(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6468519号
(P6468519)

(45)発行日　平成31年2月13日(2019.2.13)　(24)登録日　平成31年1月25日(2019.1.25)

(51)Int.Cl.　　F I
G10L 25/90 (2013.01)　　G10L 25/90
G10L 13/10 (2013.01)　　G10L 13/10　111C
　　G10L 13/10　113Z

請求項の数 8　(全 39 頁)

(21)出願番号　特願2016-32412(P2016-32412)
(22)出願日　平成28年2月23日(2016.2.23)
(65)公開番号　特開2017-151224(P2017-151224A)
(43)公開日　平成29年8月31日(2017.8.31)
審査請求日　平成29年12月8日(2017.12.8)

(73)特許権者　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(73)特許権者　504143441
国立大学法人　奈良先端科学技術大学院大学
奈良県生駒市高山町８９１６－５
(74)代理人　110001519
特許業務法人太陽国際特許事務所
(72)発明者　亀岡　弘和
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　田中　宏
奈良県生駒市高山町８９１６－５　国立大学法人奈良先端科学技術大学院大学内
最終頁に続く

(54)【発明の名称】基本周波数パターン予測装置、方法、及びプログラム

(57)【特許請求の範囲】
【請求項１】
学習サンプルのソース音声の時系列データとターゲット音声の時系列データとからなるパラレルデータを入力として、前記ソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の時系列データから抽出される、各時刻の基本周波数とに基づいて、前記ソース音声の各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の各時刻の基本周波数との間の関係をモデル化した第１確率分布のパラメータを学習する第１モデルパラメータ学習部と、
前記ターゲット音声の各時刻の基本周波数に基づいて、基本周波数パターン生成過程をモデル化した第２確率分布のパラメータを学習する第２モデルパラメータ学習部と、
予測対象のソース音声の時系列データを入力として、前記予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記第１モデルパラメータ学習部によって学習された前記第１確率分布のパラメータと、前記第２モデルパラメータ学習部によって学習された前記第２確率分布のパラメータとに基づいて、前記第１確率分布と前記第２確率分布とを用いて表される規準を大きくするように、前記予測対象のソース音声に対応する前記ターゲット音声の各時刻の基本周波数を予測する基本周波数予測部と、
を含む基本周波数パターン予測装置。
【請求項２】
前記第１確率分布を、前記ソース音声の各時刻のスペクトル特徴量ベクトルと前記ター

ゲット音声の各時刻の基本周波数と前記基本周波数の動的成分との同時確率分布を表す混合正規分布とし，
前記第２確率分布を、
各時刻の基本周波数と、
隠れマルコフモデルの各時刻の状態からなる状態系列、又は各時刻における甲状軟骨の平行移動運動によって生じる基本周波数パターンを表すフレーズ指令及び甲状軟骨の回転運動によって生じる基本周波数パターンを表すアクセント指令のペアからなる指令関数との組み合わせの確率分布とした請求項１記載の基本周波数パターン予測装置。
【請求項３】
前記規準を、
前記第１確率分布と前記第２確率分布との積を用いて表される、各時刻の基本周波数と、前記状態系列との組み合わせに応じた関数、または
前記第１確率分布と前記第２確率分布との積を用いて表される、各時刻の基本周波数と、各時刻の前記指令関数との組み合わせに応じた関数とした請求項２記載の基本周波数パターン予測装置。
【請求項４】
前記第１モデルパラメータ学習部は、ＥＭ（Expectation-Maximization）アルゴリズムにより、前記第１確率分布から求められる前記ソース音声の各時刻のスペクトル特徴量ベクトル及び前記ターゲット音声の各時刻の基本周波数の尤もらしさが大きくなるように、前記第１確率分布のパラメータを学習する請求項１～請求項３の何れか１項記載の基本周波数パターン予測装置。
【請求項５】
前記第２モデルパラメータ学習部は、ＥＭ（Expectation-Maximization）アルゴリズムにより、前記第２確率分布から求められる、各時刻の基本周波数と、隠れマルコフモデルの各時刻の状態からなる状態系列との尤もらしさが大きくなるように、前記第２確率分布のパラメータとして、前記状態系列における状態遷移確率、及び各時刻における状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータ群を学習するか、又は
各時刻の基本周波数が与えられたときの、各時刻のフレーズ指令及びアクセント指令のペアからなる指令関数及び前記パラメータ群の対数事後確率を目的関数として、前記目的関数を増加させるように、前記指令関数及び前記パラメータ群を、前記第２確率分布のパラメータとして学習する請求項１～請求項４の何れか１項記載の基本周波数パターン予測装置。
【請求項６】
前記基本周波数予測部は、
前記第１確率分布と前記第２確率分布との積を用いて表される、各時刻の基本周波数と、隠れマルコフモデルの各時刻の状態からなる状態系列との組み合わせに応じた関数が大きくなるように、各時刻の基本周波数と前記状態系列とを推定することにより、前記予測対象のソース音声に対応する前記ターゲット音声の各時刻の基本周波数を予測するか、又は
前記第１確率分布と前記第２確率分布との積を用いて表される、各時刻の基本周波数、各時刻のフレーズ指令及びアクセント指令のペアからなる指令関数との組み合わせに応じた関数が大きくなるように、各時刻の基本周波数と、各時刻の前記指令関数を推定することにより、前記予測対象のソース音声に対応する前記ターゲット音声の各時刻の基本周波数を予測する請求項１～請求項５の何れか１項記載の基本周波数パターン予測装置。
【請求項７】
第１モデルパラメータ学習部と、第２モデルパラメータ学習部と、基本周波数予測部とを含む基本周波数パターン予測装置における基本周波数パターン予測方法であって、
前記第１モデルパラメータ学習部が、学習サンプルのソース音声の時系列データとターゲット音声の時系列データとからなるパラレルデータを入力として、前記ソース音声の時

系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の時系列データから抽出される、各時刻の基本周波数とに基づいて、前記ソース音声の各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の各時刻の基本周波数との間の関係をモデル化した第１確率分布のパラメータを学習し、

前記第２モデルパラメータ学習部が、前記ターゲット音声の各時刻の基本周波数に基づいて、基本周波数パターン生成過程をモデル化した第２確率分布のパラメータを学習し、

前記基本周波数予測部が、予測対象のソース音声の時系列データを入力として、前記予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記第１モデルパラメータ学習部によって学習された前記第１確率分布のパラメータと、前記第２モデルパラメータ学習部によって学習された前記第２確率分布のパラメータとに基づいて、前記第１確率分布と前記第２確率分布とを用いて表される規準を大きくするように、前記予測対象のソース音声に対応する前記ターゲット音声の各時刻の基本周波数を予測する

基本周波数パターン予測方法。

【請求項８】

請求項１～請求項６の何れか１項に記載の基本周波数パターン予測装置の各部としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、基本周波数パターン予測装置、方法、及びプログラムに係り、特に、ソース音声から、ターゲット音声の基本周波数パターンを予測する基本周波数パターン予測装置、方法、及びプログラムに関する。

【背景技術】

【０００２】

他者とのコミュニケーションにおいて音声は利便性に優れた手段ではあるが、時として物理的制約により様々な障壁が必然的にもたらされる。例えば、発声器官の内、わずか一か所でも正常に動作しなくなると、深刻な発声障害を患い、音声コミュニケーションに支障をきたす。また、音声生成という物理的行為は、秘匿性の高い意思伝達には不向きであるし、周囲の騒音に脆弱である。これらの障壁を無くすためには、身体的制約を超えて発声器官を動作させて音声を生成したり、適切な発音動作を指定して音声を生成したり、聴取困難なほど微かな音声発声時の発声器官動作から通常音声を生成するなど、物理的・身体的制約を超えた音声生成機能の拡張が必要である。

【０００３】

例えば、喉頭癌などで喉頭を失った喉頭摘出者に対して、残存器官を用いた代替発声法により生成される自然性に乏しい音声を、より自然な音声へと変換する発声補助技術が提案されている（非特許文献１～非特許文献３を参照）。この他にも、非可聴つぶやき音声を自然な音声に変換する技術も提案されており、秘匿性に優れた通話技術としての応用が期待されている。上述の技術はいずれも音声のスペクトル特徴量系列から自然音声の基本周波数(F_0) パターンを予測する問題を扱っている点で共通しており、学習処理と変換処理で構成される。学習処理では、対象音声（前者であれば電気音声、後者であれば非可聴つぶやき音声）と通常音声の同一発話データを用いる。まず各離散時刻（以後，フレーム）において、前後数フレームから得られる対象音声のスペクトル特徴量と、通常音声の対数F_0とその動的成分（時間微分または時間差分）を抽出し、スペクトル距離尺度に基づく動的時間伸縮によりこれらを対応付けた結合ベクトルを得る。これをパラレルデータと呼ぶ。各フレームのパラレルデータを用い、対象音声のスペクトル特徴量と通常音声の対数F_0 の静的・動的成分の結合確率密度関数を混合正規分布モデル（Gaussian Mixture Model; GMM）で表現する。GMM のパラメータはExpectation-Maximization アルゴリズムにより学習することができる。変換処理では、学習されたGMM を用いて、系列内変動を考慮した最尤系列変換法により、対象音声のスペクトル特徴量系列から通常音声のF_0 パターン

へと変換することができる。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano, "Speaking-aid systems using GMM-based voice conversion for electrolaryngeal speech," Speech Communication, vol. 54, no. 1, pp. 134-146, 2012.
【非特許文献２】Kou Tanaka, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, "A hybrid approach to electrolaryngeal speech enhancement based on noise reduction and statistical excitation generation," IEICE Transactions on Information and Systems, vol. E97-D, no. 6, pp. 1429-1437, Jun. 2014.
【非特許文献３】Kou Tanaka, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, "Direct F0 controlof an electrolarynx based on statistical excitation feature prediction and its evaluation through simulation," Proc. INTERSPEECH, pp. 31-35, Sep. 2014.
【非特許文献４】Hirokazu Kameoka, Jonathan Le Roux, Yasunori Ohishi, "A statistical model of speech F0 contours," ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition (SAPA 2010), pp. 43-48, Sep. 2010.
【非特許文献５】Kota Yoshizato, Hirokazu Kameoka, Daisuke Saito, Shigeki Sagayama, "Hidden Markov convolutive mixture model for pitch contour analysis of speech," in Proc. The 13th Annual Conference of the International Speech Communication Association (Interspeech 2012), Sep. 2012.
【発明の概要】
【発明が解決しようとする課題】
【０００５】
従来技術では、学習処理や変換処理において音声のF_0 パターンの物理的な生成過程を考慮したモデルが用いられていなかったため、物理的に人間が発声しえないような不自然なF_0 パターンを生成することが起こりえた。この問題に対し、F_0パターンの物理的な生成過程を考慮した予測を行うことで、より自然なF_0 パターンを生成できる可能性がある。
【０００６】
F_0パターンは声帯に張力を与える甲状軟骨の運動によって生み出されており、非特許文献４、５ではその制御機構の確率モデルに基づき、フレーズ・アクセント指令と呼ぶ甲状軟骨の運動に関係するパラメータを推定する技術が提案されている。この技術では，フレーズ・アクセント指令の時系列の生成プロセスを隠れマルコフモデル(HMM) により表現した点がポイントの一つであり，HMM のトポロジーの設計や遷移確率の学習を通して、指令列に関する言語学的ないし先験的な知識をパラメータ推定に組み込むことが可能である。
【０００７】
本発明は、上記事情を鑑みてなされたものであり、F_0 パターンの物理的な生成過程の制約を考慮しながらスペクトル特徴量系列に対応する最適なF_0パターンを推定することができる基本周波数パターン予測装置、方法、及びプログラムを提供することを目的とする。
【課題を解決するための手段】
【０００８】
上記の目的を達成するために本発明に係る基本周波数パターン予測装置は、学習サンプルのソース音声の時系列データとターゲット音声の時系列データとからなるパラレルデータを入力として、前記ソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の時系列データから抽出される、各時刻の基本周波数とに基づいて、前記ソース音声の各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の各時刻の基本周波数との間の関係をモデル化した第１確率分布のパラメータを学習

する第１モデルパラメータ学習部と、前記ターゲット音声の各時刻の基本周波数に基づいて、基本周波数パターン生成過程をモデル化した第２確率分布のパラメータを学習する第２モデルパラメータ学習部と、予測対象のソース音声の時系列データを入力として、前記予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記第１モデルパラメータ学習部によって学習された前記第１確率分布のパラメータと、前記第２モデルパラメータ学習部によって学習された前記第２確率分布のパラメータとに基づいて、前記第１確率分布と前記第２確率分布とを用いて表される規準を大きくするように、前記予測対象のソース音声に対応する前記ターゲット音声の各時刻の基本周波数を予測する基本周波数予測部と、を含んで構成されている。

