(12) 特許公報(B2)

(11) 特許番号

(24) 登録日 平成22年9月17日 (2010.9.17)

特許第4586675号

(P4586675)

(45) 発行日 平成22年11月24日(2010.11.24)

(19) **日本国特許庁(JP)**

| (51) Int.Cl. | | | FΙ | | |
|--------------|-------|-----------|---------|-------|------|
| A61B | 5/11 | (2006.01) | A 6 1 B | 5/10 | 310K |
| A 6 1 B | 5/107 | (2006.01) | A 6 1 B | 5/10 | 300Z |
| G1OL | 11/00 | (2006.01) | G1OL | 11/00 | 501 |
| A61B | 5/055 | (2006.01) | A 6 1 B | 5/05 | 380 |
| | | | | | |

請求項の数 4 (全 18 頁)

| (01)山岡平日 | | (79) 株式 日本 オケナ | × 909091E00 | |
|---------------------------|------------------------------|--------------------|--------------------|--|
| (21) 出願會亏 | 特旗2005-238235 (P2005-238235) | (73)符計権者 393031586 | | |
| (22) 出願日 | 平成17年8月19日 (2005.8.19) | | 株式会社国際電気通信基礎技術研究所 | |
| (65)公開番号 | 特開2007-50143 (P2007-50143A) | | 京都府相楽郡精華町光台二丁目2番地2 | |
| (43) 公開日 | 平成19年3月1日 (2007.3.1) | (74)代理人 | 100099933 | |
| 審査請求日 | 平成19年11月16日 (2007.11.16) | | 弁理士 清水 敏 | |
| | | (72)発明者 | パーハム・モクタリ | |
| (出願人による申告)平成17年度独立行政法人情報通 | | | 京都府相楽郡精華町光台二丁目2番地2 | |
| 信研究機構、研究テーマ「人間情報コミュニケーション | | | 株式会社国際電気通信基礎技術研究所内 | |
| の研究開発」に関する委託研究、産業活力再生特別措置 | | (72)発明者 | 北村 達也 | |
| 法第30条の適用を受ける特許出願 | | | 京都府相楽郡精華町光台二丁目2番地2 | |
| | | | 株式会社国際電気通信基礎技術研究所内 | |
| | | (72)発明者 | 竹本 浩典 | |
| | | | 京都府相楽郡精華町光台二丁目2番地2 | |
| | | | 株式会社国際電気通信基礎技術研究所内 | |
| | | | | |
| | | | 最終頁に続く | |

(54) 【発明の名称】声道断面積関数の推定装置及びコンピュータプログラム

(57)【特許請求の範囲】

【請求項1】

複数個の母音の各々についての、所定の話者の個々の発話時における静止透過画像と、 前記複数個の母音の連続発話時の、前記所定の話者の透過動画像と、前記連続発話時の前 記所定の話者の音声データとに基づいて、前記個々の発話時の前記話者の声道断面積関数 の平均及び当該平均回りの変化を表わす第1~第m(mは2以上の自然数)の主成分と、 前記連続発話時の前記発話者の声道断面積関数を前記平均及び前記第1~第mの主成分に より近似するための前記第1~第mの主成分の重みを、前記所定の話者の音声データに対 するケプストラム分析により得られる第1次~第n次(nは予め定められた10以上の自 然数)のケプストラム係数の線形和で近似するための係数とを準備するためのパラメータ 準備手段と、

10

20

話者の音声をキャプチャして得られる、フレーム化された音声信号に対し、フレームご とに前記ケプストラム分析を行ない、前記第1次~第n次のケプストラム係数をフレーム ごとに算出するための第1のケプストラム分析手段と、

前記第1のケプストラム分析手段の算出する前記第1次~第n次のケプストラム係数に 対し、前記パラメータ準備手段により準備された前記係数を用いた前記線形和を用いて、 前記話者の音声信号の各フレームに対する前記第1~第mの主成分のための重みを算出す るための線形変換部と、

前記線形変換部により算出された前記第1~第mの主成分のための重みと、前記パラメ ータ準備手段により準備された前記平均及び前記第1~第mの主成分とを用いて、前記話 者の前記音声信号の各フレームに対し、声道断面積を推定するための声道断面積推定手段 とを含む、声道断面積関数の推定装置。

【請求項2】

前記パラメータ準備手段は、

前記複数個の母音の各々についての、前記所定の話者の個々の発話時の声道部分の静止 透過画像により得られる声道断面積関数に基づいて、当該複数個の母音に関する声道断面 積関数の平均と、前記第1~第mの主成分とを主成分分析により求めるための主成分分析 手段と、

前記透過動画像よりフレームごとに得られる前記連続発話時の声道断面積関数を前記第 1~第mの主成分により形成される空間に射影する事により、前記連続発話時の声道断面 10 積関数を前記第1~第mの主成分と前記平均とにより表わすための、前記第1~第mの主 成分の重みを算出するための重み算出手段と、

前記連続発話時の前記所定の話者の音声データに対する前記ケプストラム分析を行なう 事により、前記第1次~第n次のケプストラム係数を求めるための第2のケプストラム分 析手段と、

前記第2のケプストラム分析手段により得られる前記第1次~第n次のケプストラム係 数の線形和で、前記連続発話時の前記第1~第mの主成分の重みを近似するための重回帰 分析を行なって前記係数を算出するための重回帰分析手段と、

前記平均、前記第1及び第2の主成分、並びに前記第1~第mの主成分の重みを近似す るための前記係数を記憶するための記憶手段とを含む、請求項1に記載の声道断面積関数 ²⁰ の推定装置。

【請求項3】

前記パラメータ準備手段は、前記声道断面積関数の平均及び前記第1~第mの主成分と、 前記係数とを記憶するための記憶手段を含む、請求項1に記載の声道断面積関数の推定装 置。

【請求項4】

コンピュータにより実行されると、当該コンピュータを請求項1~請求項3のいずれかに 記載の声道断面積関数の推定装置として動作させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明はMRI(Magnetic Resonance Imaging)画像の ような透過画像を用いた人間の調音機構の分析技術に関し、特に、静止透過画像と透過動 画像とに基づき、人間の発話から得られる音声に基づき、発話時の声道断面積関数をリア ルタイムで推定する声道断面積関数の推定装置に関する。

【背景技術】

【0002】

音響から調音器官の構成への変換は逆推定法と呼ばれるが、一般的に非線形で1対1の 関係が成立しない事が知られている。これは(音源を含む)調音器官の間での補償的関係 から生じ、そのために互いに異なる声道形状から非常に良く似た音質の音声が生じる事が あるためである。

