラメータも同時推定し、声道特性と音源をより高精度に分離して抽出できる新しい音声分析法を確立する。そして、健常者音声から食道発声音声まで、より広範囲な音声を分析可能な音声分析法の実現を目的とする。また提案法のアプリケーションとして、食道発声音声を声道特性とその声質劣化要因である音源とに分離し、喉頭全摘出前に収録した音声から推定した声帯音源などに入れ替えて再合成することにより、もとの声質に近い発声を可能にする声質改善装置の研究開発を行う。

３．研究の方法

AR-HMM のパラメータとトポロジを観測音声からボトムアップに推定する分析法を図１に示すように構築する。音源 HMM の状態数が１と２のトポロジから開始して、それぞれのトポロジの AR-HMM パラメータを推定する。そして、確率統計モデルの評価指標に基づき、良いモデルを選択する。状態数が１のモデルが良いと判断した場合はそこで終了する。もし状態数が２のモデルの方が良い場合は、尤度に基づいてどちらか一方の状態を選択し、並列分割と直列分割の２種類の方法で状態を分割する。このようにして状態数が３のトポロジが２種類得られ、それぞれについて AR-HMM パラ

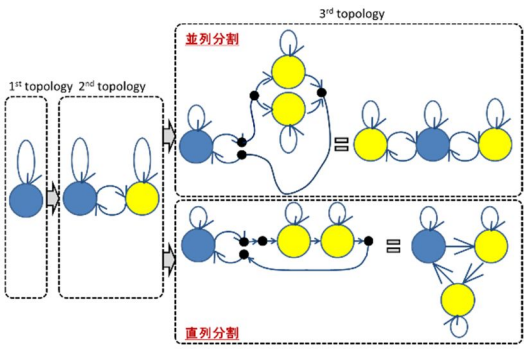

図１ トポロジ自動生成のための逐次的状態分割

メータを再推定する。そして評価指標に基づいて良いモデルを選択することで、状態数が３のトポロジが決定する。この時、状態数が２のモデルの方が良いと判断された場合は、これを最終モデルとして採用して終了する。もし状態数が３のモデルの方が良いと判断されたら、更に状態を分割して状態数が４のモデルを検討する。これを最適なモデルが選択されるまで繰り返す。そしてこの逐次状態分割による分析法を健常者音声および食道発声音声に適用し、その有効性を検証する。

上記の分析法は、音源 HMM のパラメータを繰り返し推定するため計算量が膨大となるため、リアルタイムで声質改善を行う装置のオンライン分析法として適切ではない。この問題に対しては図２に示す分析方法を検討する。予めオフラインで食道発声話者の少量音声サンプルから、その特定話者の音源 HMM のトポロジとパラメータを学習しておく。そしてリアルタイムでの声質改善処理では、音源 HMM のトポロジとパラメータを繰返し推定する分析法に代わり、時変 AR 係数と、学習音声データとオンラインでの入力音声データの振幅レベルの差を調整するための時変利得とを学習済み音源 HMM に基づき適応的に推定する分析法の使用を検討する。

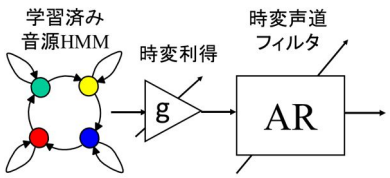

図2　事前学習HMMに基づくAR-HMM分析

逐次状態分割によるトポロジ推定 AR-HMM 分析により食道発声音声を声道特性と音源とに分離し、喉頭全摘出前に収録した自分の音声から抽出した声帯音源に入れ替えて、声質改善音声を再合成する方式を検討する。再合成音声の音韻性を決定する声道特性は、食道発声音声のオンライン AR-HMM 分析で得られる AR フィルタを使用する。食道発声音声から抽出した韻律情報は声質劣化の要因と考えられるため、そのまま声質改善音声の再合成には利用できない。一方、喉頭全摘出前に収録した自分の音声から抽出した声帯音源もそのまま音声の再合成に使用するのではなく、入力の発話内容や話者の意図に合うように韻律情報を修正する必要がある。特に食道発声の初心者などは、音源の基本周波数(F0)パターンを自身の意図するように制御することが困難である。このため F0 パターンは別途用意し、新たに生成する声帯音源に付与する必要がる。通常発話における F0 パターンについては、古くから研究の蓄積があり、その生成モデルも幾つか提案されている。本研究では声質改善音声における感情の表出を可能とするための基礎技術として、各感情における F0 パターンの生成モデルの構築を検討する。

４．研究成果

合成した疑似食道発声音声に逐次状態分割によるトポロジ推定 AR-HMM 音声分析法を適用して、声道特性の推定精度を評価した。疑似食道発声音声の合成に用いた声道フィルタとトポロジ推定 AR-HMM 分析により推定した声道フィルタとの間で LPC メルケプストラム距離を算出することで精度を評価した。また比較のためにトポロジをリング状で固定した AR-HMM 分析と従来法による分析も行った。その結果、トポロジ推定 AR-HMM 分析が最も誤差が少ないことを確認した。またトポロジ推定 AR-HMM 分析により、リング状のトポロジがどの程度の割合で生成されるかを、健常者音声と食道発声音声で比較した。その結果、健常者音声は 80.16%、食道発声音声は 62.45%でリング状トポロジが生成されていた。健常者音声は声帯振動による周期的な音源で発声しているため、食道発声音声に比べて高い割合でリング状トポロジが生成されるという期待通りの妥当な結果が得られた。

