

【特許請求の範囲】

【請求項 1】

発話音声の F 0 実測値時系列を、ベース成分と、フレーズ成分 F 0 時系列と、アクセント成分 F 0 時系列に分離する成分分離部と、

前記発話音声の各音声単位の始端および終端の時刻が記述された時刻情報付き音声記号ラベル系列と、これに対応する指令応答型モデルのパラメータの言語学的制約を記述したモデルパラメータ制約情報とを取得し、フレーズ指令、アクセント指令の各パラメータが取りうる値の範囲を、言語学的に妥当な範囲に制限するパラメータ制限値を設定するパラメータ制限値設定部と、

前記フレーズ成分 F 0 時系列とベース成分の和と、その推定値の誤差、および前記アクセント成分 F 0 時系列とその推定値の誤差が前記パラメータ制限値の範囲で最小化するように、指令応答型モデルのパラメータを推定するパラメータ推定部を含む

10

基本周波数モデルパラメータ推定装置。

【請求項 2】

請求項 1 に記載の基本周波数モデルパラメータ推定装置であって、

所定の区間における前記 F 0 実測値時系列の区間最小値とこれに対応する時刻とを、区間最小周波数系列として抽出する区間最小周波数系列抽出部を含み、

前記成分分離部は、

前記区間最小周波数系列の最小値に基づいて前記ベース成分を分離し、前記区間最小周波数系列に基づく区間最小周波数関数から前記ベース成分を差し引いたものを前記フレーズ成分 F 0 時系列とし、前記 F 0 実測値時系列から前記区間最小周波数関数を差し引いたものを前記アクセント成分 F 0 時系列とする

20

基本周波数モデルパラメータ推定装置。

【請求項 3】

請求項 1 または 2 に記載の基本周波数モデルパラメータ推定装置であって、

前記フレーズ成分 F 0 時系列を F_{0p} とし、フレーズ成分の変動を示す関数を G_p とし、前記モデルパラメータ制約情報から得られるフレーズ指令開始モーラの直前モーラの終了時刻から C 秒 ($0 < C < 0.333$) を差し引いた時刻を T_0 とし、 $F_{0p}(T_0 + C) = A_p \times G_p(C)$ を満たすフレーズ指令の大きさ A_p をパラメータ推定の初期値として求める初期パラメータ生成部を含む

30

基本周波数モデルパラメータ推定装置。

【請求項 4】

請求項 1 または 2 に記載の基本周波数モデルパラメータ推定装置であって、

前記時刻情報付き音声記号ラベル系列と、前記モデルパラメータ制約情報から得られるアクセント指令開始モーラの直前モーラの終了時刻を T_1 、アクセント指令終了モーラの直前モーラの終了時刻を T_2 とし、前記アクセント成分 F 0 時系列を F_{0a} とし、アクセント成分の変動を示す関数を G_a とし、

【数 2】

$$F_{0a}\left(\frac{T_1 + T_2}{2}\right) = A_a \times G_a\left(\frac{T_2 - T_1}{2}\right)$$

を満たすアクセント指令の大きさ A_a をパラメータ推定の初期値として求める初期パラメータ生成部を含む

基本周波数モデルパラメータ推定装置。

【請求項 5】

発話音声の F 0 実測値時系列を、ベース成分と、フレーズ成分 F 0 時系列と、アクセント成分 F 0 時系列に分離するステップと、

前記発話音声の各音声単位の始端および終端の時刻が記述された時刻情報付き音声記号ラベル系列と、これに対応する指令応答型モデルのパラメータの言語学的制約を記述した

モデルパラメータ制約情報とを取得し、フレーズ指令、アクセント指令の各パラメータが取りうる値の範囲を、言語学的に妥当な範囲に制限するパラメータ制限値を設定するステップと、

前記フレーズ成分 F 0 時系列とベース成分の和と、その推定値の誤差、および前記アクセント成分 F 0 時系列とその推定値の誤差が前記パラメータ制限値の範囲で最小化するように、指令応答型モデルのパラメータを推定するステップを含む 基本周波数モデルパラメータ推定方法。

【請求項 6】

コンピュータを請求項 1 から 4 の何れかに記載の基本周波数モデルパラメータ推定装置として機能させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、人間の声の基本周波数の変化を記述した指令応答型モデルのパラメータを推定する基本周波数モデルパラメータ推定装置、基本周波数モデルパラメータ推定方法、プログラムに関する。

【背景技術】

【0002】

人間の声の高さの変化、すなわち、基本周波数（以下 F 0 と略す）の変化は、声帯付近での声の生成過程をモデル化した指令応答型モデル、いわゆる藤崎モデルによってうまく記述できることが知られている（非特許文献 1）。

【0003】

非特許文献 1 のモデルでは、1) ベースラインの声の高さを表わすベース成分、2) イントネーション句単位での緩やかな F 0 変化を表現する役割を担っていると考えられるフレーズ成分、3) アクセント句単位での急峻な F 0 変化を表現する役割を担っていると考えられるアクセント成分、の 3 つの成分が重畳した結果として、観測された F 0 が生成され则认为る。

【0004】

ベース成分は固定値をパラメータとして用いる場合が多い。フレーズ成分はフレーズ指令という句の冒頭付近で生じさせるスパイクの高さと位置をパラメータとして、それらに応じて時間変化を生じさせる関数モデルから生成される。アクセント成分はアクセント指令というアクセント表出のための声の高さの立ち上がり立ち下りの大きさと時間軸上の位置をパラメータとして、それらに応じて時間変化を生じさせる関数モデルから生成される。