【 0 0 0 3 】

しかし、これらの問題の困難さは結局のところ選択する調音機構の形状パラメータと音響的パラメータとに依存しており、適切な制約を課す事により解決できる。従来技術において提案された最も重要な制約のタイプには二つある。人間の発声機構をより厳密に模擬した人間型のモデルを提唱するもの(非特許文献1)と、調音器官の時間的軌跡に連続性を課すものとである(非特許文献2)。

【0004】

これに代えて、仮に調音器官の間で自然に生ずる共変動を、より少ない自由度で捕らえ る事ができれば、制約をさらに加える事なく、容易にモデルを逆推定法に用いる事ができ ⁵⁰

る可能性がある。実際のところ、そのような、測定された調音器官のデータの次元数を削減する事は、母音生成に関するいくつかの研究での関心事となっている。アメリカ英語の 母音については、声道形状の変化のうち90%については、平均を中心とする変化を記述 する、わずか二つの直交成分で説明できる事が明らかにされている。同様の成分について はアイスランド語の母音でも報告されており、ドイツ語の母音に関する研究でそれらにつ いての言語間に共通した有効性が支持されている。実際、非特許文献3に記載された、モ デルを用いた実験により得られた結果の類似性によれば、それら二つの成分が言語に対し て独立であるだけではなく、人間の声道の解剖学的及び生体力学的な属性に本質的に備わ っているものである、という仮説が妥当と思われる。

【非特許文献1】J.シュレータ他、「音声信号から声道形状を推定する技術」、IEE 10 E音声及び音響処理論文集、2(1)、pp.133-150、1994年 (Schroeter , J. & Sondhi, M. M., "Techniques for estimating vocal-tract shapes f rom the speech signal", IEEE Trans. Speech & Audio Proc., 2(1), 133-1 50, 1994.)

【非特許文献2】J.ダン他、「生理的調音モデルを用いた、発話音声からの声道形状推定」、ジャーナル・オブ・フォネティクス、30(3)、pp.511-532、200 2年 (Dang, J. & Honda, K., "Estimation of vocal tract shapes from s peech sounds with a physiological articulatory model", J. Phonetics., 3 0(3): 511-532, 2002.)

【非特許文献3】P.ペリエ他、「発話における舌運動の自由度は生体力学により制限さ²⁰ れている可能性がある」、第6回音声言語処理国際会議(ICSLP)予稿集、北京、中国 、第II巻、pp.162-165、2000年 (Perrier, P., Perkell, J., Pay an, Y., Zandipour, M., Guenther, F. & Khalighi, A., "Degrees of freed om of tongue movements in speech may be constrained by biomechanics", in Proc. 6th Int. Conf. Spoken Lang. Process. (ICSLP), Beijing, Chin a, Vol. II: 162-165, 2000.)

【非特許文献4】B.H.ストーリ他、「経験的直交モードによる声道断面積関数のパラ メータ化」、ジャーナル・オブ・フォネティクス、26(3)、pp.223-260、 1998年 (Story, B. H., & Titze, I. R., "Parameterization of vocal tract area functions by empirical orthogonal modes", J. Phonetics, 26(3): 223-260, 1998.)

【非特許文献5】P.ラドフォジッド他、「フォルマント周波数からの声道形状生成」、 米国音響学会ジャーナル、64(4)、pp.1027-1035、1978年 (Ladef oged, P., Harshman, R., Goldstein, L. & Rice, L., "Generating vocal t ract shapes from formant frequencies", J. Acoust. Soc. Am., 64(4): 102 7-1035, 1978.)

【非特許文献6】H.ワキタ、「声道長を用いた母音の正規化及びその母音識別への応用」、IEEE音響・音声・信号処理論文集、25(2)、pp.183-192、197 7年 (Wakita, H. "Normalization of vowels by vocal-tract length and it s application to vowel identification", IEEE Transactions on Acoustics, Speech, and Signal Processing, 25(2), 183-192, 1977)

40

30

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 5 】

従って、声道形状に関する上記した二つの成分を逆推定法に用いる事により、よい結果 を得られる可能性がある。こうした試み自体、過去に行なわれた事はある(非特許文献4)。しかしこの従来技術では合成音声のフォルマントを用いており、従って逆推定法では フォルマントを得るために使用した特定の声道伝達線モデルの特性を学習したものとなり 、人間の声道の音響的特徴を学習したものとはならない。

[0006]

対照的に、非特許文献5では、透過撮影法による撮影と同時に録音された音声から容易 ではない手段で測定したフォルマントに対する線形回帰を適用し、各フォルマントに対す る舌形状を定める第1及び第2の要素とそれぞれ0.935及び0.902の相関がある 事を見出している。

【 0 0 0 7 】

しかし、音声の発生期間が長く安定している母音についてさえ、フォルマントを簡単な 処理で安定して精度高く測定できない事は良く知られている。従って、今まで、多くの研 究により、逆推定法のために、安定して精度高く測定できる音響パラメータを用いる事の 重要性が強調されてきたが、そのための具体的な方策は与えられていない。

従って、本発明の目的は、測定の容易なパラメータを用い、精度の高い推定をリアルタ イムで行なう事ができる声道断面積関数の推定装置を提供する事である。

【課題を解決するための手段】

【0009】

本発明の第1の局面によれば、声道断面積関数の推定装置は、複数個の母音の各々につ いての、所定の話者の個々の発話時における静止透過画像と、複数個の母音の連続発話時 の、所定の話者の透過動画像と、連続発話時の所定の話者の音声データとに基づいて、個 々の発話時の話者の声道断面積関数の平均及び当該平均回りの変化を表わす第1~第m(mは2以上の自然数)の主成分と、連続発話時の発話者の声道断面積関数を平均及び第1 ~ 第mの主成分により近似するための第1~第mの主成分の重みを、所定の話者の音声デ ータに対するケプストラム分析により得られる第1次~第n次(nは予め定められた10 以上の自然数)のケプストラム係数の線形和で近似するための係数とを準備するためのパ ラメータ準備手段と、話者の音声をキャプチャして得られる、フレーム化された音声信号 に対し、フレームごとにケプストラム分析を行ない、第1次~第n次のケプストラム係数 をフレームごとに算出するための第1のケプストラム分析手段と、第1のケプストラム分 析手段の算出する第1次~第n次のケプストラム係数に対し、パラメータ準備手段により 準備された係数を用いた線形和を用いて、話者の音声信号の各フレームに対する第1~第 mの主成分のための重みを算出するための線形変換部と、線形変換部により算出された第 1~第mの主成分のための重みと、パラメータ準備手段により準備された平均及び第1~ 第mの主成分とを用いて、話者の音声信号の各フレームに対し、声道断面積を推定するた めの声道断面積推定手段とを含む。