逐次状態分割によるトポロジ推定 AR-HMM 分析は、計算量が膨大となるため、リアルタイム処理には向いていない。予めオフラインで食道発声話者の少量音声サンプルから、特定話者の音源 HMM のトポロジとパラメータを学習し、リアルタイムでの声質改善処理（図３）では、事前学習の音源 HMM に基づいて時変利得と時変 AR 係数のみを適応的に推定する分析法を利用する。声道特性が時間的にほぼ不変である単母音の食道発声音声に対して、事前学習した音源 HMM に基づく AR-HMM 分析を行った。また比較のために従来法である Recursive Least Squares(RLS)による声道特性抽出も行った。その結果、提案法の AR-HMM 分析により得られた声道特性は、時間的に変動が少なく単母音音声のフォルマント構造を明確に抽出しているのに対し、従来法の RLS では食道入り口部の振動による音源の非定常性の影響を強く受け、得られた声道特性も時間的に大きく変動する傾向が見られた。次に、食道発声音声を声道特性と音源とに分離し、得られた音源を健常者音声から抽出した声帯音源に入れ替えて、声質改善音声を再合成する方式を検討し

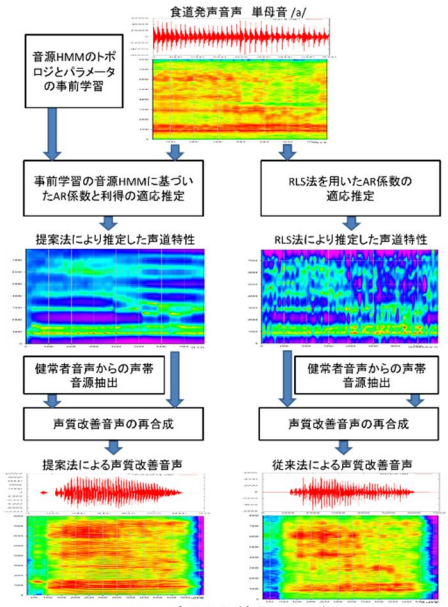

図3　声質改善処理の例

た。その結果、食道入り口部の振動による音源の非定常性の影響を強く受けた RLS の声道特性から再合成した音声は、大幅な声質改善効果が得られなかった。これに対して提案法の AR-HMM 分析で得られた声道特性から再合成した音声は、明瞭な音韻性を持ち雑音感の少ない音声が再合成されることを確認した。

健常者音声の逐次状態分割によるトポロジ推定 AR-HMM 分析結果を詳細に調べた結果、推定した声道特性に音源の特徴が残留する場合が生じることを確認した。この現象について更に詳細な解析を行った結果、音源の残留特徴が声道フィルタの実軸上の極として現れることを明らかにした。この結果を踏まえて、そのような極の発生を抑制する制約条件を導出し、それを組み込んだ新しい AR-HMM 分析法を構築した。そして、従来の GIF と比較して、音源の推定精度が良くなることを実験的に確認した。また計算量を削減するために、事前学習した音源 HMM に基づいて時変利得と時変 AR 係数を適応推定する分析法においても、同様の制約条件を導入した AR 係数の適応推定法を構築した。

食道発声音声から声道特性を高精度に推定する音声分析法の研究開発の他に、声質改善音声の合成に必要は基本周波数(F0)パターンの生成に関する検討を行った。本研究では、電気式人工喉頭にあるような発話開始から緩やかに F0 が下降するような単純な F0 生成ではなく、平常音声も含めた 4 種類の感情音声の F0 時系列を、Generative Adversarial Networks(GAN)により生成する手法を検討した。具体的には、各感情の統計的性質を保持しつつ出来るだけ多様な F0 パターンを生成可能な Generator のモデル構造を、生成した F0 パターンの感情識別率と局所密度の 2 つの指標を用いて評価した。提案法を含めて５種類のモデル構造を用いて実験した結果、提案法が生成する F0 パターンの感情識別率が最も高く、局所密度は２番目に小さい値となった。これより提案法は、異なる４つの感情の特徴を最もよく反映する F0 パターンを生成し、更に各感情のクラス内で変化に富む F0 パターンを生成できる傾向があることを確認した。

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Akira Sasou	４．巻 104
２．論文標題 Glottal inverse filtering by combining a constrained LP and an HMM-based generative model of glottal flow derivative	５．発行年 2018年
３．雑誌名 Speech Communication	６．最初と最後の頁 113-128
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1016/j.specom.2018.07.002	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計4件（うち招待講演　0件／うち国際学会　3件）

１．発表者名 Akira Sasou
２．発表標題 Automatic Identification of Pathological Voice Quality Based on the GRBAS Categorization
３．学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2017)（国際学会）
４．発表年 2017年

１．発表者名 Akira Sasou, Nyamerdene Odontsengel, Shumpei Matsuoka
２．発表標題 An Acoustic-based Tracking System for Monitoring Elderly People Living Alone
３．学会等名 4th International Conference on Information and Communication Technologies for Ageing well and e-Health（国際学会）
４．発表年 2018年

１．発表者名 松岡 駿平, 佐宗 晃
２．発表標題 病的音声のためのGRBAS尺度の自動推定に関する検討
３．学会等名 電子情報通信学会大会
４．発表年 2018年

１．発表者名 Shumpei Matsuoka, Yao Jiang, Akira Sasou
２．発表標題 Generation of Artificial F0-contours of Emotional Speech with Generative Adversarial Networks
３．学会等名 2019 IEEE Symposium Series on Computational Intelligence (SSCI)（国際学会）
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