【０００９】

本発明に係る基本周波数パターン予測方法は、第１モデルパラメータ学習部と、第２モデルパラメータ学習部と、基本周波数予測部とを含む基本周波数パターン予測装置における基本周波数パターン予測方法であって、前記第１モデルパラメータ学習部が、学習サンプルのソース音声の時系列データとターゲット音声の時系列データとからなるパラレルデータを入力として、前記ソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の時系列データから抽出される、各時刻の基本周波数とに基づいて、前記ソース音声の各時刻のスペクトル特徴量ベクトルと、前記ターゲット音声の各時刻の基本周波数との間の関係をモデル化した第１確率分布のパラメータを学習し、前記第２モデルパラメータ学習部が、前記ターゲット音声の各時刻の基本周波数に基づいて、基本周波数パターン生成過程をモデル化した第２確率分布のパラメータを学習し、前記基本周波数予測部が、予測対象のソース音声の時系列データを入力として、前記予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、前記第１モデルパラメータ学習部によって学習された前記第１確率分布のパラメータと、前記第２モデルパラメータ学習部によって学習された前記第２確率分布のパラメータとに基づいて、前記第１確率分布と前記第２確率分布とを用いて表される規準を大きくするように、前記予測対象のソース音声に対応する前記ターゲット音声の各時刻の基本周波数を予測する。

【００１０】

本発明に係るプログラムは、上記の基本周波数パターン予測装置の各部としてコンピュータを機能させるためのプログラムである。

【発明の効果】

【００１１】

以上説明したように、本発明の基本周波数パターン予測装置、方法、及びプログラムによれば、ソース音声の各時刻のスペクトル特徴量ベクトルと、ターゲット音声の各時刻の基本周波数との間の関係をモデル化した第１確率分布のパラメータを学習し、基本周波数パターン生成過程をモデル化した第２確率分布のパラメータを学習し、予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルから、第１確率分布と第２確率分布とを用いて表される規準を大きくするように、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数を予測することにより、F_0 パターンの物理的な生成過程の制約を考慮しながらスペクトル特徴量系列に対応する最適なF_0パターンを推定することができる、という効果が得られる。

【図面の簡単な説明】

【００１２】

【図１】ＨＭＭの状態遷移ネットワークの一例を説明するための図である。

【図２】ＨＭＭの状態遷移ネットワークの一例を説明するための図である。

【図３】ＨＭＭの状態遷移ネットワークの一例を説明するための図である。

【図４】状態の分割を説明するための図である。

【図５】本発明の第１の実施の形態に係る基本周波数パターン予測装置の構成を示す概略図である。

【図６】本発明の第１の実施の形態に係る基本周波数パターン予測装置の学習部の構成を

示す概略図である。
【図７】本発明の第１の実施の形態に係る基本周波数パターン予測装置の変換処理部の構成を示す概略図である。
【図８】本発明の第１の実施の形態に係る基本周波数パターン予測装置における学習処理ルーチンの内容を示すフローチャートである。
【図９】本発明の第１の実施の形態に係る基本周波数パターン予測装置における基本周波数パターン予測処理ルーチンの内容を示すフローチャートである。
【図１０】本発明の第２の実施の形態に係る基本周波数パターン予測装置の変換処理部の構成を示す概略図である。
【図１１】本発明の第２の実施の形態に係る基本周波数パターン予測装置における学習処理ルーチンの内容を示すフローチャートである。
【図１２】本発明の第２の実施の形態に係る基本周波数パターン予測装置における基本周波数パターン予測処理ルーチンの内容を示すフローチャートである。
【図１３】実験データに用いた音声のF_0パターンを示す図である。
【図１４】実験結果を示す図である。
【発明を実施するための形態】
【００１３】
以下、図面を参照して本発明の実施の形態を詳細に説明する。本発明で提案する技術は、音声の特徴量系列から基本周波数パターンを予測し、原音声の基本周波数パターンを予測した基本周波数パターンに置き換えることで音声の自然性を向上させることを目的とした音声処理技術である。
【００１４】
＜関連技術１：スペクトル特徴量系列からのF_0 パターン予測方法＞
まず、スペクトル特徴量系列からのF_0 パターン予測方法について説明する。
【００１５】
上記非特許文献１～非特許文献３では、スペクトル特徴量系列からF_0 パターンを予測する方法が提案されている。当該従来手法はスペクトル特徴量系列とF_0 パターンの同時確率分布モデルのパラメータを学習する処理と学習した当該モデルを用いて所与のスペクトル特徴量系列からF_0 パターンに変換する処理からなる。
【００１６】
＜学習処理＞
ソース音声（例えば電気音声）とターゲット音声（例えば自然音声）のパラレルデータが与えられているものとする。ソース音声のスペクトル特徴量ベクトルをc[k] とし，ターゲット音声の対数F_0 とその動的成分（時間微分または時間差分）の結合ベクトル（F_0 特徴量と呼ぶ。）をq[k] = (y[k]; Δy[k])^T とする。ここでk は離散時刻のインデックスである。音声特徴量c[k] としては例えば時刻k を中心とした前後数フレーム分のメルケプストラム（ベクトル）の系列を連結したベクトルに対し主成分分析により次元圧縮を行ったものを用いる。本手法ではc[k] とq[k] の同時確率分布を混合正規分布モデル（Gaussian Mixture Model; GMM）
【００１７】
【数１】

$$P(\boldsymbol{c}[k], \boldsymbol{q}[k]|\boldsymbol{\gamma}) = \sum_m \alpha_m \mathcal{N}\left(\begin{bmatrix}\boldsymbol{c}[k]\\ \boldsymbol{q}[k]\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_m^{(c)}\\ \boldsymbol{\mu}_m^{(q)}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_m^{(cc)} & \boldsymbol{\Sigma}_m^{(cq)}\\ \boldsymbol{\Sigma}_m^{(qc)} & \boldsymbol{\Sigma}_m^{(qq)}\end{bmatrix}\right) \quad (1)$$

$$\sum_m \alpha_m = 1 \quad (2)$$

【００１８】
でモデル化し、学習処理では所与のパラレルデータ {c[k]; q[k]} $^K_{k=1}$ から当該GMM の

パラメータ（各正規分布の重み，平均、分散共分散行列）を学習する。ただし、N(x;μ，Σ）は、xの確率密度関数が平均がμ、分散共分散行列がΣの正規分布で与えられることを意味する。

【００１９】

GMM のパラメータはExpectation-Maximization (EM) アルゴリズムにより推定することができる。学習したGMM パラメータをγとすると，条件付分布P(q[k]｜c[k]，γ）はスペクトル特徴量c[k] からF_0 特徴量q[k] を予測するための分布と見ることができ、

【００２０】

【数２】

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) = \sum_m P(m|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}})P(\boldsymbol{q}[k]|\boldsymbol{c}[k],m,\hat{\boldsymbol{\gamma}}) \quad (3)$$

【００２１】

のようにP(c[k]，q[k]｜γ）と同様にGMM で与えられる。ただし、

【００２２】

【数３】

$$P(m|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) = \frac{\hat{\alpha}_m \mathcal{N}(\boldsymbol{c}[k];\hat{\boldsymbol{\mu}}_m^{(c)},\hat{\boldsymbol{\Sigma}}_m^{(cc)})}{\sum_n \hat{\alpha}_n \mathcal{N}(\boldsymbol{c}[k];\hat{\boldsymbol{\mu}}_n^{(c)},\hat{\boldsymbol{\Sigma}}_n^{(cc)})} \quad (4)$$

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k],m,\hat{\boldsymbol{\gamma}}) = \mathcal{N}\left(\boldsymbol{q}[k];\boldsymbol{e}_m^{(q|c)},\boldsymbol{D}_m^{(q|c)}\right) \quad (5)$$

【００２３】

であり$e^{(q|c)}_m$および$D^{(q|c)}_m$ は

【００２４】

【数４】

$$\boldsymbol{e}_m^{(q|c)} = \hat{\boldsymbol{\mu}}_m^{(q)} + \hat{\boldsymbol{\Sigma}}_m^{(qc)}\hat{\boldsymbol{\Sigma}}_m^{(cc)-1}(\boldsymbol{c}[k] - \hat{\boldsymbol{\mu}}_m^{(c)}) \quad (6)$$

$$\boldsymbol{D}_m^{(q|c)} = \hat{\boldsymbol{\Sigma}}_m^{(qq)} - \hat{\boldsymbol{\Sigma}}_m^{(qc)}\hat{\boldsymbol{\Sigma}}_m^{(cc)-1}\hat{\boldsymbol{\Sigma}}_m^{(cq)} \quad (7)$$

【００２５】

で与えられる。

【００２６】

＜変換処理＞

変換処理では。所与のスペクトル特徴量系列

【００２７】

【数５】

$$\boldsymbol{c} = (\boldsymbol{c}[1]^\mathsf{T},\ldots,\boldsymbol{c}[K]^\mathsf{T})^\mathsf{T}$$

【００２８】

の下で，最尤のF_0パターン

【００２９】

【数６】

$$\boldsymbol{y} = (y[1],\ldots,y[K])^\mathsf{T}$$

【００３０】

を以下の式（８）により求めることが目的である。

【００３１】

【数７】

$$\hat{\boldsymbol{y}} = \underset{\boldsymbol{y}}{\operatorname{argmax}} \log P(\boldsymbol{q}|\boldsymbol{c},\boldsymbol{\gamma}) \quad \text{subject to} \quad \boldsymbol{q} = \boldsymbol{W}\boldsymbol{y} \tag{8}$$

【００３２】

ただし、

【００３３】

【数８】

$$\boldsymbol{q} = (\boldsymbol{q}[1]^{\mathsf{T}}, \ldots, \boldsymbol{q}[K]^{\mathsf{T}})^{\mathsf{T}}$$

【００３４】

であり、Ｗはｙとｑの関係を記述した変換行列（定数）である。ここで、Ｐ（ｑ｜ｃ，γ）は学習処理により学習したパラメータγで与えられるＧＭＭ であり、

【００３５】

【数９】

$$P(\boldsymbol{q}|\boldsymbol{c},\hat{\boldsymbol{\gamma}}) = \prod_k P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \tag{9}$$

$$= \prod_k \sum_{m_k} P(m_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) P(\boldsymbol{q}[k]|\boldsymbol{c}[k],m_k,\hat{\boldsymbol{\gamma}}) \tag{10}$$

【００３６】

で与えられる。ただし、ｍ=（m_1，...，m_K）であり、m_k は時刻ｋ におけるＧＭＭの成分インデックスを表す。ここで、Ｐ(ｑ[ｋ]｜ｃ[ｋ]，γ）が

【００３７】

【数１０】

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \simeq P(\hat{m}_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{m}_k,\hat{\boldsymbol{\gamma}}) \tag{11}$$

$$\hat{m}_k = \underset{m_k}{\operatorname{argmax}} P(m_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \tag{12}$$

【００３８】

により近似できるとする。式(12)より、$\hat{m}_k$は、データｃ[ｋ]を生成したらしい確率が最も高い正規分布のインデックスを意味する。よって、式(9)よりＰ(ｑ｜ｃ，ｍ，＾γ）はすべてのｋについて

【００３９】

【数１１】

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{m}_k,\hat{\boldsymbol{\gamma}}) = \mathcal{N}(\boldsymbol{q}[k]; \boldsymbol{e}^{(q|c)}_{\hat{m}_k}, \boldsymbol{D}^{(q|c)}_{\hat{m}_k})$$

【００４０】

の積をとったもので与えられる。式(11)の近似によりＰ(ｑ｜ｃ，γ）は

【００４１】

【数１２】

$$\boldsymbol{e}^{(q|c)}_{\hat{m}_1}, \ldots, \boldsymbol{e}^{(q|c)}_{\hat{m}_K}$$

【００４２】
を連結したベクトルｅ(ｑ｜ｃ)を平均、
【００４３】
【数１３】

$$\boldsymbol{D}_{\hat{m}_1}^{(q|c)}, \ldots, \boldsymbol{D}_{\hat{m}_K}^{(q|c)}$$

【００４４】
を対角成分としたブロック対角行列Ｄ(ｑ｜ｃ) を分散共分散行列とした正規分布
【００４５】
【数１４】

$$P(\boldsymbol{q}|\boldsymbol{c},\boldsymbol{\gamma}) = \mathcal{N}(\boldsymbol{q}; \boldsymbol{e}^{(q|c)}, \boldsymbol{D}^{(q|c)}) \tag{13}$$