[0010]

静止透過画像からは、個々の母音の発話時の声道断面積関数の平均値と、平均値を中心 とする変動を表わす第1~第mの主成分が準備される。透過動画像からは、連続発話時の 声道断面積関数を、静止透過画像から得た声道断面積関数の平均値と、第1~第mの主成 分とで近似するための、第1~第mの主成分に対する重みがフレームごとに準備される。 音声データに対するケプストラム分析により得られた第1次~第n次のケプストラム係数 でこれら重みを線形和で近似するための係数が準備される。これらが準備された後、話者 の音声をキャプチャして得られたフレーム化された音声信号から第1次~第n次のケプス トラム係数が得られる。このケプストラム係数に対し、パラメータ準備手段に準備された 係数を用いた線形変換を行なう事により、フレームごとの第1~第mの主成分のための重 みが得られる。この重みと、声道断面積関数の平均と、第1~第mの主成分とを用いて、 声道断面積関数の近似値がフレームごとに算出される。フォルマントのように測定が困難 なパラメータではなく、ケプストラム係数という測定の容易なパラメータを用い、精度の 高い推定をリアルタイムで行なう事ができる。

[0011]

さらに好ましくは、パラメータ準備手段は、複数個の母音の各々についての、所定の話 者の個々の発話時の声道部分の静止透過画像により得られる声道断面積関数に基づいて、 当該複数個の母音に関する声道断面積関数の平均と、第1~第mの主成分とを主成分分析 により求めるための主成分分析手段と、透過動画像よりフレームごとに得られる連続発話 10

20

時の声道断面積関数を第1~第mの主成分により形成される空間に射影する事により、連 続発話時の声道断面積関数を第1~第mの主成分と平均とにより表わすための、第1~第 mの主成分の重みを算出するための重み算出手段と、連続発話時の所定の話者の音声デー タに対するケプストラム分析を行なう事により、第1次~第n次のケプストラム係数を求 めるための第2のケプストラム分析手段と、第2のケプストラム分析手段により得られる 第1次~第n次のケプストラム係数の線形和で、連続発話時の第1~第mの主成分の重み を近似するための重回帰分析を行なって係数を算出するための重回帰分析手段と、平均、 第1及び第2の主成分、並びに第1~第mの主成分の重みを近似するための係数を記憶す るための記憶手段とを含む。

【0012】

10

20

上記した各パラメータを算出するための手段を備える事により、一つの声道断面積関数の推定装置で、パラメータの学習と実際の処理との双方を行なう事ができる。 【0013】

第1のケプストラム分析手段と、第2のケプストラム分析手段とは、同じ周波数帯域の 音声信号に対するケプストラム分析を行なうようにしてもよい。

[0014]

パラメータ準備手段により準備されるのと同じ周波数帯域の音声信号から得られたケプ ストラム係数を動作時に用いる事により、声道断面積関数の推定を正しく行なう事ができ る。

【0015】

好ましくは、第1のケプストラム分析手段と、第2のケプストラム分析手段とは、とも に0~4 k H z の周波数帯域の音声信号に対するケプストラム分析を行なう。

[0016]

話者にもよるが、この周波数帯域の音声信号に対するケプストラム分析の結果を用いる 事により、声道断面積関数の推定を精度良く行なえる事が判った。

【0017】

さらに好ましくは、パラメータ準備手段は、声道断面積関数の平均及び第1~第mの主 成分と、係数とを記憶するための記憶手段を含む。m=2、n=24でもよい。

【0018】

実験によれば、主成分としては第1及び第2の主成分を用いると十分な精度で推定が行 30 なえる。またケプストラム係数として第1次~第24次までを用いた場合が最も推定の精 度が高い。

【0019】

声道断面積関数が、所定の話者の声道の中心線上の、互いの等しい間隔をおいた所定個 数のセクション位置での声道の断面積と、セクション位置間の距離とを含むようにしても よい。

【0020】

セクション位置間の距離まで含めて主成分分析を行なう事により、母音により異なる声 道長まで含めた推定を精度高く行なう事ができる。

【0021】

40

本発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行される と、当該コンピュータを上記したいずれかの声道断面積関数の推定装置として動作させる ものである。従って、上記した声道断面積関数の推定装置と同様の効果を得る事ができる

【発明を実施するための最良の形態】

[0022]

< 構 成 >

図1に、本発明の一実施の形態に係る、逆推定法を用いて音声から声道断面積関数をリ アルタイムで推定し表示する声道断面積関数推定システム100のブロック図を示す。図 1を参照して、声道断面積関数推定システム100は、ある特定の話者について「ア」「 イ」「ウ」「エ」「オ」の各母音を発話しているときの声道画像を撮影して得られる5組 の静止MRI画像110と、同じ話者が「アイウエオ」という一連の発話をしているとき の声道の動画像を撮影したMRI動画像114と、同じ話者が「アイウエオ」という一連 の発話をしているときの音声を録音した音声データ116とを用い、逆推定法に必要なパ ラメータを算出するためのパラメータ算出部118を含む。理想的には、音声データ11 6はMRI動画像114を撮影するときの話者の音声を録音すればよいが、実際にはMR Iの撮影には非常に大きな音が発生するため、撮影と録音とを同時に行なう事は困難であ る。従って本実施の形態では、MRI動画像114を撮影するときとは別の時点に音声デ ータ116の録音をする。しかし、両者は互いに同期している必要があるため、話者にへ ッドフォンを装着させ、同期音発生装置112により発生させた規則的な同期音をこのへ ッドフォンを通じて話者に聞かせ、その同期音にあわせて「アイウエオ」の発声を行なう ようにする。なお、音声データ116の録音時には、声道の形状がMRI動画像114の 撮影時のそれとできるだけ一致するよう、話者の姿勢もMRI動画像114の撮影時と同 様にした。

【 0 0 2 3 】

声道断面積関数推定システム100はさらに、パラメータ算出部118により算出され たパラメータを用いて、マイクロフォン122により音声信号に変換された上記話者の音 声からその声道断面積関数をリアルタイムで逆推定し、表示装置124に表示するための 逆推定処理部120を含む。

[0024]

パラメータ算出部118が推定するパラメータは、静止MRI画像110から推定され た、各母音の発声中の発話者の声道断面積関数の平均、その変動に関する主成分分析(P CA)の第1成分と第2成分の値、及びMRI動画像114から得られる上記二つの主成 分に対する重みの変化と、音声データ116から得られるケプストラム係数との間の重回 帰分析により得られる回帰係数である。この重回帰分析の詳細については後述するが、得 られる回帰係数は、5個の母音を連続して発声したときの二つの主成分に対する重み(の 変化)を、音声データ116から得られるケプストラム係数の線形和で近似するための、 各ケプストラム係数の重みである。