【0005】

これらのパラメータは言語学的な知見との対応があり、それらを高速かつ高精度に推定することができれば、音声に含まれる F 0 変化の理解にとって有用である。

【0006】

非特許文献 2 の方法では、値が仮設定された前述のパラメータを用いて前述の指令応答型モデルから出力された F 0 推定値と、読み込んだ F 0 実測値との 2 乗誤差が最小化されるように、前述のパラメータの値を随時更新する。非特許文献 2 の方法では、F 0 の推定値と実測値との誤差がしきい値を越えた時に新たなパラメータを導入する。この処理を音声の先頭から末尾に向かって繰り返すことで、F 0 実測値に合致した F 0 推定値を出力できるパラメータの数と値を決定する。

【0007】

非特許文献 3 の方法は、前述のパラメータが、隠れマルコフモデルから生成されると仮定する。非特許文献 3 の方法では、隠れマルコフモデルが指令応答型モデルの各パラメータを正しく生成できるように、言語音声に対して人手で正確に各パラメータの時刻と大きさを付与した大量の学習データを用いて、隠れマルコフモデルの確率値の調整が行われる。学習データの中では正しい順序と間隔で指令応答型モデルのパラメータが並んでいるた

10

20

30

40

50

め、正しく学習されれば、言語学的にも妥当なパラメータを生成する隠れマルコフモデルとなる。

【0008】

非特許文献4の方法では、最初に、時系列データであるF0実測値をフーリエ変換によって、フレーズ成分とアクセント成分に分離する。具体的には、F0実測値時系列に対しフーリエ変換を用い、ローパスフィルタをかけてF0実測値の大まかな変化を作り出す成分のみを通過させ、それらから再構成された系列をフレーズ成分、一方の遮断された成分から再構成された系列をアクセント成分とする。続いて、山登り法を用いて、各成分の2乗誤差を減らすことで、前述の指令応答型モデルのパラメータの最適化を実現する。これを発話を構成する句毎に実施する。なお、ローパスフィルタによる処理に関しては非特許文献5に説明がある。

10

【0009】

非特許文献4を含め、符号化された実データの時系列をフーリエ変換する際には計算量を減らすため高速フーリエ変換（非特許文献5）が用いられる。

【先行技術文献】

【非特許文献】

【0010】

【非特許文献1】 Hiroya Fujisaki and Keikichi Hirose, Analysis of Voice fundamental frequency contours for declarative sentences of Japanese, J.Acoust. Soc. Jpn. (E) 5, 4 (1984).

20

【非特許文献2】 Edouard Geoffrois, A pitch contour analysis guided by prosodic event detection, Proc. of Eurospeech, pp.793-796, 1993.

【非特許文献3】 亀岡弘和、ルルージョナトン、大石康智、「音声F0パターン生成過程の確率モデル」、日本音響学会秋季研究発表会講演論文集、1-1-3、pp.207-210、2010

【非特許文献4】 Hartmut R. Pfitzinger et al., Comparison of Fujisaki-Model Extractors and F0 Stylizers, Proc. of Interspeech, pp.2455-2458, 2009.

【非特許文献5】 Cooley, James W., and John W. Tukey. "An algorithm for the machine calculation of complex Fourier series." Mathematics of computation 19.90 (1965): 297-301.

【発明の概要】

30

【発明が解決しようとする課題】

【0011】

非特許文献2の方法では、F0の推定値と実測値との誤差削減のみに基づいてパラメータを最適化し、誤差の大きさに基づいてパラメータの導入、つまり、パラメータの数を決定している。そのため、アクセント持続時間が負の値になるような言語学的に不適切なパラメータの調節結果が得られるという課題や、アクセント句数やイントネーションフレーズ数といった発話された文言の言語学的な情報と対応しない場合が生じるという課題がある。

【0012】

そこで本発明では、音声情報だけでなく言語情報を利用したパラメータ推定ができる基本周波数モデルパラメータ推定装置を提供することを目的とする。

40

【課題を解決するための手段】

【0013】

本発明の基本周波数モデルパラメータ推定装置は、成分分離部と、パラメータ制限値設定部と、パラメータ推定部を含む。

【0014】

成分分離部は、発話音声のF0実測値時系列を、ベース成分と、フレーズ成分F0時系列と、アクセント成分F0時系列に分離する。パラメータ制限値設定部は、発話音声の各音声単位の始端および終端の時刻が記述された時刻情報付き音声記号ラベル系列と、これに対応する指令応答型モデルのパラメータの言語学的制約を記述したモデルパラメータ制

50

約情報とを取得し、フレーズ指令、アクセント指令の各パラメータが取りうる値の範囲を、言語学的に妥当な範囲に制限するパラメータ制限値を設定する。パラメータ推定部は、フレーズ成分 F 0 時系列とベース成分の和と、その推定値の誤差、およびアクセント成分 F 0 時系列とその推定値の誤差がパラメータ制限値の範囲で最小化するように、指令応答型モデルのパラメータを推定する。