【００４６】
となる。これにｑ＝Ｗｙを代入し、ｙの分布となるように正規化すると、
【００４７】
【数１５】

$$\mathcal{N}(\boldsymbol{W}\boldsymbol{y}; \boldsymbol{e}^{(q|c)}, \boldsymbol{D}^{(q|c)}) \propto \mathcal{N}\left(\boldsymbol{y}; (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{e}^{(q|c)}, (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W})^{-1}\right) \tag{14}$$

【００４８】
となるため、
【００４９】
【数１６】

$$P(\boldsymbol{y}|\boldsymbol{c},\boldsymbol{\gamma}) = \mathcal{N}\left(\boldsymbol{y}; (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{e}^{(q|c)}, (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W})^{-1}\right) \tag{15}$$

【００５０】
のようにｙの条件付き分布を得る。よって、式(8)の解は、
【００５１】
【数１７】

$$\hat{\boldsymbol{y}} = (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{e}^{(q|c)} \tag{16}$$

【００５２】
となる。
【００５３】
＜関連技術２：F_0パターン生成過程モデル＞
次に、F_0パターン生成過程の確率モデルについて説明する。
【００５４】
音声のF_0パターンの生成過程を記述したモデルに、藤崎の基本周波数（F_0）パターン生成過程モデル（藤崎モデル）が知られている（非特許文献６）。
【００５５】
［非特許文献６］：H. Fujisaki, "In Vocal Physiology: Voice Production, Mechanisms and Functions," Raven Press, 1988.
【００５６】
藤崎モデルとは、甲状軟骨の運動によるF_0 パターンの生成過程を説明した物理モデルである。藤崎モデルでは、甲状軟骨の二つの独立な運動（平行移動運動と回転運動）にそれぞれ伴う声帯の伸びの合計がF_0の時間的変化をもたらすと解釈され、声帯の伸びとF_0パ

ターンの対数値y(t) が比例関係にあるという仮定に基づいてF_0パターンがモデル化される。甲状軟骨の平行移動運動によって生じるF_0パターン$x_p(t)$をフレーズ成分、回転運動によって生じるF_0 パターン$x_a(t)$ をアクセント成分と呼ぶ。藤崎モデルでは、音声のF_0パターンy(t) は、これらの成分に声帯の物理的制約によって決まるベースライン成分bを足し合わせたものとして、

【００５７】

【数１８】

$$y(t) = x_{\mathrm{p}}(t) + x_{\mathrm{a}}(t) + \mu_b \tag{17}$$

【００５８】

と表現される。これら二つの成分は二次の臨界制動系の出力と仮定され、

【００５９】

【数１９】

$$x_{\mathrm{p}}(t) = g_{\mathrm{p}}(t) * u_{\mathrm{p}}(t), \tag{18}$$

$$g_{\mathrm{p}}(t) = \begin{cases} \alpha^2 t e^{-\alpha t} & (t \geq 0) \\ 0 & (t < 0) \end{cases}, \tag{19}$$

$$x_{\mathrm{a}}(t) = g_{\mathrm{a}}(t) * u_{\mathrm{a}}(t), \tag{20}$$

$$g_{\mathrm{a}}(t) = \begin{cases} \beta^2 t e^{-\beta t} & (t \geq 0) \\ 0 & (t < 0) \end{cases}, \tag{21}$$

【００６０】

と表される（＊ は時刻t に関する畳み込み演算）。ここで$u_p(t)$ はフレーズ指令関数と呼ばれ、デルタ関数（フレーズ指令）の列からなり、$u_a(t)$ はアクセント指令関数と呼ばれ、矩形波（アクセント指令）の列からなる。これらの指令列には、発話の最初にはフレーズ指令が生起する、フレーズ指令は二連続で生起しない、異なる二つの指令は同時刻に生起しない、という制約条件がある。またαとβはそれぞれフレーズ制御機構、アクセント制御機構の固有角周波数であり、話者や発話内容によらず、おおよそα=3 rad/s、β=20 rad/s 程度であることが経験的に知られている。

【００６１】

＜関連技術３：F_0パターン生成過程モデルパラメータ推定法＞

上述の藤崎モデルは以下のような確率モデルで記述することができる（非特許文献４、５、７参照）。

【００６２】

[非特許文献７]：石原達馬，吉里幸太，亀岡弘和，齋藤大輔，嵯峨山茂樹，\音声基本周波数の藤崎モデル指令列の統計的語彙モデル," 日本音響学会2013 年春季研究発表会講演論文集，1-7-9，pp. 283-286，Mar. 2013.

【００６３】

まずフレーズ、アクセント指令関数のペア$o[k] = (u_p[k], u_a[k])^T$ を出力するHMM を考える。ただし、k は離散時刻のインデックスを表す。状態出力分布は正規分布とし、各時刻の状態が与えられた下で

【００６４】

【数２０】

$$\boldsymbol{o}[k] \sim \mathcal{N}(\boldsymbol{o}[k]; \boldsymbol{\rho}[k], \boldsymbol{\Upsilon}[k]) \tag{22}$$

$$\boldsymbol{\rho}[k] = \begin{bmatrix} \mu_{s_k}^{(\mathrm{p})}[k] \\ \mu_{s_k}^{(\mathrm{a})} \end{bmatrix}, \quad \boldsymbol{\Upsilon}[k] = \begin{bmatrix} \upsilon_{s_k}^{(\mathrm{p})2} & 0 \\ 0 & \upsilon_{s_k}^{(\mathrm{a})2} \end{bmatrix} \tag{23}$$

【００６５】

により指令関数ペアo[k] が生成されるものとする。ここで $\{s_k\}_{k=1}^{K}$ はHMMの状態系列であり、平均ベクトルρ[k] はHMM の状態遷移の結果として定まる値である。具体的なHMM の構成の例を図１～図３に示す。

【００６６】

図1に示すHMM の状態遷移ネットワークの例では、状態$t = r_0$ において$\mu^{(p)}_t$ [k] と$\mu^{(a)}_t$ はいずれも0 である。状態$t = r_0$ からは状態p_0 にのみ遷移することができ、状態$t = p_0$ において$\mu^{(p)}_t$ [k] は非負値$A^{(p)}$[k] をとり、$\mu^{(a)}_t$ は0 となる。状態$t = p_0$の次は状態r_1 にのみ遷移することが許される。状態$t = r_0$ 同様、状態$t = r_1$ において$\mu^{(p)}_t$ [k] と$\mu^{(a)}_t$ はいずれも0 である。状態$t = r_1$ からは状態$a_0, \ldots, a_{N-1}$のいずれかにのみ遷移することができ、状態$t = a_n$ において$\mu^{(a)}_t$ は非負値$A^{(a)}_n$をとり、$\mu^{(p)}_t$ [k] は0 となる。状態$t = a_n$の次は状態r_0 またはr_1にのみ遷移することが許される。これよりμ_a[k] が矩形パルス列となることが保証される。

【００６７】

図2に示すHMMの状態遷移ネットワークの例では、状態$t = r_0$ において$\mu^{(p)}_t$ [k] と$\mu^{(a)}_t$ はいずれも0 である。状態$t = r_0$ からは状態$p_0, \ldots, p_{M-1}$のいずれかにのみ遷移することができ、状態$t = p_m$ において$\mu^{(p)}_t$ は非負値$A^{(p)}_m$をとり、$\mu^{(a)}_t$ は0 となる。状態$t = p_m$の次は状態r_1にのみ遷移することが許される。状態$t = r_0$ 同様、状態$t = r_1$ において$\mu^{(p)}_t$ [k] と$\mu^{(a)}_t$ はいずれも0 である。状態$t = r_1$ からは状態$a_0, \ldots, a_{N-1}$のいずれかにのみ遷移することができ、状態$t = a_n$ において$\mu^{(a)}_t$ は非負値$A^{(a)}_n$をとり、$\mu^{(p)}_t$ は0 となる。状態$t = a_n$の次は状態r_0 またはr_1 にのみ遷移することが許される。これよりμ_a[k] が矩形パルス列となることが保証される。

【００６８】

図3に示すHMMの状態遷移ネットワークの例では、それぞれの終点と始点が連結された複数のLeft-to-Right 型HMM からなる。図１、２と同様、状態$t = r_l$において$\mu^{(p)}_t$ [k] と$\mu^{(a)}_t$ はいずれも0 である。また、状態$t = p_m$ において$\mu^{(p)}_t$は非負値$A^{(p)}_m$ をとり、$\mu^{(a)}_t$ は0 となる。状態$t = a_n$において$\mu^{(a)}_t$は非負値$A^{(a)}_n$をとり、$\mu^{(p)}_t$は0となる。

【００６９】

ρ[k]は、図1 の例では、以下の式（２４）で表わされる。

【００７０】

【数２１】

$$\boldsymbol{\rho}[k] = \begin{bmatrix} \mu_{s_k}^{(\mathrm{p})}[k] \\ \mu_{s_k}^{(\mathrm{a})} \end{bmatrix} = \begin{cases} (0,0)^{\mathsf{T}} & (s_k \in \mathrm{r}_l) \\ (A^{(\mathrm{p})}[k], 0)^{\mathsf{T}} & (s_k \in \mathrm{p}_0) \\ (0, A_n^{(\mathrm{a})})^{\mathsf{T}} & (s_k \in \mathrm{a}_n) \end{cases} \tag{24}$$

【００７１】

また、図2，3 の例では、ρ[k]は、以下の式（２５）で表わされる。

【００７２】

【数２２】

$$\boldsymbol{\rho}[k] = \begin{bmatrix} \mu_{s_k}^{(\mathrm{p})} \\ \mu_{s_k}^{(\mathrm{a})} \end{bmatrix} = \begin{cases} (0,0)^{\mathsf{T}} & (s_k \in \mathrm{r}_l) \\ (A_m^{(\mathrm{p})},0)^{\mathsf{T}} & (s_k \in \mathrm{p}_m) \\ (0,A_n^{(\mathrm{a})})^{\mathsf{T}} & (s_k \in \mathrm{a}_n) \end{cases} \tag{25}$$

【００７３】

いずれの例においても、図４のようにそれぞれの状態を同じ出力分布をもついくつかの小状態に分割し、Left-to-Right 型の状態遷移経路を制約することで同一状態に停留する時間長の確率をパラメータ化することができる。図4 は状態a_n を分割した例である。例えばこの図のように全ての$m \neq 0$ に対して$a_{n,n'}$ から$a_{n,n'+1}$ への状態遷移確率を1に設定することで，$a_{n,0}$から$a_{n,n'}$への遷移確率が状態a_n がn′ステップだけ持続する確率に対応し、アクセント指令の持続長の確率を設定したり学習したりできるようになる。同様にp_m とr_l も小状態に分割することで、フレーズ指令の持続長と指令間の間隔の長さの分布をパラメータ化することが可能になる。以後、状態集合を

【００７４】

【数２３】

$$\mathrm{r}_l = \{\mathrm{r}_{l,0}, \mathrm{r}_{l,1}, \ldots\},\ \ \mathrm{p}_m = \{\mathrm{p}_{m,0}, \mathrm{p}_{0,1}, \ldots\},\ \ \mathrm{a}_n = \{\mathrm{a}_{n,0}, \mathrm{a}_{n,1}, \ldots\}$$

【００７５】

と表記する。上記のHMM の構成は次のように書ける。

【００７６】

【数２４】

$$\begin{aligned} \text{出力値系列:}\ & \{\boldsymbol{o}[k]\}_{k=1}^{K} \\ \text{状態集合:}\ & \mathcal{S} = \{\mathrm{r}_0, \ldots, \mathrm{r}_{L-1}, \mathrm{p}_0, \ldots, \mathrm{p}_{M-1}, \mathrm{a}_0, \ldots, \mathrm{a}_{N-1}\} \\ \text{状態系列:}\ & \{s_k\}_{k=1}^{K} \\ \text{状態出力分布:}\ & P(\boldsymbol{o}[k]|s_k) = \mathcal{N}(\boldsymbol{o}[k]; \boldsymbol{\rho}[k], \boldsymbol{\Upsilon}[k]) \\ & \boldsymbol{\rho}[k] = \begin{bmatrix} \mu_{s_k}^{(\mathrm{p})}[k] \\ \mu_{s_k}^{(\mathrm{a})} \end{bmatrix} \\ & \boldsymbol{\Upsilon}[k] = \begin{bmatrix} v_{s_k}^{(\mathrm{p})2} & 0 \\ 0 & v_{s_k}^{(\mathrm{a})2} \end{bmatrix} \\ \text{状態遷移確率:}\ & \phi_{i',i} = \log P(s_k = i | s_{k-1} = i') \end{aligned}$$

【００７７】

状態系列$s = \{s_k\}_{k=1}^{K}$ が与えられたとき、このHMM はフレーズ指令関数$u_p[k]$ とアクセント指令関数$u_a[k]$ のペアを出力する。式(18) と式(20) で示した通り、$u_p[k]$ と$u_a[k]$ にそれぞれ$g_p[k]$ と$g_a[k]$ が畳み込まれてフレーズ成分$x_p[k]$ とアクセント成分$x_a[k]$ が出力される。これを式で表すと、