【0025】

本実施の形態では、MRI動画像114は、発話者が「アイウエオ」を連続して発声す 30 る際の声道形状を撮影して得られる、連続した35フレーム分の画像からなっている。好 ましくは、MRI動画像114は複数回の撮影から得るようにする。同期音発生装置11 2を用いる事により、それら複数回の撮影でも一定のタイミングで「アイウエオ」の発声 を行なう事ができる。

【0026】

パラメータ算出部118は、静止MRI画像110の5個の母音に対して得られた5組 のMRI静止画像の各々から各母音の発声時の声道断面積関数を求め、それらに対するP CA処理を行なって、それら声道断面積関数の平均値と、平均値を中心とした声道断面積 関数の変動を主に支配する第1及び第2の主成分(それぞれ「PC I」及び「PC Т I」と略記する。)とを出力するためのPCA処理部140と、PCA処理部140から 出力された声道断面積関数の平均値、PC I及びII、並びにそれらに付随する後述す る所定の定数を記憶するための記憶部142と、MRI動画像114から得られた、時間 的に変化する声道断面積関数に対して記憶部142に格納されている声道断面積関数の平 均値及びPC I及びIIとを射影する事により、MRI動画像114により表わされる 声道断面積関数の変化に対応する、 PC Ⅰ及びⅠⅠに対するフレームごとの重みk 」及 び k ,の 変 化 曲 線 を 算 出 す る た め の 動 画 像 デ ー 夕 射 影 処 理 部 1 4 4 と 、 動 画 像 デ ー 夕 射 影 処理部144により算出された、PC I及びIIに対するフレームごとの重みk」及び k 。を記憶するための記憶部146とを含む。 P C A 処理部140での処理及び動画像デ -タ射影処理部144での処理の詳細については後述する。 [0027]

(6)

20

10

40

パラメータ算出部118はさらに、音声データ116により得られた「アイウエオ」に 関する連続発声について、dBスケールでFFT(高速フーリエ変換)された所定周波数 帯域内のスペクトルサンプルに対する離散コサイン変換(DCT)によって、先頭(c₀)を除く、先頭から所定数のケプストラム係数(c₁~c_N)を算出するためのケプスト ラム係数算出部148と、ケプストラム係数算出部148により算出されたケプストラム 係数を記憶するための記憶部150と、記憶部146に記憶されたPC I及びIIの重 みk₁及びk₂の時間的変化曲線及び記憶部150に記憶されたケプストラム係数の間の 重回帰分析により、次の線形式を解く事によりk₁及びk₂を近似する回帰係数 m₀~

(7)

_{m n}を算出するための重回帰分析処理部152とを含む。 【0028】

【数1】

$$k_m = \alpha_{m0} + \sum_{n=1}^{N} \alpha_{mn} c_n, \qquad m = 1,...,M$$
 (1)

ただしNは算出されたケプストラム係数の数、Mは主成分の数、 moは定数、 mn(n=1~N)はm番目の主成分を近似するための、n番目のケプストラム係数 c n にかか る回帰係数である。本実施の形態では、N=24、M=2であり、 mo及び m1~ m 24は、動画像データ射影処理部144により算出されたk1及びk2と重回帰分析処理 部152により算出されたケプストラム係数c1~c24とを式(1)に代入して m0 及び m1~ m24について式(1)を解く事により得られる。式(1)により、この 重回帰分析により得られる重回帰モデルが表わされる。

【0029】

パラメータ算出部118はさらに、重回帰分析処理部152により算出された回帰係数 m1~ m24を記憶するための記憶部154を含む。

[0 0 3 0]

一方、逆推定処理部120は、マイクロフォン122から得られた音声信号をキャプチ ャし、フレーム化したデジタル信号に変換するための音声キャプチャ処理部180と、音 声キャプチャ処理部180によりデジタル信号に変換された音声信号に対し、ケプストラ ム係数算出部148と同様の処理を行なってケプストラム係数 с1~ с24をフレームご とに算出するためのケプストラム係数算出部182と、ケプストラム係数算出部182に より算出されたケプストラム係数 c1 ~ c24、及び記憶部 154 に記憶された回帰係数 mo及び m1~ m24を用い、式(1)により示される線形変換に従ってPC I及 びIIのための重みk₁及びk₂を算出するための線形変換処理部184と、線形変換処 理部184により各フレームに対して算出された重みk₁及びkっと、記憶部142に記 憶された平均声道断面積関数及びPC I及びIIとを用い、後述する式に従って各フレ ームにおける声道断面積関数の推定値を算出するための声道断面積関数推定部186とを 含む。線形変換処理部184及び声道断面積関数推定部186の機能の詳細については後 述する。本実施の形態では、このようにして各フレームごとに算出された声道断面積関数 を、表示装置124に与え、表示する。この表示により、例えば被験者が所定の発声をす るときの声道断面積関数の変化を確認する事ができる。外国語の学習などにおいて、この 声道断面積関数とともに手本となる声道断面積関数を表示すれば、正しい発音をするため

[0031]

以下、PCA処理部140で行なわれる処理について説明する。本実施の形態では、M RI画像から0.25cmのスライス間隔で声道断面画像を作成し、さらに声道の中心線 上に沿って、互いに等しいセクション長の間隔を隔てた44個の位置での声道断面積を再 サンプリングする。さらに、こうして求められた声道断面積の平方根を算出する。これに より各声道断面積関数を表わす44次元のベクトル×が得られる。母音による声道長の変 化を補償するため、このベクトル×の45番目の要素として、上記セクション長を加える

には声道断面積関数をどのように変化させればよいかを確認しながら発音の練習ができる

10

20

30

。ただし、このセクション長の変動は、声道断面積の変動の最大値と一致するように正規 化される。

【0032】

従って、再サンプリングされた要素からなる、 i 番目(1 i 44)のセクション位置の、 v 番目(1 v 5)の母音に関する声道断面積をA_{i,v}とすると、 v 番目の母音に関する声道断面積関数を表わすベクトル x_vの各要素 x_{i、v}は次の式(2)により表わされる。

【0033】

【数 2 】

$$\mathbf{x}_{i,v} = \begin{cases} \sqrt{A_{i,v}}, & i = 1,...,44\\ \underline{\left(\delta_{v} - \overline{\delta}\right)}\sigma_{A}} \\ \underline{\sigma_{\delta}} + \overline{\delta}, & i = 45, \end{cases}$$
(2)