【発明の効果】

【0015】

本発明の基本周波数モデルパラメータ推定装置によれば、音声情報だけでなく言語情報を利用したパラメータ推定ができる。

【図面の簡単な説明】

10

【0016】

【図1】実施例1の基本周波数モデルパラメータ推定装置の構成を示すブロック図。

【図2】時刻情報付き音声記号ラベル系列の例を示す図。

【図3】モデルパラメータ制約情報の例を示す図。

【図4】実施例1の基本周波数モデルパラメータ推定装置の動作を示すフローチャート。

【図5】区間最小周波数系列抽出部と成分分離部の処理結果の例を示す図であって、図5 (a) は、F 0 実測値時系列の例を、図5 (b) は、区間最小周波数系列の例を、図5 (c) は、区間最小周波数関数の例を、図5 (d) は、ベース成分の例を、図5 (e) は、フレーズ成分 F 0 時系列の例を、図5 (f) は、アクセント成分 F 0 時系列の例を、それぞれ示す。

20

【図6】変形例1の基本周波数モデルパラメータ推定装置の構成を示すブロック図。

【図7】変形例1の基本周波数モデルパラメータ推定装置の動作を示すフローチャート。

【図8】変形例2のモデルパラメータ制約情報の例を示す図。

【発明を実施するための形態】

【0017】

以下、本発明の実施の形態について、詳細に説明する。なお、同じ機能を有する構成部には同じ番号を付し、重複説明を省略する。

【実施例1】

【0018】

以下、図1を参照して、実施例1の基本周波数モデルパラメータ推定装置の構成について説明する。同図に示すように、本実施例の基本周波数モデルパラメータ推定装置1は、時刻情報付き音声記号ラベル系列記憶部10aと、F 0 実測値時系列記憶部10bと、モデルパラメータ制約情報記憶部10cと、区間最小周波数系列抽出部11と、成分分離部12と、初期パラメータ生成部13と、パラメータ制限値設定部14と、パラメータ推定部15を含む。

30

【0019】

<基本周波数モデルパラメータ推定装置1の入出力>

本実施例の基本周波数モデルパラメータ推定装置1は、1) 発話音声と発話音声の書き起こしから得られた音声記号などの各音声単位の発話音声ファイル内での始端の時刻および終端の時刻が記述された時刻情報付き音声記号ラベル系列と、2) 発話音声のF 0 実測値時系列と、3) 指令応答型モデルのパラメータである、フレーズ指令、アクセント指令の当該発話音声内での個数、および、それらがどの音声記号の近くに生起するかを、言語学的制約に基づいて記述したモデルパラメータ制約情報を入力とする。

40

【0020】

図2に時刻情報付き音声記号ラベル系列の例を示す。また、図3にモデルパラメータ制約情報の例を示す。図3の枠内に示す記号付きのテキストは、発話内容「あと一時間で雨が降るらしい」に対応するモデルパラメータ制約情報である。同図の例では、句の発話内容をひらがなで記述するものとする。ここで、記号 "<" から記号 ">" の間にアクセント指令が1つ生成されるという制約を付す。また、記号 "■" の箇所にフレーズ指令が1つ生成されるという制約を付す。1型アクセントの場合、先頭から全て、記号 "<"、記

50

号" > "で囲む（例えば"あと"や"ふる"など）。一方、それ以外のアクセント型の場合は、2モーラ目から全てを記号" < "、記号" > "で囲む（例えば"いちじかんで"や"らしい"など）。

【0021】

本実施例の基本周波数モデルパラメータ推定装置1は前述の3つのデータを読み込み、それぞれの記憶部（時刻情報付き音声記号ラベル系列記憶部10a、F0実測値時系列記憶部10b、モデルパラメータ制約情報記憶部10c）に格納する。

【0022】

本実施例の基本周波数モデルパラメータ推定装置1は発話音声に対する指令応答型モデルのパラメータである、ベース成分の大きさ、フレーズ指令の生起時刻と大きさ、アクセント指令の生起時刻と大きさを、当該音声に対して必要な数だけ出力する。

10

【0023】

<基本周波数モデルパラメータ推定装置1の動作の詳細>

以下、図4を参照して、本実施例の基本周波数モデルパラメータ推定装置1の動作を各ステップに分けて説明する。

【0024】

<ステップS11>

まず、区間最小周波数系列抽出部11は、 $t = 0$ から発話終了までF0実測値時系列を1フレームずつ走査し、各フレームのサンプリング時刻 $t = s$ において、当該時刻の周波数が区間 $[0, s]$ において最小となる、即ち

20

$$F_0(s) = \min\{F_0(t) | t \in [0, s]\}$$

となる場合の $F_0(s)$ を区間最小値、 s を区間最小値発生時刻として抽出する（S11）。区間最小値 $F_0(s)$ とこれに対応する区間最小値発生時刻 s からなる区間最小周波数系列の例を図5（b）に□記号で示す。