【００７８】

【数２５】

$$x_{\mathrm{p}}[k] = g_{\mathrm{p}}[k] * u_{\mathrm{p}}[k], \tag{26}$$

$$x_{\mathrm{a}}[k] = g_{\mathrm{a}}[k] * u_{\mathrm{a}}[k], \tag{27}$$

【００７９】

と書ける（＊ は離散時刻k に関する畳み込み演算）。このとき，F_0 パターンx[k] は

【００８０】

【数２６】

$$x[k] = x_{\mathrm{p}}[k] + x_{\mathrm{a}}[k] + \mu_{\mathrm{b}}, \tag{28}$$

【００８１】

と三種類の成分の重ね合わせで書ける。ただしb は時刻によらないベースライン成分である。

【００８２】

また、実音声においては、いつも信頼のできるF_0 の値が観測できるとは限らない。藤崎モデルのパラメータ推定を行うにあたっては、信頼のおける観測区間のF_0 値のみを考慮に入れて、そうでない区間は無視することが望ましい。例えば音声の無声区間においては通常声帯の振動に伴う周期的な粗密波は観測されないので、仮に自動ピッチ抽出によって音声の無声区間から何らかの値がF_0 の推定値として得られたとしても、その値を声帯から発せられる信号のF_0 の値と見なすのは適当ではない。そこで、提案モデルに観測F_0 値の時刻k における不確かさの程度$v^2_n[k]$ を導入する。具体的には、観測F_0 値y[k] を、真のF_0 値x[k] とノイズ成分

【００８３】

【数２７】

$$x_{\mathrm{n}}[k] \sim \mathcal{N}\left(0, \upsilon^2_{\mathrm{n}}[k]\right)$$

【００８４】

との重ね合わせで

【００８５】

【数２８】

$$y[k] = x[k] + x_{\mathrm{n}}[k] \tag{29}$$

【００８６】

と表現することで、信頼のおける区間かどうかに関わらず全ての観測区間を統一的に扱える。

【００８７】

$x_n[k]$ を周辺化することで、出力値系列$o = \{o[k]\}^{K}_{k=1}$ が与えられたときの$y = \{y[k]\}^{K}_{k=1}$ の確率密度関数

【００８８】

【数２９】

$$P(\boldsymbol{y}|\boldsymbol{o}) = \prod_{k=1}^{K} \mathcal{N}(y[k]; x[k], \upsilon^2_{\mathrm{n}}[k]) \tag{30}$$

【００８９】

が得られる。状態系列$s = \{s_k\}^{K}_{k=1}$ と指令の振幅を表すパラメータ

【００９０】

【数３０】

$$\boldsymbol{\theta} = \begin{cases} \{\{A^{(\mathrm{p})}[k]\}_{k=1}^{K}, \{A_n^{(\mathrm{a})}\}_{n=1}^{N}\} & (\text{図 1 の場合}) \\ \{\{A_m^{(\mathrm{p})}\}_{m=1}^{M}, \{A_n^{(\mathrm{a})}\}_{n=1}^{N}\} & (\text{図 2, 3 の場合}) \end{cases} \tag{31}$$

【００９１】
および遷移確率行列φ＝（$\phi_{i,j}$）$_{I\times I}$が与えられたとき、出力値系列oは
【００９２】
【数３１】

$$P(\boldsymbol{o}|\boldsymbol{s},\boldsymbol{\theta})=\prod_{k=1}^{K}\mathcal{N}(\boldsymbol{o}[k];\boldsymbol{\rho}[k],\boldsymbol{\Upsilon}[k]) \tag{32}$$

【００９３】
に従って生成される。また、P(s｜φ）は状態遷移確率の積として
【００９４】
【数３２】

$$\log P(\boldsymbol{s})=\phi_{s_1}+\sum_{k=2}^{K}\phi_{s_{k-1},s_k} \tag{33}$$

と書ける。ただし、
【００９５】
【数３３】

$$\phi_{s_1}$$

【００９６】
は初期状態がs_1である確率をあらわす。式(30)、(32）および式(33）よりP(y,o,s｜θ,φ）は
【００９７】
【数３４】

$$P(\boldsymbol{y},\boldsymbol{o},\boldsymbol{s}|\boldsymbol{\theta},\boldsymbol{\phi})=P(\boldsymbol{y}|\boldsymbol{o})P(\boldsymbol{o}|\boldsymbol{s},\boldsymbol{\theta})P(\boldsymbol{s}|\boldsymbol{\phi}) \tag{34}$$

【００９８】
と書ける。これをo に関して周辺化すると
【００９９】
【数３５】

$$P(\boldsymbol{y},\boldsymbol{s}|\boldsymbol{\theta},\boldsymbol{\phi})=P(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{\theta})P(\boldsymbol{s}|\boldsymbol{\phi}) \tag{35}$$

$$P(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{\theta})=\mathcal{N}(\boldsymbol{y};\boldsymbol{G}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}}+\boldsymbol{G}_{\mathrm{a}}\boldsymbol{\mu}_{\mathrm{a}}+\mu_{\mathrm{b}}\mathbf{1},\boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{\mathsf{T}}+\boldsymbol{G}_{\mathrm{a}}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{G}_{\mathrm{a}}^{\mathsf{T}}+\boldsymbol{\Upsilon}_{\mathrm{n}}) \tag{36}$$

【０１００】
が得られる。ただし、
【０１０１】

【数３６】

$$\boldsymbol{G}_{\mathrm{p}} = \begin{bmatrix} g_{\mathrm{p}}[1] & & & & O \\ g_{\mathrm{p}}[2] & g_{\mathrm{p}}[1] & & & \\ g_{\mathrm{p}}[3] & g_{\mathrm{p}}[2] & g_{\mathrm{p}}[1] & & \\ \vdots & \ddots & \ddots & \ddots & \\ g_{\mathrm{p}}[K] & \cdots & g_{\mathrm{p}}[3] & g_{\mathrm{p}}[2] & g_{\mathrm{p}}[1] \end{bmatrix} \tag{37}$$

$$\boldsymbol{G}_{\mathrm{a}} = \begin{bmatrix} g_{\mathrm{a}}[1] & & & & O \\ g_{\mathrm{a}}[2] & g_{\mathrm{a}}[1] & & & \\ g_{\mathrm{a}}[3] & g_{\mathrm{a}}[2] & g_{\mathrm{a}}[1] & & \\ \vdots & \ddots & \ddots & \ddots & \\ g_{\mathrm{a}}[K] & \cdots & g_{\mathrm{a}}[3] & g_{\mathrm{a}}[2] & g_{\mathrm{a}}[1] \end{bmatrix} \tag{38}$$

$$\boldsymbol{\mu}_{\mathrm{p}} = \begin{cases} (\mu^{(\mathrm{p})}_{s_1}[1], \mu^{(\mathrm{p})}_{s_2}[2], \ldots, \mu^{(\mathrm{p})}_{s_K}[K])^{\mathsf{T}} & (\text{図 1 の場合}) \\ (\mu^{(\mathrm{p})}_{s_1}, \mu^{(\mathrm{p})}_{s_2}, \ldots, \mu^{(\mathrm{p})}_{s_K})^{\mathsf{T}} & (\text{図 2, 3 の場合}) \end{cases} \tag{39}$$

$$\boldsymbol{\mu}_{\mathrm{a}} = (\mu^{(\mathrm{a})}_{s_1}, \mu^{(\mathrm{a})}_{s_2}, \ldots, \mu^{(\mathrm{a})}_{s_K})^{\mathsf{T}} \tag{40}$$

$$\mathbf{1} = (\underbrace{1, 1, \ldots, 1}_{K})^{\mathsf{T}} \tag{41}$$

$$\boldsymbol{\Upsilon}_{\mathrm{p}} = \begin{bmatrix} v^{(\mathrm{p})}_{s_1} & & O \\ & \ddots & \\ O & & v^{(\mathrm{p})}_{s_K} \end{bmatrix} \tag{42}$$

$$\boldsymbol{\Upsilon}_{\mathrm{a}} = \begin{bmatrix} v^{(\mathrm{a})}_{s_1} & & O \\ & \ddots & \\ O & & v^{(\mathrm{a})}_{s_K} \end{bmatrix} \tag{43}$$

$$\boldsymbol{\Upsilon}_{\mathrm{n}} = \begin{bmatrix} v^{2}_{\mathrm{n}}[1] & & O \\ & \ddots & \\ O & & v^{2}_{\mathrm{n}}[K] \end{bmatrix} \tag{44}$$

【０１０２】

である。一方、ｓ に関して周辺化すると

【０１０３】

【数３７】

$$P(\boldsymbol{y}, \boldsymbol{o}|\boldsymbol{\theta}, \boldsymbol{\phi}) = P(\boldsymbol{y}|\boldsymbol{o}) \sum_{\boldsymbol{s}} P(\boldsymbol{o}|\boldsymbol{s}, \boldsymbol{\theta}) P(\boldsymbol{s}|\boldsymbol{\phi}) \tag{45}$$

が得られる。ただし、Σ_{s} はあらゆる状態系列に関して和をとる操作を意味する。

【０１０４】

＜パラメータ推定アルゴリズム１＞

ｙとｏを完全データと見なすことで、式(35) を局所最大化するｓとθ をExpectation-

Maximization アルゴリズムにより探索することができる。導出は省略するが、

【０１０５】

【数３８】

$$Q(\boldsymbol{s},\boldsymbol{\theta}) = -\frac{1}{2}\Big[\log|\boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{\mathsf{T}}| + \log|\boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{\mathsf{T}}| + \mathrm{tr}(\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{R}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{-\mathsf{T}}) + \mathrm{tr}(\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{R}_{\mathrm{a}}\boldsymbol{G}_{\mathrm{a}}^{-\mathsf{T}}) - 2\boldsymbol{\mu}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{G}_{\mathrm{p}}^{-1}\bar{\boldsymbol{x}}_{\mathrm{p}} - 2\boldsymbol{\mu}_{\mathrm{a}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{G}_{\mathrm{a}}^{-1}\bar{\boldsymbol{x}}_{\mathrm{a}} - 2\mu_{\mathrm{b}}\mathbf{1}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{b}}^{-1}\bar{\boldsymbol{x}}_{\mathrm{b}} + \boldsymbol{\mu}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{\mu}_{\mathrm{a}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}^{2}\mathbf{1}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{b}}^{-1}\mathbf{1}\Big] + \log P(\boldsymbol{s}) \tag{46}$$

【０１０６】

が大きくなるようにｓとθを更新するステップと、更新したｓとθを用いて

【０１０７】

【数３９】

$$\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_{\mathrm{p}}^{\mathsf{T}}, \bar{\boldsymbol{x}}_{\mathrm{a}}^{\mathsf{T}}, \bar{\boldsymbol{x}}_{\mathrm{b}}^{\mathsf{T}})^{\mathsf{T}}$$

【０１０８】

とＲを

【０１０９】

【数４０】

$$\bar{\boldsymbol{x}} = \boldsymbol{a} + \boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}})^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{a}) \tag{47}$$

$$\boldsymbol{R} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}})^{-1}\boldsymbol{H}\boldsymbol{\Lambda} + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathsf{T}} \tag{48}$$

【０１１０】

により更新するステップを繰り返すことで式(35) を単調増加させることができる（詳細は、上記非特許文献４参照）。

【０１１１】

具体的には、以下の初期設定、Ｅステップ、及びＭステップが実行される。

【０１１２】

（初期設定）

ｓとθを初期設定する。

【０１１３】

（E ステップ）

フレーズ成分、アクセント成分、ベースライン成分の条件付き期待値

【０１１４】

【数４１】

$$\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_{\mathrm{p}}^{\mathsf{T}}, \bar{\boldsymbol{x}}_{\mathrm{a}}^{\mathsf{T}}, \bar{\boldsymbol{x}}_{\mathrm{b}}^{\mathsf{T}})^{\mathsf{T}}$$

【０１１５】

と条件付き分散共分散行列Ｒを

【０１１６】

【数４２】

$$\bar{\boldsymbol{x}} \leftarrow \boldsymbol{a} + \boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}})^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{a}) \tag{49}$$

$$\boldsymbol{R} \leftarrow \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathsf{T}})^{-1}\boldsymbol{H}\boldsymbol{\Lambda} + \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathsf{T}} \tag{50}$$