ただし/ (「/」は式中の上線を表わす。)及び はそれぞれ、 v 個の母音に対応す る v 個のセクション長 v の平均及び標準偏差である。 A は断面積の平方根 x i, v (i = 1 ~ 4 4)の標準偏差の最大値を表わす。さらに、 P C A 処理部 1 4 0 は、これらベ クトル x の平均値ベクトル / x も算出する。こうした値はいずれも記憶部 1 4 2 に記憶さ れる。

【0034】

このベクトルに対する主成分分析により、平均ベクトル / x と4 5 個の固有ベクトル e m とが得られる。これらの固有ベクトルのうち少なくとも(v - 1) = 4 個の固有ベクト ルが一意に求められ、非零の固有値と関連付けられる。

【0035】

このようにして求められた最初の二つの主成分の例を図2(PC I)及び図3(PC II)に示す。さらに、第3の主成分(PC III)についても図4に示す。 【0036】

図2(A)、図3(A)及び図4(A)はそれぞれ固有ベクトルを示し、図2(B)、 図3(B)及び図4(B)は、各固有ベクトルによって平均声道断面積関数がどのような 影響を受けるかを示す。図2(B)、図3(B)及び図4(B)において、実線は上記の ようにして求められた声道断面積関数を示す。破線は正の方向に、点線は負の方向に、各 主成分を、対応する固有ベクトルの方向に標準偏差に対応する量だけ変化させたときに得 られる、変化後の声道断面積関数を示す。このグラフから、各主成分を変化させたときに 声道断面積関数がどのように変化するかが判る。図2(A)、図3(A)及び図4(A) において、「」は正の固有ベクトルの44個の要素とセクション長、「」は負に符号 を変えた固有ベクトルの44個の要素の値とセクション長をそれぞれ示す。 【0037】

発明者らが行なった事前実験によれば、二つの成分 PC I及び IIにより、全変動量の88.8%及び8.5%がそれぞれ表わされている。従って、これら二つの成分により、合計で発話者の母音生成空間内における全変動量の97%が表わされる。

【0038】

PC Iは開口時の咽頭の狭窄と、咽頭腔を開いた状態での口腔狭窄との間での変動を 表わす。PC Iはまた、口腔前部での母音については短く、後部での母音については長 くなるように、(発話者にとって)正しく声道長を調整するための共変動も表わす。 【0039】

図3に示すPC IIは、上部咽頭及び軟口蓋域付近の領域(声門から7ないし13c m)と、唇とにおける、付随する狭窄の程度の変動を表わす。これら二つの位置における 狭窄が大きくなると、それに伴い声道長が長くなる。これと対照的に、図4に示す第3番 目の主成分PC IIIIは、全変動のうちの残りのうち、1.5%しか表わさず、その状態もエラー及び雑音を含むように見受けられ、調音上の用語で説明するのは困難である。

10

20

30

[0040]

なお、上記したように静止MRI画像から得たPC I及びIIにより表わされた声道 面積と声道長とが、日本語の5つの母音に関して、後述するように動画像MRIから得た 実際のデータとよく一致する事は、発明者たちが確認している。これら二つの主成分を用 いる事により、以下に述べるように音声から声道断面積関数への逆推定を行なう事が可能 と考えられる。

【0041】

次に、動画像データ射影処理部144で行なわれる処理について説明する。本実施の形態では、MRI動画像114は毎秒30フレーム(フレーム間隔33ミリ秒)で記録する。そして、各フレームにおける3D画像から、声門から唇の放射面までの声道断面を、0.25cm間隔でサンプリングし、さらに声門から唇までの声道を互いに等しい長さに等分した44個のセクション位置で声道断面積を再サンプリングする。ただしこの前に、歯の影響を考慮するために3D画像にデジタル的に歯の映像を追加する。この結果、「ア」から「オ」までの連続発声に関し、35フレームの画像の各々について、声道断面積関数が得られる。

[0042]

動画像データ射影処理部144は、このMRI動画像の35フレーム分の声道断面積関数の各々を、PC I-PC II空間に射影する事により、声道断面積関数を逆推定するために必要なPC I及びIIの重みk₁及びk₂のフレームごとの値を算出する。一般に、あるベクトル×を、主成分に対応する固有ベクトルで表すための、各固有ベクトルに対する重みk_m(m=1~使用する固有ベクトルの数)は、次の式(3)で求められる

【0043】

【数3】

 $\mathbf{k} = \mathbf{E}^{\mathrm{T}} \left(\mathbf{x} - \mathbf{x} \right)$

(3)

ただしベクトル k は重み k m を要素とするベクトル、 E は固有ベクトル e m を各列に持つ 行列、「 E ^T」は行列 E の転置行列、 / x はベクトル x の平均値ベクトルを、それぞれ示 す。

【0044】

ケプストラム係数算出部148は、本実施の形態においては、発話者の音声のうち、0 ~4kHzの周波数帯域から24個のケプストラム係数を算出する。これは、ある被験者 について繰返した実験において、動画像MRIから得られた声道断面積関数を再構成する 上で、この組合せが最も高い精度を示したためである。従ってこの組合せは人により異な る可能性がある。ただし、ケプストラム係数は、ほぼ実用的な精度を得るためには少なく とも10個、より実用的な精度を得るためには少なくとも11個は必要である事が実験に より確認された。

[0045]

重回帰分析処理部152は、記憶部146に記憶されているk₁及びk₂の値と、記憶 部150に記憶されたケプストラム係数c₁~c₂₄とを用いて、前述した式(1)を立 4 て、その式を解く事で回帰係数 _{m0}~ _{mN}(m=1,2、N=24)を得る。 【0046】

逆推定処理部120の線形変換処理部184は次のような機能を持つ。マイクロフォン 122からの音声信号を音声キャプチャ処理部180及びケプストラム係数算出部182 で処理する事により、各フレームごとの発話者の音声のケプストラム係数 c₁ ~ c₂₄が 得られる。一方、記憶部154にはパラメータ算出部118により式(1)で使用される 回帰係数 m₀ ~ m₂₄(m=1,2)が記憶されている。線形変換処理部184は、 このケプストラム係数 c₁ ~ c₂₄及び回帰係数 m₀ ~ m₂₄を式(1)に代入して 得られる式を用い、各フレームにおける P C I 及び I I に対する重みk₁及びk₂を算 出する機能を持つ。 10

20

30

[0047]

声道断面積関数推定部186は、線形変換処理部184から各フレームごとに与えられた重みk₁及びk₂と、記憶部142に記憶された平均声道断面積関数及びPC I及び IIとを用い、各フレームごとにまず次の式によって声道断面積の平方根からなるベクト ル ^ xを推定する。なお本明細書中における記号「^」は、式中では直後の文字の上に置 かれている。