【0025】

<ステップS12>

成分分離部12は、F0実測値時系列を、ベース成分と、フレーズ成分F0時系列と、アクセント成分F0時系列に分離する（S12）。具体的には、成分分離部12は、ステップS11で抽出された区間最小値 $F_0(s)$ と、区間最小値発生時刻 s とのペアから成る区間最小周波数系列を用いる。成分分離部12は、区間最小周波数系列の各座標を補間し、区間最小周波数関数を得る。区間最小周波数関数の例を図5（c）に破線で示す。なお、補間は直線に限定しない。成分分離部12は、区間最小周波数系列、または区間最小周波数関数の最小値を求め、これをベース成分とする。図5（d）にベース成分の例を細破線で示す。次に、成分分離部12は、区間最小周波数関数からベース成分を差し引いたものをフレーズ成分F0時系列として求める。図5（e）に△記号でフレーズ成分F0時系列の例を示す。さらに、成分分離部12は、F0実測値時系列から区間最小周波数関数を差し引いたものをアクセント成分F0時系列とする。なお、成分分離部12は、F0実測値時系列からベース成分およびフレーズ成分F0時系列を差し引いて、アクセント成分F0時系列としてもよい。図5（f）に○記号でアクセント成分F0時系列の例を示す。成分分離部12は、このようにして成分の分離を行なう。

30

40

【0026】

<ステップS13>

次に、初期パラメータ生成部13は、指令応答型モデルのパラメータの初期値を生成する（S13）。まず、初期パラメータ生成部13は、ステップS12で得られた区間最小周波数系列の最小値の自然対数を、指令応答型モデルのパラメータの1つであるベース成分 $\ln F_{min}$ の初期値とする。続いて、初期パラメータ生成部13は、時刻情報付き音声記号ラベル系列とモデルパラメータ制約情報に基づき、フレーズ指令、および、アクセント指令のパラメータの初期値を生成する。

【0027】

より具体的には、初期パラメータ生成部13は、以下のようにパラメータの初期値を生

50

成する。まず、初期パラメータ生成部 13 は、 T_0 として、時刻情報付き音声記号ラベル系列とモデルパラメータ制約情報から得られる指令の生起範囲指定データで指定されたフレーズ指令開始モーラの直前モーラの終了時刻から一定の時間（例えば 0.15 秒、後述）を差し引いた時刻を設定する。

【0028】

＜フレーズ指令開始モーラ、その直前モーラ＞

図 3 における "■＜あと＞い＜ちじかんで＞■＜あめが＞＜ふる＞ら＜しい＞" の場合、1 つ目のフレーズ "あと いちじかんで" では、フレーズ指令開始モーラが "あ"（の区間）、フレーズ指令開始モーラの直前モーラが発話前無音区間である *sil*（の区間）である。2 つ目のフレーズ "あめがふるらしい" では、フレーズ指令開始モーラが "あ"（の区間）、フレーズ指令開始モーラの直前モーラが "で"（の区間）となる。ただし、ここでは便宜上、無音区間 *sil* も *a* や *to* と同様のモーラとして扱う。

10

【0029】

初期パラメータ生成部 13 は、 T_0 を初期値としてフレーズ成分実測値と推定値が一致するように次式を満たすようなフレーズ指令の大きさ A_P を設定する。なお、 G_P は非特許文献 5 にあるフレーズ成分の変動を示す関数である。

$$F_{ij}(T_i+0.15)=A_i \times G_i(0.15)$$

なお、 F_{op} は F_0 中のフレーズ成分のみ（フレーズ成分 F_0 時系列、例えば、図 5（e）の△記号参照）を意味する。初期パラメータ生成部 13 は、この式が成り立つよう、すなわち、時刻 $T_0+0.15$ での F_{op} の値である $F_{op}(T_0+0.15)$ と $A_P \times G_P(0.15)$ が等しくなるように A_P を定め、これをフレーズ指令の大きさ A_P の初期値とする。

20

【0030】

ここで、0.15 秒はフレーズ成分の変動を示す関数 G_P 値の変動と F_0 実測値の変動とがうまく一致するための時間軸上での位置合わせとして寄与する値であって、過去の文献や慣例から定めた値である。具体的には、フレーズ成分の変動を示す関数 $G_P(t)$ を時間 t で偏微分して 0 とおくことで G_P がどの時刻で最大値を取るかが解析的に定まり、この時刻は図 5 の時間軸（横軸）で $t=1/\alpha$ 、およそ $1/3 (=0.333)$ である。 $t=T_0$ すなわち図 5 のカーブの始まる左端の立ち上がりの位置（時刻 0）においては、ほぼ F_{op} も G_P も値がゼロとなるため、フレーズ指令の大きさ A_P の初期値推定に不向きである。さらに最大値を取る時刻 $t=0.333$ の場合、 T_0 推定値がずれていた場合（実測値よりも大きくても小さくても） G_P を過小評価してしまい A_P の推定結果に誤差が入りやすくなる。そこで、0 と 0.333 の間の値として 0.15 を選択した。例えば、

30

$$F_{ij}(T_i+C)=A_i \times G_i(C)$$

として、前述の式を一般化し、 C を $0 < C < 0.333$ を満たすものとしてもよい。

【0031】

続いて、初期パラメータ生成部 13 は、 T_1 として、時刻情報付き音声記号ラベル系列と、モデルパラメータ制約情報から得られる指令の生起範囲指定データで指定されたアクセント指令開始モーラの直前モーラの終了時刻を設定する。さらに、 T_2 として、同データで指定されたアクセント指令終了モーラの直前モーラの終了時刻を設定する。

40

【0032】

＜アクセント指令開始モーラと直前モーラ、アクセント指令終了モーラと直前モーラ＞

例えば、図 3 の文「あと一時間で雨が降るらしい」の "いちじかんで" の場合、その制約情報は "い＜ちじかんで＞" となっており、アクセント開始モーラが "ち"、アクセント開始モーラの直前モーラが "い"、アクセント終了モーラが "あ"、アクセント終了モーラの直前モーラが "で" となる。また、"あめが" の場合、その制約情報は "＜あめが＞" となっており、アクセント開始モーラが "あ"、アクセント開始モーラの直前モーラが "で"、アクセント終了モーラが "ふ"、アクセント終了モーラの直前モーラが "が" となる。