により更新する。ただし、

【０１１７】

【数４３】

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{G}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}} \\ \boldsymbol{G}_{\mathrm{a}}\boldsymbol{\mu}_{\mathrm{a}} \\ \mu_{\mathrm{b}}\boldsymbol{1} \end{bmatrix} \tag{51}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{\mathsf{T}} & O & O \\ O & \boldsymbol{G}_{\mathrm{a}}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{G}_{\mathrm{a}}^{\mathsf{T}} & O \\ O & O & \boldsymbol{\Upsilon}_{\mathrm{n}} \end{bmatrix} \tag{52}$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} \end{bmatrix} \tag{53}$$

【０１１８】

である。また、Ｒにおける

【０１１９】

【数４４】

$$\bar{\boldsymbol{x}}_{\mathrm{p}},\ \bar{\boldsymbol{x}}_{\mathrm{a}},\ \bar{\boldsymbol{x}}_{\mathrm{b}}$$

【０１２０】

に対応するブロック対角成分を

【０１２１】

【数４５】

$$\boldsymbol{R}_{\mathrm{p}},\ \boldsymbol{R}_{\mathrm{a}},\ \boldsymbol{R}_{\mathrm{b}}$$

【０１２２】

とする。

【０１２３】

すなわち、

【０１２４】

【数４６】

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{\mathrm{p}} & * & * \\ * & \boldsymbol{R}_{\mathrm{a}} & * \\ * & * & \boldsymbol{R}_{\mathrm{b}} \end{bmatrix} \tag{54}$$

【０１２５】

である（＊ は以後用いないブロック成分である）。

【０１２６】

（Ｍステップ）

Ｑ(ｓ，θ）が最大となる状態系列ｓ=（$s_1,\ldots,s_K$）を探索する。γ_p と γ_aは対角行

【０１２７】

列なので、

【０１２８】

【数４７】

$$\log|\boldsymbol{\Upsilon}_\mathrm{p}| \text{ と } \log|\boldsymbol{\Upsilon}_\mathrm{a}|,\ \mathrm{tr}(\boldsymbol{\Upsilon}_\mathrm{p}^{-1}\boldsymbol{G}_\mathrm{p}^{-1}\boldsymbol{R}_\mathrm{p}\boldsymbol{G}_\mathrm{p}^{-\mathsf{T}}) \text{ と } \mathrm{tr}(\boldsymbol{\Upsilon}_\mathrm{a}^{-1}\boldsymbol{G}_\mathrm{a}^{-1}\boldsymbol{R}_\mathrm{a}\boldsymbol{G}_\mathrm{a}^{-\mathsf{T}}),\ \boldsymbol{\mu}_\mathrm{p}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{p}^{-1}\boldsymbol{G}_\mathrm{p}^{-1}\bar{\boldsymbol{x}}_\mathrm{p}$$

$$\text{と } \boldsymbol{\mu}_\mathrm{a}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{a}^{-1}\boldsymbol{G}_\mathrm{a}^{-1}\bar{\boldsymbol{x}}_\mathrm{a},\ \boldsymbol{\mu}_\mathrm{p}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{p}^{-1}\boldsymbol{\mu}_\mathrm{p} \text{ と } \boldsymbol{\mu}_\mathrm{a}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{a}^{-1}\boldsymbol{\mu}_\mathrm{a}$$

【０１２９】

はいずれも

【０１３０】

【数４８】

$$\log|\boldsymbol{\Upsilon}_\mathrm{p}| = \sum_k \log \upsilon_{s_k}^{(p)2} \tag{55}$$

$$\log|\boldsymbol{\Upsilon}_\mathrm{a}| = \sum_k \log \upsilon_{s_k}^{(p)2} \tag{56}$$

$$\mathrm{tr}(\boldsymbol{\Upsilon}_\mathrm{p}^{-1}\boldsymbol{G}_\mathrm{p}^{-1}\boldsymbol{R}_\mathrm{p}\boldsymbol{G}_\mathrm{p}^{-\mathsf{T}}) = \sum_k \frac{[\boldsymbol{G}_\mathrm{p}^{-1}\boldsymbol{R}_\mathrm{p}\boldsymbol{G}_\mathrm{p}^{-\mathsf{T}}]_{k,k}}{\upsilon_{s_k}^{(p)2}} \tag{57}$$

$$\mathrm{tr}(\boldsymbol{\Upsilon}_\mathrm{a}^{-1}\boldsymbol{G}_\mathrm{a}^{-1}\boldsymbol{R}_\mathrm{a}\boldsymbol{G}_\mathrm{a}^{-\mathsf{T}}) = \sum_k \frac{[\boldsymbol{G}_\mathrm{a}^{-1}\boldsymbol{R}_\mathrm{a}\boldsymbol{G}_\mathrm{a}^{-\mathsf{T}}]_{k,k}}{\upsilon_{s_k}^{(p)2}} \tag{58}$$

$$\boldsymbol{\mu}_\mathrm{p}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{p}^{-1}\boldsymbol{G}_\mathrm{p}^{-1}\bar{\boldsymbol{x}}_\mathrm{p} = \sum_k \frac{\mu_{s_k}^{(p)}[k][\boldsymbol{G}_\mathrm{p}^{-1}\bar{\boldsymbol{x}}_\mathrm{p}]_k}{\upsilon_{s_k}^{(p)2}} \tag{59}$$

$$\boldsymbol{\mu}_\mathrm{a}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{a}^{-1}\boldsymbol{G}_\mathrm{a}^{-1}\bar{\boldsymbol{x}}_\mathrm{a} = \sum_k \frac{\mu_{s_k}^{(a)}[\boldsymbol{G}_\mathrm{a}^{-1}\bar{\boldsymbol{x}}_\mathrm{a}]_k}{\upsilon_{s_k}^{(a)2}} \tag{60}$$

$$\boldsymbol{\mu}_\mathrm{p}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{p}^{-1}\boldsymbol{\mu}_\mathrm{p} = \sum_k \frac{\mu_{s_k}^{(p)}[k]^2}{\upsilon_{s_k}^{(p)2}} \tag{61}$$

$$\boldsymbol{\mu}_\mathrm{a}^{\mathsf{T}}\boldsymbol{\Upsilon}_\mathrm{a}^{-1}\boldsymbol{\mu}_\mathrm{a} = \sum_k \frac{\mu_{s_k}^{(a)2}}{\upsilon_{s_k}^{(a)2}} \tag{62}$$

【０１３１】

のようにｋごとの項の和の形で書ける。従って、Ｑ(ｓ，θ）はｓに依らない項を除けば

【０１３２】

【数４９】

$$\mathcal{J}(\boldsymbol{s}) = -\frac{1}{2}\sum_{k=1}^{K}\left(\log \upsilon_{s_k}^{(p)2} + \log \upsilon_{s_k}^{(p)2} + \frac{[\boldsymbol{G}_{\mathrm{p}}^{-1}\boldsymbol{R}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{-\mathsf{T}}]_{k,k}}{\upsilon_{s_k}^{(p)2}} + \frac{[\boldsymbol{G}_{\mathrm{a}}^{-1}\boldsymbol{R}_{\mathrm{a}}\boldsymbol{G}_{\mathrm{a}}^{-\mathsf{T}}]_{k,k}}{\upsilon_{s_k}^{(p)2}}\right.$$
$$\left. -2\frac{\mu_{s_k}^{(p)}[k][\boldsymbol{G}_{\mathrm{p}}^{-1}\bar{\boldsymbol{x}}_{\mathrm{p}}]_k}{\upsilon_{s_k}^{(p)2}} - 2\frac{\mu_{s_k}^{(a)}[\boldsymbol{G}_{\mathrm{a}}^{-1}\bar{\boldsymbol{x}}_{\mathrm{a}}]_k}{\upsilon_{s_k}^{(a)2}} + \frac{\mu_{s_k}^{(p)}[k]^2}{\upsilon_{s_k}^{(p)2}} + \frac{\mu_{s_k}^{(a)2}}{\upsilon_{s_k}^{(a)2}}\right) + \phi_{s_1} + \sum_{k=2}^{K}\phi_{s_{k-1},s_k} \quad (63)$$

【０１３３】

と書ける。従って、Q(ｓ,θ) を最大にする状態系列ｓ=（ｓ$_1$,...,ｓ$_K$)はViterbiアルゴリズムにより求めることができる（詳細は上記非特許文献４参照）。ただし、[・]$_{k,k}$ は行列のk 行k 列の要素、[・]$_k$はベクトルの第k 要素を表す。

【０１３４】

続いて、Q(ｓ,θ)を最大にするようにθを更新する。Q(ｓ,θ)を最大にするθは、Q(ｓ,θ)の各変数に関する偏微分を0 と置くことにより得られる（（詳細は上記非特許文献４参照）。

【０１３５】

また、推定された状態系列ｓから、状態遷移確率φが求められる。

【０１３６】

［第１の実施の形態］

＜本発明の実施の形態の概要＞

本発明の実施の形態の技術は、上述した関連技術１のスペクトル特徴量系列からのF$_0$パターン予測方法と同様、学習処理と変換処理からなるが、式(8) の代わりに、上述した関連技術１のスペクトル特徴量系列からのF$_0$ パターン予測方法の確率分布と、上述した関連技術３のF$_0$ パターン生成過程モデルパラメータ推定法の確率分布の積を規準とすることにより、上述したF$_0$パターン生成過程モデルにできるだけ即したF$_0$パターンをスペクトル特徴量から統計的に予測することを可能にする技術である。

【０１３７】

学習処理ではパラレルデータの学習サンプル {c[k], q[k]} $^{K}_{k=1}$ が与えられた下で

【０１３８】

【数５０】

$$P(\boldsymbol{c},\boldsymbol{q}|\boldsymbol{\gamma}) = \prod_k P(\boldsymbol{c}[k],\boldsymbol{q}[k]|\boldsymbol{\gamma})$$

【０１３９】

ができるだけ大きくなるようにγを学習する。また、学習サンプルの基本周波数パターン {y[k]} $^{K}_{k=1}$ が与えられた下でP(ｙ，ｓ｜θ、φ)ができるだけ大きくなるようにθ とφを学習する。

【０１４０】

一方、変換処理では入力音声の {c[k]} $^{K}_{k=1}$ が与えられた下でP(q｜c，γ)P(ｙ，ｓ｜θ，φ)またはこれらを近似する確率密度関数ができるだけ大きくなるようにｙを求める。

【０１４１】

＜システム構成＞

次に、ソース音声のスペクトル特徴量系列から、ターゲット音声の基本周波数パターンを予測する基本周波数パターン予測装置に、本発明を適用した場合を例にして、本発明の実施の形態を説明する。

【０１４２】

図５に示すように、本発明の第１の実施の形態に係る基本周波数パターン予測装置は、ＣＰＵと、ＲＡＭと、後述する学習処理ルーチン、及び基本周波数パターン予測処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、

機能的には次に示すように構成されている。

【０１４３】

図５に示すように、基本周波数パターン予測装置１００は、入力部１０と、演算部２０と、出力部９０とを備えている。

【０１４４】

入力部１０は、学習サンプルのソース音声（例えば電気音声）の時系列データとターゲット音声（例えば自然音声）の時系列データとからなるパラレルデータを受け付ける。また、入力部１０は、予測対象のソース音声の時系列データを受け付ける。

【０１４５】

演算部２０は、学習部３０と、パラメータ記憶部４０と、変換処理部５０とを備えている。

【０１４６】

図６に示すように、学習部３０は、特徴量抽出部３２と、基本周波数系列抽出部３４と、第１モデルパラメータ学習部３６と、第２モデルパラメータ学習部３８とを備えている。

【０１４７】

特徴量抽出部３２は、入力部１０によって受け付けた学習サンプルのソース音声の時系列データから、ソース音声のスペクトグラム特徴量ベクトルc[k]を抽出する。ここでk は離散時刻のインデックスである。例えば、非特許文献１～３と同様に時刻k を中心とした前後数フレーム分のメルケプストラム（ベクトル）の系列を連結したベクトルに対し主成分分析により次元圧縮を行ったものをc[k] として用いる。

【０１４８】

基本周波数系列抽出部３４は、入力部１０によって受け付けた学習サンプルのターゲット音声の時系列データから、ターゲット音声の各時刻ｋにおける基本周波数y[k]を抽出し、$y = (y[1], \ldots, y[K])^T$とする。

【０１４９】

この基本周波数の抽出処理は、周知技術により実現でき、例えば、非特許文献８（H. Kameoka, "Statistical speech spectrum model incorporating all-pole vocal tract model and F0 contour generating process model," in Tech. Rep. IEICE, 2010, in Japanese.）に記載の手法を利用して、８ｍｓごとに基本周波数を抽出する。

【０１５０】

また、y とその動的成分（時間微分または時間差分）の結合ベクトル（F_0 特徴量と呼ぶ。）をq[k] = $(y[k], \Delta y[k])^T$ とする。