(10)

【0048】 【数4】

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \sum_{m=1}^{M} k_m \mathbf{e}_m, \qquad (4)$$

ただし本実施の形態では式(4)におけるM=2、/×はPCA処理部140により算出 されたベクトル×の平均値、 e m はPCA処理部140で算出されたm番目の固有ベクト ルを、それぞれ示す。

【0049】

声道断面積関数推定部186は、こうして各フレームに対しベクトル × を推定した後、以下の式(5)により単純に式(1)の逆変換を行なう事により、各フレームにおける i番目(1 i 44)のセクション位置における推定声道断面積 ^ A _i と、推定セクション長 ^ とを算出する。

【0050】

【数5】

$$\hat{A}_{i} = \hat{x}_{i}^{2} \qquad i = 1,...,44$$
$$\hat{\delta} = \frac{\left(\hat{x}_{45} - \overline{\delta}\right)\sigma_{\delta}}{\sigma_{A}} + \overline{\delta}$$
(5)

ただし^×_iはベクトル^×のi番目の要素を、^×₄ 5 はベクトル^×の45番目の要素を、 _AはPCA処理部140により算出された断面積の平方根×_{i,∨}(i=1~4 4)の標準偏差の最大値を、/ 及び はそれぞれ、∨個の母音に対応する∨個のセク ション長 _Vの平均及び標準偏差を、それぞれ示す。i個の推定声道断面積^A_iを要素 とするベクトル^Aが、当該フレームにおける推定声道断面積関数を表わす。 【0051】

<動作>

声道断面積関数推定システム100は以下のように動作する。声道断面積関数推定シス テム100の動作には二つのフェーズがある。第1のフェーズは、パラメータ算出部11 8によるパラメータの学習を行なう学習フェーズである。より具体的には、図1を参照し て、ある話者に対する「ア」「イ」「ウ」「エ」「オ」という発話時の静止MRI画像1 10、「アイウエオ」という連続発声時のMRI動画像114、及びMRI動画像114 の撮影時と時間的に同期するようにして測定した音声データ116から、「ア」「イ」「 ウ」「エ」「オ」という発話時の発話者の平均声道断面積関数及びその第1及び第2の主 成分PC I及びIIと、式(1)に示される回帰係数 m 0 ~ m 2 4 (m = 1,2) を算出する。第2のフェーズは推定フェーズである。推定フェーズでは、学習フェーズで 得られた発話者の平均声道断面積関数及びPC I及びIIと、回帰係数 m 0 ~ m 2 4 とを用い、マイクロフォン122から入力される、同じ発話者の音声に基づいてその声 道断面積関数をリアルタイムに推定し、表示装置124に表示する。以下、これら第1の フェーズと第2のフェーズとにおける声道断面積関数推定システム100の動作を順に説 明する。

【0052】

- 学習フェーズにおけるパラメータ算出部118の動作 -

この学習フェーズでは、ある話者についてまず静止MRI画像110、MRI動画像1 50

20

10

14、及び音声データ116を以下のようにして準備する。

[0053]

静止MRI画像110は、発話者に母音「ア」「イ」「ウ」「エ」「オ」をそれぞれ長めに発音してもらい、その間の発話者のMRI画像を撮影する。静止MRI画像110画像の撮影自体は通常のものと変わりない。

【0054】

MRI動画像114の画像にあたっては、発話者にヘッドフォンを装着させ、同期音発 生装置112により発生される同期音にあわせて「アイウエオ」という連続した発話を一 定時間かけて繰返させ、その間にMRI動画像を撮影する。本実施の形態では、1回の繰 返しで35フレームからなるMRI動画像を撮影する。好ましくは、この撮影を多数回繰 返し、それら繰返しによって得られた多数の動画像を合成して35フレームの画像を得る ようにする。

【 0 0 5 5 】

音声データ116の収録にあたっては、MRI動画像114の撮影と同じ条件で発話者 が発話する事が必要である。そこで、発話者にはMRI動画像114の撮影時と同じ姿勢 (仰向け)をとってもらい、ヘッドフォンを装着させ、MRI動画像114の収録時と同 じように同期音発生装置112からの同期音に従って「アイウエオ」という発話を繰返さ せる。

【0056】

静止MRI画像110が準備されると、PCA処理部140が、「ア」「イ」「ウ」「20 エ」「オ」という5種類の母音の各々に対し、互いに等しいセクション長を隔てた44箇 所における声道断面積をMRI画像から算出する。さらにPCA処理部140は、母音に より異なるセクション長の変動を、声道断面積の変動の最大値と一致するように正規化す る。PCA処理部140は、こうして各母音に対して、44個の声道断面積と1個の正規 化後のセクション長とからなるベクトル×を算出する。ベクトル×の各成分については式 (2)により表わされている。

[0057]

[0058]

PCA処理部140はさらに、得られた5個のベクトル×に対しPCA処理を行ない、 ベクトル×の平均値ベクトル/×と、平均値ベクトル/×を中心とするベクトル×の変動 を表わす第1及び第2の主成分PC I及びIIとを算出する。得られた平均値ベクトル /×とPC I及びIIとは記憶部142に格納される。

30

10

動画像データ射影処理部144は、MRI動画像114の各フレームから声道断面積関数を算出する。算出された各フレームの声道断面積関数を、記憶部142に記憶されている平均声道断面積関数並びにPC I及びIIを用い、式(3)に従ってPC I-PC II空間に射影する。その結果、声道断面積関数を第1及び第2の主成分に対応する固有ベクトルで表わすための重みk₁及びk₂が各フレームにおいて算出される。各フレームごとの重みk₁及びk₂は記憶部146に記憶される。

【 0 0 5 9 】

 ケプストラム係数算出部148は、音声データ116の各フレームの音声データについ
 40

 て、0~4kHzの周波数帯域についてケプストラム係数c₁~c₂₄を算出し、記憶部

 150に記憶させる。重回帰分析処理部152は、記憶部146に記憶された重みk₁及びk₂と、記憶部150に記憶されたケプストラム係数c₁~c₂₄とを式(1)に代入し、この式を解く事により、回帰係数 m₀及び m₁~ m₂₄を算出する。算出された

 回帰係数 m₀及び m₁~ m₂₄は記憶部154に記憶される。

[0060]

以上がパラメータ算出部118の動作である。

[0061]