50

【0033】

初期パラメータ生成部13は、前述の T_1 と T_2 とを用いて、アクセント成分実測値と推定値とが一致するよう次式を満たすようなアクセント指令の大きさ A_a をパラメータ推定の初期値として求める。なお、 G_a は非特許文献5にあるアクセント成分の変動を示す関数である。なお、 F_{0a} は F_0 中のアクセント成分のみ（アクセント成分 F_0 時系列、例えば、図5（f）の○記号参照）を意味する。

【0034】

【数1】

$$F_{0a}\left(\frac{T_1+T_2}{2}\right)=A_a \times G_a\left(\frac{T_2-T_1}{2}\right)$$

【0035】

$(T_1+T_2)/2$ の時刻で実測値と G_a の値を合わせる理由は、その時刻が A_a の初期値推定に際しノイズが入りにくい箇所であるという理由による。理論上、アクセント成分 F_0 時系列は $t=T_1$ のとき0で、 $t=T_2$ のとき最大値を取る。 $F_{0a}(t)=A_a \times G_a(t)$ を用いて A_a を求めるとき、 $t=T_1$ であれば、 $F_{0a}(t)$ が小さいため A_a の推定精度が下がり、 $t=T_2$ であれば、 T_2 推定値の初期値が T_2 の真の値より大きい場合、 $F_{0a}(t)$ の最大値を過小評価してしまい A_a の推定精度が下がるという問題が生ずる。よって T_1 と T_2 の中間の値を推定に用いれば好適である。

【0036】

<ステップS14>

次に、パラメータ制限値設定部14は、時刻情報付き音声記号ラベル系列記憶部10aに記憶された時刻情報付き音声記号ラベルと、モデルパラメータ制約情報格納部10cに格納されたモデルパラメータ制約情報とを取得し、これらを利用して、フレーズ指令、アクセント指令の各パラメータが取りうる値の範囲を、言語学的に妥当な範囲に制限できるようにパラメータ制限値を設定する（ステップS14）。ここで、言語学的に妥当な範囲とは、指令応答型モデルのパラメータ（フレーズ指令およびアクセント指令）が置かれるべき妥当な位置の範囲のことである。この範囲外となると、言語学的に意味のないパラメータとなり、例えば、アクセント持続時間が負の値になるような言語学的に不適切なパラメータが得られたり、アクセント句数やイントネーション句数の点で発話された文言の言語学的な情報と対応しない場合が生じる。

【0037】

<ステップS15>

パラメータ推定部15は、フレーズ成分 F_0 時系列とその推定値の誤差、およびアクセント成分 F_0 時系列とその推定値の誤差がパラメータ制限値の範囲で最小化するように、指令応答型モデルのパラメータを推定する（ステップS15）。パラメータ推定部15は、フレーズ成分 F_0 推定値（指令応答型モデルのフレーズ成分を決める G_p 関数によって計算された値の系列）と $\ln F_{min}$ の和と、ステップS12で得られたフレーズ成分 F_0 時系列とベース成分の和、の2乗誤差をステップS14の範囲で最小化するように、パラメータ $\ln F_{min}$ 、 T_0 、 A_p を最小2乗法で調整する。パラメータ推定部15は、アクセント成分 F_0 推定値とステップS12で得られたアクセント成分 F_0 時系列の2乗誤差をステップS14の範囲で最小化するように、パラメータ T_1 、 T_2 、 A_a を最小2乗法で調整する。

【0038】

上記の演算では F_0 の値は自然対数に変換された値を用いる。ステップS12の成分分離においては、最初のアクセント句が頭高（1型）である場合、アルゴリズム上アクセント成分をフレーズ成分と誤認識するおそれがある。そこで最初のアクセント句が頭高の場合のみ例外処理として、フレーズ成分推定には開始後0.2秒までの区間 F_0 を無視した $\min\{F_0(t)|t \in [0.2, s]\}$ を利用する。 s は発話末尾の時刻である。また、ステップS14では、例えば、以下の閾値を使用することが可能である。

20

30

40

50

$T_i: [t'_i-0.3, t'_i-0.05]$ (句末のみ $[t'_i-0.05, t'_i+0.05]$)

$A_i: [0, 1.0]$

$T_i: [u-1/6(t'_i-u), u+1/6(t'_i-u)]$ (アクセント型が 1 型なら $u=t'_i$, その他なら $u=t'_i$)

$T_i: [u+1/6(t'_i-u), t'_i]$

$\ln F_{min}: [0, 6.0] (= [1, 403] (\text{Hz}))$

【0039】

なお、 t'_i はフレーズ内の i 番目モーラの開始時刻を意味する。例えば図 3 の " あ と " の場合、 t'_1 は 1 番目のモーラ " あ " の発話開始時刻、 t'_2 は 2 番目のモーラ " と " の発話開始時刻である。 u は任意の時刻である。 n はフレーズ内の最終モーラの番号である。式に出てくる $1/6$ 倍などの係数は経験的に得られたものを用いている。

10

【0040】

[変形例 1]