【０１５１】

以上より、$\{c[k], q[k]\}_{k=1}^{K}$ というデータが得られる。

【０１５２】

第１モデルパラメータ学習部３６は、特徴量抽出部３２によって抽出された各時刻ｋのスペクトル特徴量ベクトルc[k]と、基本周波数系列抽出部３４によって抽出された各時刻ｋの基本周波数の結合ベクトルｑ［ｋ］とに基づいて、ソース音声の各時刻のスペクトル特徴量ベクトルc[k]とターゲット音声の各時刻ｋの基本周波数の結合ベクトルｑ［ｋ］との同時確率分布を表す混合正規分布である第１確率分布のパラメータを学習する。

【０１５３】

具体的には、第１モデルパラメータ学習部３６は、上述したスペクトル特徴量系列からのF_0パターン予測方法の学習処理と同様に、式(1) のGMM のパラメータγを学習する。学習したGMM パラメータを$\hat{\gamma}$とする。

【０１５４】

第２モデルパラメータ学習部３８は、基本周波数系列抽出部３４によって抽出された各時刻ｋの基本周波数y[k]に基づいて、各時刻ｋの基本周波数y[k]と、隠れマルコフモデルの各時刻の状態からなる状態系列ｓとの組み合わせの確率分布である第２確率分布のパラメータを学習する。

【０１５５】
具体的には、第２モデルパラメータ学習部３８は、上述した関連技術３のF_0パターン生成過程モデルパラメータ推定法のパラメータ推定アルゴリズム１に従って、F_0パターン生成過程モデルのパラメータθ、φを学習する。
【０１５６】
もし学習サンプルのフレーズ指令系列とアクセント指令系列のペアのデータ$o = \{o_k\}_{k=1}^{K}$が入手できるのであれば，oからθ、φを学習しても良い（HMMの通常の学習に相当）。学習したF_0パターン生成過程モデルのパラメータを$\hat{\theta}$、$\hat{\phi}$とする。
【０１５７】
変換処理部５０は、予測対象のソース音声の時系列データを入力として、ソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルと、第１モデルパラメータ学習部３６によって学習された第１確率分布のパラメータγと、第２モデルパラメータ学習部３８によって学習された第２確率分布のパラメータθ、φとに基づいて、第１確率分布と第２確率分布との積を用いて表される規準を大きくするように、各時刻の基本周波数ｙと、各時刻の状態からなる状態系列ｓを推定することにより、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数ｙを予測する。
【０１５８】
ここで、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数ｙを予測する原理について説明する。
【０１５９】
学習処理で学習したパラメータ$\hat{\gamma}$、$\hat{\theta}$、$\hat{\phi}$と、予測対象のソース音声の特徴量系列$c = \{c[k]\}_{k=1}^{K}$を用いて、式(9)と式(35)の積
【０１６０】
【数５１】

$$\mathcal{I}_1(\boldsymbol{y}, \boldsymbol{s}) := P(\boldsymbol{q}|\boldsymbol{c}, \hat{\boldsymbol{\gamma}}) P(\boldsymbol{y}, \boldsymbol{s}|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})^{\omega} \tag{77}$$

ができるだけ大きくなるようにｙ、ｓを推定する。ただし、ｑ＝Ｗｙであり、
【０１６１】
【数５２】

$$P(\boldsymbol{q}|\boldsymbol{c}, \hat{\boldsymbol{\gamma}}) = \prod_k \sum_{m_k} P(m_k|\boldsymbol{c}[k], \hat{\boldsymbol{\gamma}}) P(\boldsymbol{q}[k]|\boldsymbol{c}[k], m_k, \hat{\boldsymbol{\gamma}}) \tag{78}$$

$$P(\boldsymbol{y}, \boldsymbol{s}|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) = P(\boldsymbol{s}|\hat{\boldsymbol{\psi}}) P(\boldsymbol{y}|\boldsymbol{s}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \tag{79}$$

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k], m_k, \hat{\boldsymbol{\gamma}}) = \mathcal{N}(\boldsymbol{q}[k]; \boldsymbol{e}_{m_k}^{(q|c)}, \boldsymbol{D}_{m_k}^{(q|c)}) \tag{80}$$

$$P(\boldsymbol{y}|\boldsymbol{s}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{G}_{\mathrm{p}}\boldsymbol{\mu}_{\mathrm{p}} + \boldsymbol{G}_{\mathrm{a}}\boldsymbol{\mu}_{\mathrm{a}} + \mu_{\mathrm{b}}\mathbf{1}, \boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{\mathsf{T}} + \boldsymbol{G}_{\mathrm{a}}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{G}_{\mathrm{a}}^{\mathsf{T}} + \boldsymbol{\Upsilon}_{\mathrm{n}}) \tag{81}$$

【０１６２】
である。ωはF_0パターンの予測においてF_0パターン生成過程のモデルをどれだけ考慮に入れるかを意味した非負の定数である。
【０１６３】
以下に、
【０１６４】
【数５３】

$$\mathcal{I}_1(\boldsymbol{y}, \boldsymbol{s})$$

【０１６５】
を大きくするためのアルゴリズムについて述べる。上述した関連技術１のスペクトル特徴

量系列からのF_0 パターン予測方法と同様、

【０１６６】

【数５４】

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \simeq P(\hat{m}_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}})P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{m}_k,\hat{\boldsymbol{\gamma}}) \tag{86}$$

$$\hat{m}_k = \underset{m_k}{\operatorname{argmax}}\, P(m_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \tag{87}$$

【０１６７】

と近似することで以下の反復処理によりｙ、ｓを推定することができる（ステップ１と２の実行順序は任意）。

【０１６８】

（ステップ１）上述した関連技術１のスペクトル特徴量系列からのF_0 パターン予測方法の変換処理によりｙを初期設定する。

【０１６９】

（ステップ２）cを用いて^mを式(87) により求める。

【０１７０】

（ステップ３）ｙを固定し、上述した関連技術３のF0 パターン生成過程モデルパラメータ推定法のパラメータ推定アルゴリズム１によりｓを推定する。

【０１７１】

（ステップ４）ｓと^mを固定して以下の式によりｙを更新し、ステップ３に戻る。

【０１７２】

【数５５】

$$\boldsymbol{y} \leftarrow (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W} + \omega(\boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{G}^{\mathsf{T}})^{-1})^{-1}(\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{e}^{(q|c)} + \omega(\boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{G}^{\mathsf{T}})^{-1}\boldsymbol{G}\boldsymbol{\mu}) \tag{88}$$

ただし,

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_{\mathrm{p}} & \boldsymbol{G}_{\mathrm{a}} & \boldsymbol{I} \end{bmatrix} \tag{89}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathrm{p}} \\ \boldsymbol{\mu}_{\mathrm{a}} \\ \mu_{\mathrm{b}}\mathbf{1} \end{bmatrix} \tag{90}$$

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{\Upsilon}_{\mathrm{p}} & O & O \\ O & \boldsymbol{\Upsilon}_{\mathrm{a}} & O \\ O & O & \boldsymbol{\Upsilon}_{\mathrm{n}} \end{bmatrix} \tag{91}$$

【０１７３】

（ＧγG^T）$^{-1}$は大きなサイズの行列の逆行列であるが、以下に示すやり方で効率的に計算することができる。ＧγG^Tは

【０１７４】

【数５６】

$$\boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{G}^{\mathsf{T}} = \boldsymbol{G}_{\mathrm{p}}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{G}_{\mathrm{p}}^{\mathsf{T}} + \boldsymbol{G}_{\mathrm{a}}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{G}_{\mathrm{a}}^{\mathsf{T}} + \boldsymbol{\Upsilon}_{\mathrm{n}} \tag{92}$$

【０１７５】

であること、G^{-1}_p とG^{-1}_a が

【０１７６】

【数５７】

$$\boldsymbol{G}_{\mathrm{p}}^{-1}=\begin{bmatrix} f_0^{(p)} & & & & O \\ f_1^{(p)} & f_0^{(p)} & & & \\ f_2^{(p)} & f_1^{(p)} & f_0^{(p)} & & \\ & \ddots & \ddots & \ddots & \\ O & & f_2^{(p)} & f_1^{(p)} & f_0^{(p)} \end{bmatrix} =: \boldsymbol{F}_{\mathrm{p}} \tag{93}$$

$$\boldsymbol{G}_{\mathrm{a}}^{-1}=\begin{bmatrix} f_0^{(a)} & & & & O \\ f_1^{(a)} & f_0^{(a)} & & & \\ f_2^{(a)} & f_1^{(a)} & f_0^{(a)} & & \\ & \ddots & \ddots & \ddots & \\ O & & f_2^{(a)} & f_1^{(a)} & f_0^{(a)} \end{bmatrix} =: \boldsymbol{F}_{\mathrm{a}} \tag{94}$$

【０１７７】

のような下三角帯行列で近似できることより、（ＧγG^{T}）$^{-1}$は

【０１７８】

【数５８】

$$(\boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{G}^{\mathsf{T}})^{-1}=(\boldsymbol{F}_{\mathrm{p}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{F}_{\mathrm{p}}^{-\mathsf{T}}+\boldsymbol{F}_{\mathrm{a}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{F}_{\mathrm{a}}^{-\mathsf{T}}+\boldsymbol{\Upsilon}_{\mathrm{n}})^{-1} \tag{95}$$

【０１７９】

と書け、さらにWoodbury の公式

【０１８０】

【数５９】

$$(A+B)^{-1}=A^{-1}-A^{-1}(B^{-1}+A^{-1})^{-1}A^{-1} \tag{96}$$

【０１８１】

を式(95) 右辺に適用することで（ＧγG^{T}）$^{-1}$は

【０１８２】

【数６０】

$$(\boldsymbol{G}\boldsymbol{\Upsilon}\boldsymbol{G}^{\mathsf{T}})^{-1}=\boldsymbol{\Upsilon}_{\mathrm{n}}^{-1}-\boldsymbol{\Upsilon}_{\mathrm{n}}^{-1}((\boldsymbol{F}_{\mathrm{p}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{F}_{\mathrm{p}}^{-\mathsf{T}}+\boldsymbol{F}_{\mathrm{a}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{F}_{\mathrm{a}}^{-\mathsf{T}})^{-1}+\boldsymbol{\Upsilon}_{\mathrm{n}}^{-1})^{-1}\boldsymbol{\Upsilon}_{\mathrm{n}}^{-1} \tag{97}$$

【０１８３】

と書ける。さらにWoodbury の公式より

【０１８４】

【数６１】

$$\left(\boldsymbol{F}_{\mathrm{p}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{F}_{\mathrm{p}}^{-\mathsf{T}}+\boldsymbol{F}_{\mathrm{a}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{F}_{\mathrm{a}}^{-\mathsf{T}}\right)^{-1}$$

【０１８５】

は

【０１８６】

【数６２】

$$(\boldsymbol{F}_{\mathrm{p}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{p}}\boldsymbol{F}_{\mathrm{p}}^{-\mathsf{T}}+\boldsymbol{F}_{\mathrm{a}}^{-1}\boldsymbol{\Upsilon}_{\mathrm{a}}\boldsymbol{F}_{\mathrm{a}}^{-\mathsf{T}})^{-1}= \boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}}-\boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}}(\boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}}+\boldsymbol{F}_{\mathrm{a}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{F}_{\mathrm{a}})^{-1}\boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}} \quad (98)$$

【０１８７】

と書ける。式(93)、(94) より

【０１８８】

【数６３】

$$\boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}} \text{ と } \boldsymbol{F}_{\mathrm{a}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{F}_{\mathrm{a}}$$

【０１８９】

はいずれも帯行列になるので

【０１９０】

【数６４】

$$(\boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}}+\boldsymbol{F}_{\mathrm{a}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{F}_{\mathrm{a}})^{-1}\boldsymbol{a} \text{ または } (\boldsymbol{F}_{\mathrm{p}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{p}}^{-1}\boldsymbol{F}_{\mathrm{p}}+\boldsymbol{F}_{\mathrm{a}}^{\mathsf{T}}\boldsymbol{\Upsilon}_{\mathrm{a}}^{-1}\boldsymbol{F}_{\mathrm{a}})^{-1}\boldsymbol{A}$$

【０１９１】

の形の計算はCholesky 分解により効率的に計算することができる。ただし、a は任意のベクトル，A は任意の行列である。

【０１９２】

以上説明した原理を実現するために、本実施の形態では、図７に示すように、変換処理部５０は、特徴量抽出部５２と、基本周波数系列予測部５４と、正規分布系列予測部５６と、状態系列推定部５８と、基本周波数系列更新部６０と、収束判定部６２とを備えている。