- 推定フェーズにおける逆推定処理部120の動作 -

発話者がマイクロフォン122前で発話すると、その音声はマイクロフォン122によ 50

り電気信号に変換されて音声キャプチャ処理部180に与えられる。音声キャプチャ処理 部180は、この音声信号をデジタル化し、フレーム化してケプストラム係数算出部18 2に与える。ケプストラム係数算出部182は、音声キャプチャ処理部180から与えら れるデジタル化された各フレームの音声信号のうち、0~4kHzの周波数帯域からケプ ストラム係数 c₁~ c₂₄を算出し、線形変換処理部184に与える。 【0062】

(12)

線形変換処理部184は、ケプストラム係数算出部182から与えられる各フレームの ケプストラム係数 c₁ ~ c₂₄と記憶部154に記憶されている回帰係数 m₀及び m₁ ~ m₂₄とを式(1)に代入する。これにより、PC I及びIIにそれぞれ対応する 重み k₁及び k₂が各フレームに対し算出される。線形変換処理部184は、算出された 重み k₁及び k₂を、フレームごとに声道断面積関数推定部186に与える。 【0063】

声道断面積関数推定部186は、各フレームごとに、線形変換処理部184から与えられた重みk₁及びk₂と、記憶部142に記憶された平均声道断面積関数/×及びPC I及びIIに対応する固有ベクトルe₁及びe₂とを式(4)に代入する事により、推定 声道断面積の平方根からなるベクトル^×を算出する。声道断面積関数推定部186はさ らに、得られたベクトル^×の各要素と、記憶部142に記憶された定数(セクション長 の平均/ 及び標準偏差 、断面積の平方根×_{i,v}(iはセクション番号で、i=1 ~44)の標準偏差の最大値 _A)とを用い、式(5)によって各フレームにおける推定 声道断面積 ^A_i(i=1~44)を推定する。これにより、各フレームにおける推定声 道断面積関数を表わすベクトル^Aが得られる。声道断面積関数予測部186は、各フレ ームにおける声道断面積関数をリアルタイムで視覚化し、表示装置124に表示させる。 視覚化の例については図6を参照して後述する。

20

10

[0064]

< 実験 >

上記した実施の形態では、ある発話者についての音声データ116のうち、0~4kH zの周波数帯域の音声についてケプストラム係数c₁~c₂4を算出している。しかし、 この値は発話者により変動する可能性がある。そのため、使用する周波数帯域と算出する ケプストラム係数の個数とを様々に変えて得られる重回帰モデルを用いて、他は同じ条件 で声道断面積関数の推定を行なった。その結果、いずれの周波数帯域を使用した場合も同 様によい結果が得られた。しかし、4kHz以下の周波数帯域を用いると、それ以上の周 波数帯域を用いた場合と比較して一般的に良い結果が得られた。最もよい結果が得られた のが、上記した実施の形態で述べた0~4kHzの周波数帯域を用いた場合である。ケプ ストラム係数としては、ある程度の効果を得るためには最低でもc₁~c₁を用いる必要があっ た。最もよい結果が得られたのが、上記した実施の形態の説明で述べたとおりc₁~c₂

[0065]

図5(A)及び(B)は、最も正確な予測が得られたモデルを示し、35フレームにわたる、測定値から得られた実際のPC I及びIIの値の変動(実線)と、0~4kHz の周波数帯域で定義された24個のケプストラム定数を用いて推定したそれらの値(印))とをそれぞれ示す。図5(A)及び(B)にはさらに、対比のために、0~3kHzの 周波数帯域で10個のケプストラム係数を用いて推定したPC I及びIIの値(印) をそれぞれ示す。最も良い回帰モデルでは、PC I及びIIの重みk₁及びk₂を推定 するに際し、それぞれ0.991及び0.968の相関、並びに標準偏差にして0.08 6と0.226の二乗平均誤差という結果であった。

【0066】

図 6 に、ケプストラムを用いた最も性能のよい回帰モデルにより得られた結果を視覚化 した例を示す。図 6 (A) ~ (E)はそれぞれ、MRI動画像から得られた、各母音の発 音時の声道断面積関数(点線)と、声道断面積関数の計測値から得られたPC I及びI

30

Iを用いて表わされた声道断面積関数(細線)と、上記実施の形態に従う最も性能のよい 回帰モデルにより得られた重みを用いて推定された声道断面積関数(太線)とを示す。 【0067】

(13)

図6では、「アイウエオ」と連続して発話した場合の、各母音に対応するフレームのみ を示しているが、連続する各フレームに対してこうしたグラフが得られる。従って、それ らフレームのグラフをフレームごとに切替えて表示する事により、発話者の発話と同時に 、リアルタイムでその声道断面積関数の変化を表示する事ができる。

【0068】 PCAとケプストラムとを組合わせた上記実施の形態によれば、35個のフレーム全て

に対する平均として、声道断面積において0.367 cm²、声道長において0.150 cmの誤差という結果が得られた。この誤差は、元のMRI動画像から計算されたPCA により表わした声道断面積及び声道長の誤差と比較してわずかに大きいだけである。一方 、非特許文献4に紹介された技術では、クローズドテストで得られた声道断面積の誤差が 0.334 cm² となっており、本実施の形態によるモデルはこれと遜色がない。さらに 、非特許文献6において線形予測モデルにより推定された日本語の5つの母音の声道長で の平均誤差0.84 cmと比較して、本実施の形態によるモデルでの誤差ははるかに小さ い。実際、本実施の形態での最も性能のよいモデルを用いたケースの声道長の誤差0.1 50 cmは、MRI画像で声道断面積画像を得るときのスライス長0.25 cmよりも短 く、十分に正確であるといえる。

[0069]

以上のように、本実施の形態によれば、正確な声道断面積関数を、発話者の音声にあわ せてリアルタイムで計算する事ができる。上記実施の形態では、学習時の発話者と、動作 時の発話者とが同じ人物である事が想定されているが、地声の声質が類似した発話者であ れば、別の発話者についてもリアルタイムで音声から声道断面積関数を計算できる。さら に、発話者による音質の相違を解消するように、発話者の音声を正規化できれば、適用で きる発話者の範囲はさらに広がる事が期待される。

[0070]

<コンピュータによる実現 >

この実施の形態のシステムは、コンピュータハードウェアと、そのコンピュータハード ウェアにより実行されるプログラムと、コンピュータハードウェアに格納されるデータと により実現される。図7はこのコンピュータシステム330の外観を示し、図8はコンピ ュータシステム330の内部構成を示す。

[0071]

図7を参照して、このコンピュータシステム330は、FD(フレキシブルディスク) ドライブ352及びCD-ROM(コンパクトディスク読出専用メモリ)ドライブ350 を有するコンピュータ340と、キーボード346と、マウス348と、モニタ342と 、マイク370と、スピーカ372とを含む。