実施例 1 において、ベース成分 $\ln F_{min}$ は音声毎に求めたが、これを話者が同じであるとか、発話様式が同じであるなどの同一条件の場合に共有することが可能である。以下、変形例 1 の基本周波数モデルパラメータ推定装置 1 a について説明する。図 6 に示すように、本変形例の基本周波数モデルパラメータ推定装置 1 a は、実施例 1 と共通して、時刻情報付き音声記号ラベル系列記憶部 10 a と、 F_0 実測値時系列記憶部 10 b と、モデルパラメータ制約情報記憶部 10 c と、区間最小周波数系列抽出部 11 と、成分分離部 12 と、初期パラメータ生成部 13 と、パラメータ制限値設定部 14 と、パラメータ推定部 15 を含み、さらに、グループ分け部 16 と、グループ内平均値取得部 17 を含む。

20

【0041】

まず、変形例 1 の周波数モデルパラメータ推定装置 1 a は、前述の実施例 1 の処理（ステップ S11～S15）を実行する。

【0042】

続いて、周波数モデルパラメータ推定装置 1 a のグループ分け部 16 は、得られたパラメータデータを同一条件のもの（話者が同じ、発話様式が同じ等）ごとに分類しグループ分けする（S16）。グループ内平均値取得部 17 は、ステップ S16 でグループ分けされた各グループ内の各データにおいて $\ln F_{min}$ パラメータの平均値を取得する（S17）。本変形例の周波数モデルパラメータ推定装置 1 a は、前述の平均値を $\ln F_{min}$ パラメータとし、各データに対し再び実施例 1 の処理（ステップ S11～S15）を実行する。

30

【0043】

[変形例 2]

実施例 1 のステップ S11 では区間最小値の抽出範囲は発話全体であったが、これを基本周波数の上下変動が激しいと予想される区間などの一部区間を発話全体から除いた範囲に置き換えることもできる。

【0044】

この際、上下変動が激しいと予想される区間に対応する指令が存在する場合、ステップ S14 で用いる閾値を、パラメータごとに他の指令と別途に設定することが可能である。なお、閾値を別途に設定しなかったパラメータに関しては共通の閾値を参照する。基本周波数の上下変動が激しいと予想される区間にフレーズ指令やアクセント指令が対応する場合、例えば、ステップ S14 で以下の閾値を他のパラメータと別途に使用することが可能である。

40

$A_i: [-1.0, 1.0]$

$A_i: [-1.0, 1.0]$

【0045】

前述の基本周波数の上下が激しいと予想される区間は、前述のモデルパラメータ制約情報のファイルで指定する。図 8 に例を示す。図 8 の ! と ! とで囲まれた区間を F_0 変化が激しいと予想される区間と定義し、その区間を最小周波数の検査対象外とする。

【0046】

50

また、このとき、！と！で囲まれたモーラ区間内に存在するアクセント指令（＜＞で囲まれた箇所）やフレーズ指令（■）を「上下が激しいと予想される区間に対応する指令」として扱う。

【0047】

図8の事例の場合、末尾の"よ"の発話開始時刻から終了時刻までを最小周波数の検査対象外とする。"よ"に対応するアクセント指令は前述のステップS14の閾値を別途指定することが可能である。なお、図3、図8ともに、用いる記号は、図3や図8で用いられた記号に限るものではない。上記枠内の記載方法は図3と同じであるので差分のみ説明する。発話内容「雨が降るらしいよ」を例として対応する制約情報を示した。"! "で囲まれた区間が前述の基本周波数の上下が激しいと予想される区間に対応する。

10

【0048】

（実施例、変形例にかかる基本周波数モデルパラメータ推定装置の効果）

モデルパラメータ制約情報に記述されたパラメータの生起範囲指定データで言語学的制約をかける点から、非特許文献2で課題となったような言語学的に意味のないパラメータが抽出される可能性が大幅に減るという効果がある。仮に抽出された場合、言語理論上生じると仮定されるアクセントが実際の音声では生じなかったことを意味しているため、データに則した実際的なパラメータになっていることが期待できる。

【0049】

また非特許文献3は、隠れマルコフモデルという統計モデルを用いるため、学習用データが大量に必要になるという課題があった。また、非特許文献3のような統計モデルによる推定は、パラメータを推定する初見のデータが学習データと同分布に従うという前提に基づいているため、両データが従う分布が異なる場合に推定が正しく行われるかどうかは不明であるという課題があった。

20

【0050】

実施例、変形例にかかる基本周波数モデルパラメータ推定装置によれば、推定する音声データのみからパラメータ推定する点から、学習データが不要である点で、非特許文献3の課題も生じない。

【0051】

また非特許文献4を含め、符号化された実データの時系列をフーリエ変換する際には計算量を減らすため高速フーリエ変換が用いられるが、この際、フーリエ変換の積分計算を総和で近似しているため、周波数 $n / (\text{時系列の時間長})$ （ n は整数）の成分しか抽出することができない。結果として、 $1 / (\text{時系列の時間長})$ 以下の周波数を抽出または分離することは不可能である。

30

【0052】

そのため、短い句ではローパスフィルタによりフレーズ成分を十分に抽出できないという問題点がある。会話音声にはさまざまな短い句が出現するが、そのような語の分析に用いることが困難であるという課題があった。