【０１９３】

特徴量抽出部５２は、入力部１０によって受け付けた予測対象のソース音声の時系列データから、特徴量抽出部３２と同様に、ソース音声の各時刻ｋのスペクトグラム特徴量ベクトルc[k]を抽出する。

【０１９４】

基本周波数系列予測部５４は、第１モデルパラメータ学習部３６によって学習された第１確率分布のパラメータγと、特徴量抽出部５２によって抽出された各時刻ｋのスペクトル特徴量ベクトルc[k]とに基づいて、上述したF_0パターン予測方法の変換処理と同様に、上記式（１６）に従って、各時刻ｋの基本周波数ｙ［ｋ］を推定することにより、各時刻ｋの基本周波数ｙ［ｋ］を初期設定する。

【０１９５】

正規分布系列予測部５６は、第１モデルパラメータ学習部３６によって学習された第１確率分布のパラメータγと、特徴量抽出部５２によって抽出された各時刻ｋのスペクトル特徴量ベクトルc[k]とに基づいて、上記式（８７）に従って、各時刻ｋのスペクトル特徴量ベクトルc[k]を生成したらしい確率が最も高い正規分布のインデックス$\hat{}\ m_k$を推定する。

【０１９６】

状態系列推定部５８は、基本周波数系列予測部５４によって初期設定された、または状態系列推定部５８によって前回更新された各時刻ｋの基本周波数ｙ［ｋ］を固定して、上述した関連技術３のF_0 パターン生成過程モデルパラメータ推定法のパラメータ推定アルゴリズム１と同様に、上記式（３５）を局所最大化する状態系列ｓと各時刻ｋにおける状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータθとを、ＥＭアルゴリズムにより探索することにより、状態系列ｓを推定する。

【０１９７】

基本周波数系列更新部６０は、状態系列推定部５８によって推定された状態系列ｓと、

正規分布系列予測部５６によって推定された各時刻の正規分布のインデックスm_kに基づいて、上記式（８８）に従って、各時刻ｋの基本周波数ｙ［ｋ］を更新する。

【０１９８】

収束判定部６２は、予め定められた収束判定条件を満たすまで、状態系列推定部５８及び基本周波数系列更新部６０による各処理を繰り返させる。収束判定条件としては、例えば、予め定められた繰り返し回数に到達することである。

【０１９９】

収束判定条件を満たしたときに、最終的に得られた各時刻ｋの基本周波数ｙ［ｋ］を、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数の予測結果として、出力部９０により出力する。

【０２００】

＜基本周波数パターン予測装置の作用＞

次に、本実施の形態に係る基本周波数パターン予測装置１００の作用について説明する。まず、学習サンプルのソース音声の時系列データ及びターゲット音声の時系列データからなるパラレルデータが、基本周波数パターン予測装置１００に入力されると、基本周波数パターン予測装置１００において、図８に示す学習処理ルーチンが実行される。

【０２０１】

まず、ステップＳ１０１において、入力されたソース音声の時系列データを読み込み、

【０２０２】

各時刻ｋのスペクトル特徴量ベクトルｃ［ｋ］を抽出する。ステップＳ１０２において、入力されたターゲット音声の時系列データを読み込み、ターゲット音声の各時刻ｋにおける基本周波数y[k]を抽出し、また、基本周波数y[k]とその動的成分の結合ベクトルq[k]を抽出する。

【０２０３】

そして、ステップＳ１０３において、上記ステップＳ１０１で抽出された各時刻ｋのスペクトル特徴量ベクトルc[k]と、上記ステップＳ１０２で抽出された各時刻ｋの基本周波数の結合ベクトルｑ［ｋ］とに基づいて、上記式(1) のGMM のパラメータγを学習する。

【０２０４】

ステップＳ１０４では、状態系列ｓと、各時刻における状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータθとを初期設定する。

【０２０５】

そして、ステップＳ１０５において、上記式（４９）、式（５０）に従って、フレーズ成分、アクセント成分、ベースライン成分の条件付き期待値 $\bar{x}$ と、条件付き分散共分散行列Ｒとを更新する。

【０２０６】

次のステップＳ１０６では、上記ステップＳ１０４で初期設定された、又は後述するステップＳ１０７で前回更新されたパラメータθと、上記ステップＳ１０５で更新されたフレーズ成分、アクセント成分、ベースライン成分の条件付き期待値 $\bar{x}$ と、条件付き分散共分散行列Ｒとに基づいて、上記式（６３）式を用いて、Q(ｓ,θ) を最大にする状態系列$s=(s_1,\ldots,s_K)$をViterbiアルゴリズムにより求めて、状態系列ｓを更新する。

【０２０７】

ステップＳ１０７では、上記ステップＳ１０６で更新された状態系列ｓと、上記ステップＳ１０５で更新されたフレーズ成分、アクセント成分、ベースライン成分の条件付き期待値 $\bar{x}$ と、条件付き分散共分散行列Ｒとに基づいて、Q(ｓ,θ)の各変数に関する偏微分を0 と置くことにより、Q(ｓ,θ) を最大にするパラメータθを求めて、各時刻における状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータθとを更新する。

【０２０８】

ステップＳ１０８において、予め定められた収束判定条件を満たしたか否かを判定し、収束判定条件を満たしていない場合には、上記ステップＳ１０５へ戻る。一方、収束判定

条件を満たした場合には、ステップＳ１０９において、上記ステップＳ１０３で学習されたパラメータγ、上記ステップＳ１０７で最終的に得られたパラメータθとを、パラメータ記憶部４０に格納する。また、上記ステップＳ１０６で最終的に得られた状態系列ｓから、状態遷移確率φを求め、パラメータ記憶部４０に格納する。

【０２０９】

次に、予測対象のソース音声の時系列データが、基本周波数パターン予測装置１００に入力されると、基本周波数パターン予測装置１００において、図９に示す基本周波数パターン予測処理ルーチンが実行される。

【０２１０】

まず、ステップＳ１２１において、入力された予測対象のソース音声の時系列データを読み込み、各時刻ｋのスペクトル特徴量ベクトルｃ［ｋ］を抽出する。ステップＳ１２２において、パラメータ記憶部４０に記憶されたパラメータγと、上記ステップＳ１２１で抽出された各時刻のスペクトル特徴量ベクトルc[k]とに基づいて、上記式（１６）に従って、各時刻ｋの基本周波数ｙ［ｋ］を推定することにより、各時刻ｋの基本周波数ｙ［ｋ］を初期設定する。

【０２１１】

そして、ステップＳ１２３では、パラメータ記憶部４０に記憶されたパラメータγと、上記ステップＳ１２１で抽出された各時刻ｋのスペクトル特徴量ベクトルc[k]とに基づいて、上記式（８７）に従って、各時刻ｋのスペクトル特徴量ベクトルc[k]を生成したらしい確率が最も高い正規分布のインデックス$\hat{m}_k$を推定する。

【０２１２】

ステップＳ１２４では、上記ステップＳ１２２で初期設定された、または後述するステップＳ１２５で前回更新された各時刻ｋの基本周波数ｙ［ｋ］を固定して、上記式（３５）を局所最大化する状態系列ｓと各時刻ｋにおける状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータθとを、ＥＭアルゴリズムにより探索することにより、状態系列ｓを推定する。

【０２１３】

ステップＳ１２５では、上記ステップＳ１２４で推定された状態系列ｓと、上記ステップＳ１２３で推定された各時刻の正規分布のインデックス$\hat{m}_k$とに基づいて、上記式（８８）に従って、各時刻ｋの基本周波数ｙ［ｋ］を更新する。

【０２１４】

ステップＳ１２６では、予め定められた収束判定条件を満たしたか否かを判定し、収束判定条件を満たしていない場合には、ステップＳ１２４へ戻る。一方、収束判定条件を満たした場合には、ステップＳ１２７において、上記ステップＳ１２５で最終的に得られた各時刻ｋの基本周波数ｙ［ｋ］を、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数の予測結果として、出力部９０により出力し、基本周波数パターン予測処理ルーチンを終了する。

【０２１５】

以上説明したように、第１の実施の形態に係る基本周波数パターン予測装置によれば、ソース音声の各時刻のスペクトル特徴量ベクトルと、ターゲット音声の各時刻の基本周波数との間の関係をモデル化したＧＭＭである第１確率分布Ｐ(q[k]，c[k]｜γ)のパラメータγを学習し、基本周波数パターン生成過程をモデル化した第２確率分布Ｐ（ｙ，ｓ｜θ，φ）のパラメータθ、φを学習し、予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルから、第１確率分布Ｐ(q[k]，c[k]｜γ)と第２確率分布Ｐ（ｙ，ｓ｜θ，φ）との積を用いて表される規準を大きくするように、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数を予測することにより、F_0パターンの物理的な生成過程の制約を考慮しながらスペクトル特徴量系列に対応する最適なF_0パターンを推定することができる。

【０２１６】

［第２の実施の形態］

次に、本発明の第２の実施の形態に係る基本周波数パターン予測装置について説明する。なお、第１の実施の形態と同様の構成となる部分については同一符号を付して説明を省略する。

【０２１７】

第２の実施の形態では、第２の確率分布及びパラメータを推定する方法と、各時刻の基本周波数の予測方法とが第１の実施の形態と異なっている。

【０２１８】

第２の実施の形態に係る基本周波数パターン予測装置の学習部３０の第２モデルパラメータ学習部３８によるパラメータを学習する原理について説明する。

【０２１９】

まず、関連技術３のF_0パターン生成過程モデルパラメータ推定法のパラメータ推定アルゴリズムについて説明する。

【０２２０】

＜パラメータ推定アルゴリズム２＞

観測F_0系列ｙが与えられたもとで、モデルパラメータθとｏの事後確率Ｐ(ｏ，θ｜ｙ)の局所最適解を求める反復アルゴリズムを以下に示す。状態系列ｓを隠れ変数とし、事後確率Ｐ(ｏ，θ｜ｙ)が、

【０２２１】

【数６５】

$$P(\boldsymbol{o},\boldsymbol{\theta},\boldsymbol{s}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{o})P(\boldsymbol{o}|\boldsymbol{s},\boldsymbol{\theta})P(\boldsymbol{s})$$

【０２２２】

をｓについて周辺化することで得られる点に注意すると、Ｑ関数Ｑ(ｏ，θ，ｏ´，θ´)は

【０２２３】

【数６６】

$$\begin{aligned} Q(\boldsymbol{o},\boldsymbol{\theta},\boldsymbol{o}',\boldsymbol{\theta}') &= \sum_{\boldsymbol{s}} P(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')\log P(\boldsymbol{o},\boldsymbol{\theta},\boldsymbol{s}|\boldsymbol{y}) \\ &\overset{c}{=} \log P(\boldsymbol{y}|\boldsymbol{o}) + \sum_{\boldsymbol{s}} P(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')\log P(\boldsymbol{o}|\boldsymbol{s},\boldsymbol{\theta})P(\boldsymbol{s}) \end{aligned}$$

$$\log P(\boldsymbol{y}|\boldsymbol{o}) = \log \prod_{k=1}^{K} \mathcal{N}(y[k]; x[k], v_{\mathrm{n}}^2[k]) = -\sum_{k=1}^{K} \frac{C[k]}{2v_{\mathrm{n}}^2[k]} \tag{64}$$

$$C[k] = \left(y[k] - \sum_{i\in\{\mathrm{p,a,b}\}} \sum_{l} g_i[k-l+1]u_i[l] \right)^2 \tag{65}$$

$$\sum_{\boldsymbol{s}} P(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')\log P(\boldsymbol{\theta},\boldsymbol{o}|\boldsymbol{s}) = -\sum_{k=1}^{K}\sum_{t} P(s_k = t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}') \sum_{i\in\{\mathrm{p,a}\}} \frac{D_t^{(i)}[k]}{2v_t^{(i)2}} \tag{66}$$

$$D_t^{(i)}[k] = \left(u_i[k] - \mu_t^{(i)}[k] \right)^2 \tag{67}$$

と置ける。ただし、

【０２２４】

【数６７】

$$\overset{c}{=}$$

【０２２５】
は定数項を除いて等しいことを表す。また、$g_b[k] = \delta[k]$（クロネッカーのデルタ）である。よって、$P(s_k = t \mid y, o', \theta')$をForward-Backward アルゴリズムにより計算するステップ、oとθについてQ(o，θ，o′，θ′)を増加させるステップを繰り返すことで、P(o，θ｜y）が局所最大となる解を得ることができる。oはフレーズ・アクセント指令系列のペアであるため、Q(o，θ，o′，θ′)を増加させるステップにおいては、oの非負制約を考慮する必要がある。oの非負制約を満たしながらQ(o，θ，o′，θ′)を増加させるような更新則は以下により導くことができる。まず、Q(o，θ，o′，θ′)の下界はJensen の不等式より