[0072]

図8を参照して、コンピュータ340は、FDドライブ352及びCD-ROMドライ ブ350に加えて、CPU(中央処理装置)356と、CPU356、FDドライブ35 4 2及びCD-ROMドライブ350に接続されたバス366と、ブートアッププログラム 等を記憶する読出専用メモリ(ROM)358と、バス366に接続され、プログラム命 令、システムプログラム、及び作業データ等を記憶するランダムアクセスメモリ(RAM)360と、バス366、マイク370及びスピーカ372に接続され、マイク370か らの電気信号をデジタル化しフレーム化する処理と、CPU356により生成された音声 をアナログ信号化し、スピーカ372に与えるためのサウンドボード368とを含む。コ ンピュータシステム330はさらに、図示しないプリンタを含んでもよい。 【0073】

ここでは示さないが、コンピュータ340はさらに、ローカルエリアネットワーク(L AN)への接続を提供するネットワークアダプタボードを含んでもよい。 10

20



(14)

コンピュータシステム330に本実施の形態に係る声道断面積関数推定システム100 、又はその一部をなすパラメータ算出部118若しくは逆推定処理部120としての動作 を行なわせるためのコンピュータプログラムは、CD-ROMドライブ350又はFDド ライブ352に挿入されるCD-ROM362又はFD364に記憶され、さらにハード ディスク354に転送される。又は、プログラムは図示しないネットワークを通じてコン ピュータ340に送信されハードディスク354に記憶されてもよい。プログラムは実行 の際にRAM360にロードされる。CD-ROM362から、FD364から、又はネ ットワークを介して、直接にRAM360にプログラムをロードしてもよい。 【0075】

このプログラムは、コンピュータ340にこの実施の形態に係る声道断面積関数推定シ ステム100、又はその一部をなすパラメータ算出部118若しくは逆推定処理部120 としての動作を行なわせる複数の命令を含む。この動作を行なわせるのに必要な基本的機 能のいくつかはコンピュータ340上で動作するオペレーティングシステム(OS)もし くはサードパーティのプログラム、又はコンピュータ340にインストールされる各種ツ ールキットのモジュールにより提供される。従って、このプログラムはこの実施の形態の システムを実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、 命令のうち、所望の結果が得られるように制御されたやり方で適切な機能又は「ツール」 を呼出す事により、上記した声道断面積関数推定システム100、又はその一部をなすパ ラメータ算出部118若しくは逆推定処理部120としての動作を実行する命令のみを含 んでいればよい。コンピュータシステム330の動作は周知であるので、ここでは繰返さ ない。

20

30

10

[0076]

< 変 形 >

なお、上記実施の形態では、MRI画像を用いて声道断面積関数の推定のためのパラメ ータの学習を行なった。しかし本発明はそのような実施の形態には限定されない。MRI 画像と同様の3D透過画像が得られ、それによって声道断面積が測定できるような画像が 得られるのであれば、どのような方法を用いてもよい。上記実施の形態では、静止透過画 像と透過動画像とはともにMRIを用いている。しかし両者が一致している必要はない。 静止透過画像と、透過動画像とを異なる手法で入手するようにしてもよい。さらに、上記 実施の形態では、MRI動画像114の撮影と、音声データ116の収録とは別々に、た だし同期音発生装置112を使用して同期させて行なった。これはMRIの撮影において は大きな騒音が発生するため、音声データを良好な状態で収録する事ができないという制 約によるものであった。従って、MRI以外の方法で、静かな環境で透過動画像が撮影で きるのであれば、同時に音声データ116の収録を行なっても差し支えない。また、MR Iを用いる場合でも、音声データ116の収録が良好にできるのであれば、撮影と音声の 収録とを同時に行なってもよい。

[0077]

又、上記実施の形態では、音声データに対しケプストラム分析を行ない、得られたケプ ストラム係数のうち第1次~第24次の係数の線形和で主成分の重みを近似している。こ ⁴⁰ れは、実験で24次までのケプストラム係数を用いるとよい結果が得られたためである。 しかし、話者によっては最適な次数が多少変わることはあり得るし、また、同一人物でも 、用いるケプストラム係数の次数を多少増減させてもよい事はいうまでもない。

【0078】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許 請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内で のすべての変更を含む。

【図面の簡単な説明】

10

20

30

【図1】本発明の一実施の形態に係る声道断面積関数推定システム100の機能的構成を 示すブロック図である。 【図2】声道断面積関数に関する第1の主成分のフレームごとの値と、対応する声道断面 積関数とを示す図である。 【図3】声道断面積関数に関する第2の主成分のフレームごとの値と、対応する声道断面 積関数とを示す図である。 【図4】声道断面積関数に関する第3の主成分のフレームごとの値と、対応する声道断面 積関数とを示す図である。 【図5】MRI動画像から得られた第1及び第2の主成分、及び推定された2種類の第1 及び第2の主成分の、フレームごとの値を示す図である。 【図6】発話者の連続音声から推定された声道断面積関数を、学習時に用いられたMRI 動画像から得られた声道断面積関数、及びその第1及び第2の主成分による表現とを示す 図である。 【図7】本発明の一実施の形態に係る声道断面積関数推定システム100、又はその一部 をなすパラメータ算出部118又は逆推定処理部120を実現するコンピュータシステム の外観図である。 【図8】図7に示すコンピュータシステムのブロック図である。 【符号の説明】 $\begin{bmatrix} 0 & 0 & 8 & 0 \end{bmatrix}$ 100 声道断面積関数推定システム 110 静止MRI画像 1 1 2 同期音発生装置 1 1 4 MRI動画像 116 音声データ 118 パラメータ算出部 120 逆推定処理部 140 PCA処理部 142,146,150,154 記憶部 144 動画像データ射影処理部 1 4 8 ケプストラム係数算出部 152 重回帰分析処理部 180 音声キャプチャ処理部 182 ケプストラム係数算出部 184 線形変換処理部 186 声道断面積関数推定部



推定値

20



10 声門からの距離 (cm)





フロントページの続き

(72)発明者 本多 清志 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官谷垣圭二

- (56)参考文献 竹本浩典 等,〔音声生成の情報処理機構の解明に向けて〕連続発話と遊離発話における日本語5
 母音の声道形状比較,日本音響学会研究発表会講演論文集,2004年 9月21日,Vol.2004
 ,秋季1,233-234
- (58)調査した分野(Int.CI., DB名)
 - A 6 1 B 5 / 1 1 A 6 1 B 5 / 0 5 5 A 6 1 B 5 / 1 0 7 G 1 0 L 1 1 / 0 0