【0053】

実施例、変形例にかかる基本周波数モデルパラメータ推定装置によれば、フレーズ成分、アクセント成分の分離にF0形状の包絡線を利用している点から、対話音声に頻出する短い音声に対してもフレーズ、アクセント成分を分離可能となり正確なパラメータ推定が可能となる。よって非特許文献4の課題も解決される。

40

【0054】

（実施例、変形例にかかる基本周波数モデルパラメータ推定装置のポイント）

時刻情報付き音声記号ラベル系列格納部10aやモデルパラメータ制約情報格納部10cによって、言語情報からパラメータの個数や各パラメータ値が取り得る範囲を予め規定することで、音声情報だけでなく言語情報を利用したパラメータ推定が可能となる。

【0055】

<補記>

本発明の装置は、例えば単一のハードウェアエンティティとして、キーボードなどが接

50

続可能な入力部、液晶ディスプレイなどが接続可能な出力部、ハードウェアエンティティの外部に通信可能な通信装置（例えば通信ケーブル）が接続可能な通信部、CPU（Central Processing Unit、キャッシュメモリやレジスタなどを備えていてもよい）、メモリであるRAMやROM、ハードディスクである外部記憶装置並びにこれらの入力部、出力部、通信部、CPU、RAM、ROM、外部記憶装置の間のデータのやり取りが可能なように接続するバスを有している。また必要に応じて、ハードウェアエンティティに、CD-ROMなどの記録媒体を読み書きできる装置（ドライブ）などを設けることとしてもよい。このようなハードウェア資源を備えた物理的実体としては、汎用コンピュータなどがある。

【0056】

10

ハードウェアエンティティの外部記憶装置には、上述の機能を実現するために必要となるプログラムおよびこのプログラムの処理において必要となるデータなどが記憶されている（外部記憶装置に限らず、例えばプログラムを読み出し専用記憶装置であるROMに記憶しておくこととしてもよい）。また、これらのプログラムの処理によって得られるデータなどは、RAMや外部記憶装置などに適宜に記憶される。

【0057】

ハードウェアエンティティでは、外部記憶装置（あるいはROMなど）に記憶された各プログラムとこの各プログラムの処理に必要なデータが必要に応じてメモリに読み込まれて、適宜にCPUで解釈実行・処理される。その結果、CPUが所定の機能（上記、…部、…手段などと表した各構成要件）を実現する。

20

【0058】

本発明は上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。また、上記実施形態において説明した処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されるとしてもよい。

【0059】

既述のように、上記実施形態において説明したハードウェアエンティティ（本発明の装置）における処理機能をコンピュータによって実現する場合、ハードウェアエンティティが有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記ハードウェアエンティティにおける処理機能がコンピュータ上で実現される。

30

【0060】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD（Digital Versatile Disc）、DVD-RAM（Random Access Memory）、CD-ROM（Compact Disc Read Only Memory）、CD-R（Recordable）／RW（ReWritable）等を、光磁気記録媒体として、MO（Magneto-Optical disc）等を、半導体メモリとしてEEPROM（Electrically Erasable and Programmable-Read Only Memory）等を用いることができる。

40

【0061】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【0062】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自

50

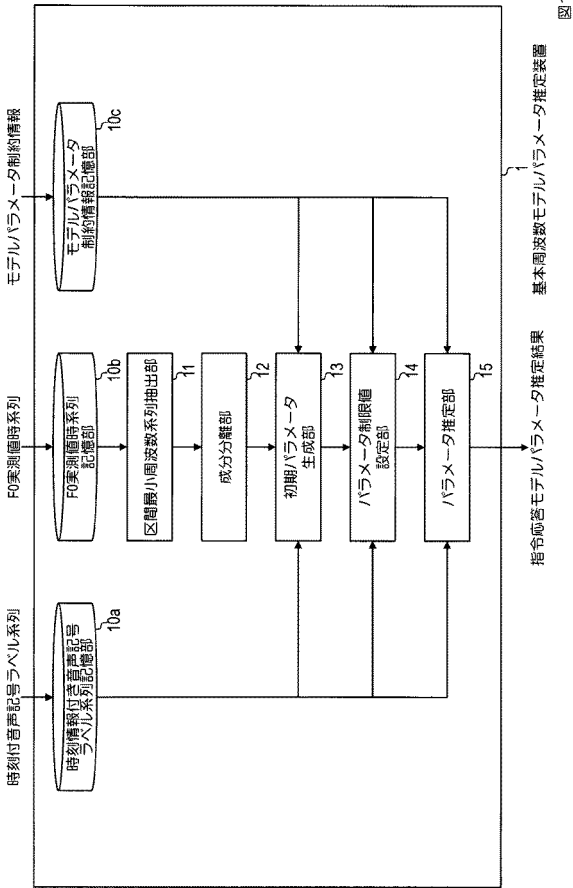
己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP（Application Service Provider）型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

10

【0063】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、ハードウェアエンティティを構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

【図1】



【図2】

モーラ音素記号ラベル (音素単位)	始測時刻[s]	終測時刻[s]
<sil>	0.000	0.500
a	0.500	0.625
to	0.625	0.765
i	0.765	0.915
chi	0.915	1.005
...	...	...

図2

【図 3】

■<あと>い<ちじかんで>■<あめが><ふる>ら<しい>

図3

【図 4】

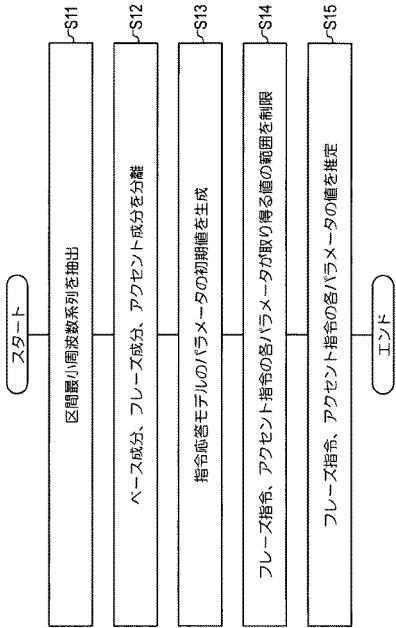


図4

【図 5】

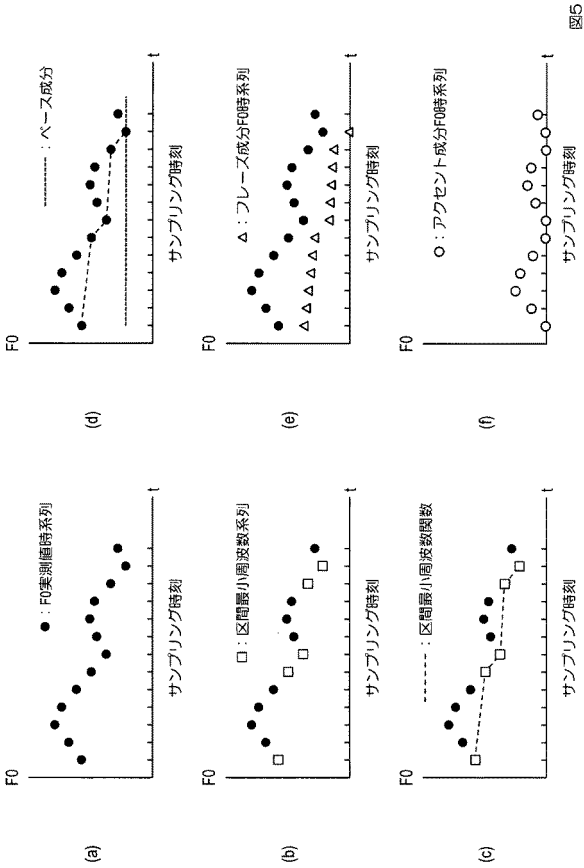


図5

【図 6】

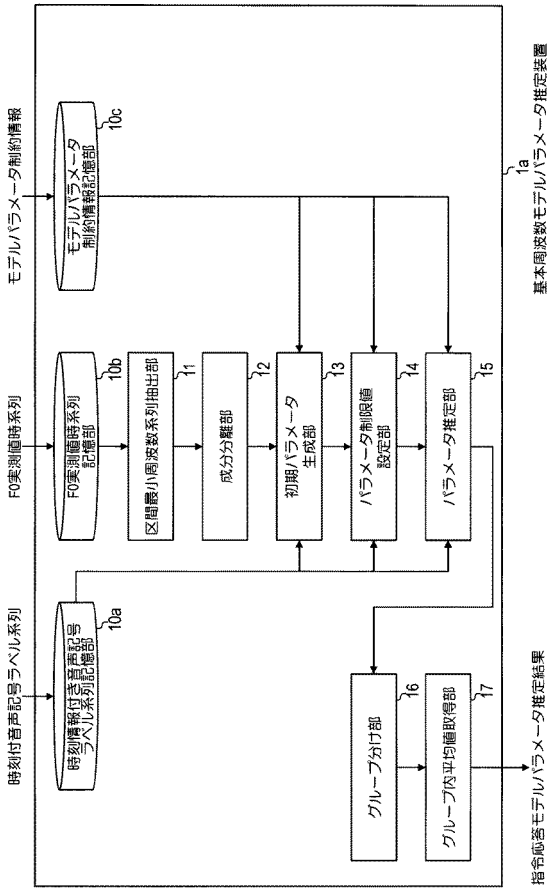


図6

【図 7】

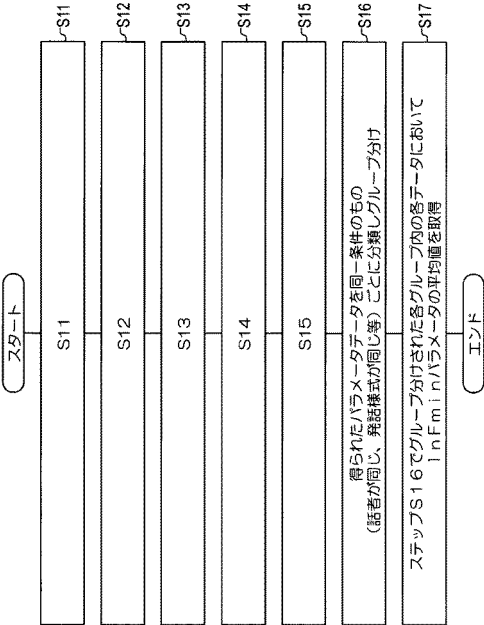


図 7

【図 8】

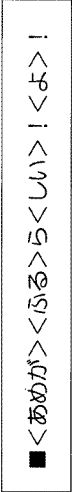


図 8

フロントページの続き

(72)発明者 高田 一真

東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内

(72)発明者 匂坂 芳典

東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内