【０２２６】
【数６８】

$$-\left(\sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l} g_i[k-l+1]u_i[l]\right)^2 \geq -\sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l}\frac{g_i^2[k-l+1]u_i^2[l]}{\lambda_{i,k,l}}, \tag{68}$$

【０２２７】
のように設計することができる。また、i，k，l は、

【０２２８】
【数６９】

$$0 < \lambda_{i,k,l} < 1, \sum_i \sum_l \lambda_{i,k,l} = 1$$

【０２２９】
を満たす任意の変数である。従ってQ関数の下界は、

【０２３０】
【数７０】

$$Q(\boldsymbol{o},\boldsymbol{\theta},\boldsymbol{o}',\boldsymbol{\theta}') \geq -\sum_{k=1}^{K}\frac{\tilde{C}[k]}{2\upsilon_{\mathrm{n}}^2[k]} - \sum_{k=1}^{K}\sum_{t} P(s_k = t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}') \sum_{i\in\{\mathrm{p,a}\}}\frac{D_t^{(i)}[k]}{2\upsilon_t^{(i)2}} \tag{69}$$

$$\tilde{C}[k] = y[k]^2 - 2y[k]\sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l} g_i[k-l+1]u_i[l] + \sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l}\frac{g_i^2[k-l+1]u_i^2[l]}{\lambda_{i,k,l}} \tag{70}$$

【０２３１】
と表される。この下界関数を$\lambda_{i,k,l} \geqq 0$ に関して最大化するステップとo に関して最大化するステップを交互に繰り返せばQ(o，θ，o′，θ′)を増加させることができる。いずれのステップの更新則も解析的に求めることができ、それぞれ

【０２３２】
【数７１】

$$\lambda_{i,k,l} = \frac{g_i[k-l+1]u_i[l]}{\sum_{i\in\{\mathrm{p,a,b}\}}\sum_{l'} g_i[k-l'+1]u_i[l']} \tag{71}$$

$$u_i[l] = \frac{\sum_t \frac{P(s_l = t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')\mu_t^{(i)}[l]}{\upsilon_t^{(i)2}} + \sum_k \frac{y[k]g_i[k-l+1]}{\upsilon_{\mathrm{n}}^2[k]}}{\sum_t \frac{P(s_l = t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')}{\upsilon_t^{(i)2}} + \sum_k \frac{g_i^2[k-l+1]}{\upsilon_{\mathrm{n}}^2[k]\lambda_{i,k,l}}} \tag{72}$$

【０２３３】
で表される。以上の反復が収束したあと、続けてθを更新する。更新式は、図１の場合、
【０２３４】
【数７２】

$$A^{(\mathrm{p})}[k]=\frac{\sum_{t\in \mathrm{p}_0}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')u_{\mathrm{p}}[k]}{{\upsilon_t^{(\mathrm{p})}}_2}}{\sum_{t\in \mathrm{p}_0}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')}{{\upsilon_t^{(\mathrm{p})}}_2}} \tag{73}$$

$$A_n^{(\mathrm{a})}=\frac{\sum_k\sum_{t\in \mathrm{a}_n}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')u_{\mathrm{a}}[k]}{{\upsilon_t^{(\mathrm{a})}}_2}}{\sum_k\sum_{t\in \mathrm{a}_n}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')}{{\upsilon_t^{(\mathrm{a})}}_2}} \tag{74}$$

【０２３５】
図２、３の場合、
【０２３６】
【数７３】

$$A_m^{(\mathrm{p})}=\frac{\sum_k\sum_{t\in \mathrm{p}_m}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')u_{\mathrm{p}}[k]}{{\upsilon_t^{(\mathrm{p})}}_2}}{\sum_k\sum_{t\in \mathrm{p}_m}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')}{{\upsilon_t^{(\mathrm{p})}}_2}} \tag{75}$$

$$A_n^{(\mathrm{a})}=\frac{\sum_k\sum_{t\in \mathrm{a}_n}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')u_{\mathrm{a}}[k]}{{\upsilon_t^{(\mathrm{a})}}_2}}{\sum_k\sum_{t\in \mathrm{a}_n}\frac{P(s_k=t|\boldsymbol{y},\boldsymbol{o}',\boldsymbol{\theta}')}{{\upsilon_t^{(\mathrm{a})}}_2}} \tag{76}$$

【０２３７】
である。これらの更新値をｏ´とθ´に代入したのちに、Ｐ(ｓ$_{k}$＝ｔ｜ｙ，ｏ´，θ´)の更新を再度行い、以後同様の処理を繰り返すことで事後確率Ｐ(ｏ，θ｜ｙ)を単調増加させることができる。
【０２３８】
以上の反復アルゴリズムが収束した後、上述したパラメータ推定アルゴリズム１のViterbi アルゴリズムにより求まる最適なｓを状態系列の推定結果とする。
【０２３９】
また、推定された状態系列ｓから、状態遷移確率φが求められる。
【０２４０】
以上説明した原理に従って、第２モデルパラメータ学習部３８は、基本周波数系列抽出部３４によって抽出された各時刻ｋの基本周波数y[k]に基づいて、各時刻ｋの基本周波数y[k]と、各時刻ｋにおける甲状軟骨の平行移動運動によって生じる基本周波数パターンを表すフレーズ指令ｕ$_{p}$［ｋ］及び甲状軟骨の回転運動によって生じる基本周波数パターン

を表すアクセント指令u_a［ｋ］のペアからなる指令関数ｏ［ｋ］との組み合わせの確率分布である第２確率分布のパラメータθ、φを学習する。

【０２４１】

第２の実施の形態における変換処理部５０は、予測対象のソース音声の時系列データを入力として、ソース音声の時系列データから抽出される各時刻ｋのスペクトル特徴量ベクトルｃ［ｋ］と、第１モデルパラメータ学習部３６によって学習された第１確率分布のパラメータγと、第２モデルパラメータ学習部３８によって学習された第２確率分布のパラメータθ、φとに基づいて、第１確率分布と第２確率分布とを用いて表される規準を大きくするように、各時刻ｋの基本周波数ｙ［ｋ］と、各時刻ｋのフレーズ指令及びアクセント指令のペアからなる指令関数ｏ［ｋ］とを推定することにより、予測対象のソース音声に対応するターゲット音声の各時刻ｋの基本周波数ｙ［ｋ］を予測する。

【０２４２】

ここで、予測対象のソース音声に対応するターゲット音声の各時刻ｋの基本周波数ｙ［ｋ］を予測する原理について説明する。

【０２４３】

＜変換処理＞

学習処理で学習したパラメータ＾γ、＾θ、＾ψと、予測対象のソース音声の特徴量系列$\mathrm{c} = \{\mathrm{c}[\mathrm{k}]\}_{\mathrm{k}=1}^{\mathrm{K}}$を用いて、式(9)と式(45)の積

【０２４４】

【数７４】

$$\mathcal{I}_2(\boldsymbol{y},\boldsymbol{o}) := P(\boldsymbol{q}|\boldsymbol{c},\hat{\boldsymbol{\gamma}})P(\boldsymbol{y},\boldsymbol{o}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\psi}})^{\omega} \tag{82}$$

【０２４５】

ができるだけ大きくなるようにｙ、ｏを推定する。ただし、ｑ＝Ｗｙであり、

【０２４６】

【数７５】

$$P(\boldsymbol{y},\boldsymbol{o}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\psi}}) = P(\boldsymbol{y}|\boldsymbol{o})\sum_{\boldsymbol{s}} P(\boldsymbol{o}|\boldsymbol{s},\hat{\boldsymbol{\theta}})P(\boldsymbol{s}|\hat{\boldsymbol{\psi}}) \tag{83}$$

$$P(\boldsymbol{y}|\boldsymbol{o}) = \prod_{k} \mathcal{N}(y[k]; x[k], \upsilon_{\mathrm{n}}^2[k]) \tag{84}$$

$$x[k] = g_{\mathrm{p}}[k] * u_{\mathrm{p}}[k] + g_{\mathrm{a}}[k] * u_{\mathrm{a}}[k] + \mu_{\mathrm{b}} \tag{85}$$

である。

【０２４７】

次に、

【０２４８】

【数７６】

$$\mathcal{I}_2(\boldsymbol{y},\boldsymbol{o})$$

【０２４９】

を大きくするためのアルゴリズムについて述べる。上述した関連技術１のスペクトル特徴量系列からのF_0パターン予測方法と同様に、

【０２５０】

【数７７】

$$P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \simeq P(\hat{m}_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}})P(\boldsymbol{q}[k]|\boldsymbol{c}[k],\hat{m}_k,\hat{\boldsymbol{\gamma}}) \tag{99}$$

$$\hat{m}_k = \underset{m_k}{\operatorname{argmax}}\, P(m_k|\boldsymbol{c}[k],\hat{\boldsymbol{\gamma}}) \tag{100}$$

【０２５１】

と近似することで以下の反復処理によりｙ、ｏを推定することができる（ステップ1 と2の実行順序は任意）。

【０２５２】

（ステップ１）上述した関連技術１のスペクトル特徴量系列からのF_0 パターン予測方法の変換処理によりｙを初期設定する。

【０２５３】

（ステップ２）cを用いて^mを式(１００）により求める。

【０２５４】

（ステップ３）ｙを固定し、上述した関連技術３のF_0 パターン生成過程モデルパラメータ推定法のパラメータ推定アルゴリズム２によりｏを推定する。

【０２５５】

（ステップ４）ｏと＾ｍを固定して以下の式によりｙを更新し、ステップ３に戻る。

【０２５６】

【数７８】

$$\boldsymbol{y} \leftarrow (\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{W} + \omega\boldsymbol{\Sigma}_{\mathrm{n}}^{-1})^{-1}(\boldsymbol{W}^{\mathsf{T}}\boldsymbol{D}^{(q|c)-1}\boldsymbol{e}^{(q|c)} + \omega\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{G}\boldsymbol{u}) \tag{101}$$

ただし，

$$\boldsymbol{G} = \begin{bmatrix}\boldsymbol{G}_{\mathrm{p}} & \boldsymbol{G}_{\mathrm{a}} & \boldsymbol{I}\end{bmatrix} \tag{102}$$

$$\boldsymbol{u} = \begin{bmatrix}\boldsymbol{u}_{\mathrm{p}} \\ \boldsymbol{u}_{\mathrm{a}} \\ \boldsymbol{u}_{\mathrm{b}}\end{bmatrix} \tag{103}$$

$$\boldsymbol{u}_{\mathrm{p}} = (u_{\mathrm{p}}[1],\ldots,u_{\mathrm{p}}[K])^{\mathsf{T}} \tag{104}$$

$$\boldsymbol{u}_{\mathrm{a}} = (u_{\mathrm{a}}[1],\ldots,u_{\mathrm{a}}[K])^{\mathsf{T}} \tag{105}$$

$$\boldsymbol{u}_{\mathrm{b}} = (u_{\mathrm{b}}[1],\ldots,u_{\mathrm{b}}[K])^{\mathsf{T}} \tag{106}$$

【０２５７】

以上説明した原理を実現するために、第２の実施の形態では、図１０に示すように、変換処理部５０は、特徴量抽出部５２と、基本周波数系列予測部５４と、正規分布系列予測部２５６と、指令系列推定部２５８と、基本周波数系列更新部２６０と、収束判定部６２とを備えている。

【０２５８】

正規分布系列予測部２５６は、第１モデルパラメータ学習部３６によって学習された第１確率分布のパラメータγと、特徴量抽出部５２によって抽出された各時刻のスペクトル特徴量ベクトルc[k]とに基づいて、上記式（１００）に従って、各時刻ｋのスペクトル特徴量ベクトルc[k]を生成したらしい確率が最も高い正規分布のインデックス＾mkを推定する。

【０２５９】

指令系列推定部２５８は、基本周波数系列予測部５４によって初期設定された、または状態系列推定部５８によって前回更新された各時刻ｋの基本周波数ｙ［ｋ］を固定して、上述した関連技術３のF0 パターン生成過程モデルパラメータ推定法のパラメータ推定アルゴリズム２と同様に、事後確率Ｐ（ｏ，θ｜ｙ）を局所最大化する指令系列ｏを推定する。

【０２６０】
基本周波数系列更新部２６０は、指令系列推定部２５８によって推定された指令系列ｏと、正規分布系列予測部５６によって推定された各時刻の正規分布のインデックス＾ｍｋとに基づいて、上記式（１０１）に従って、各時刻ｋの基本周波数ｙ［ｋ］を更新する。
【０２６１】
収束判定部６２は、予め定められた収束判定条件を満たすまで、指令系列推定部２５８及び基本周波数系列更新部２６０による各処理を繰り返させる。収束判定条件としては、例えば、予め定められた繰り返し回数に到達することである。
【０２６２】
収束判定条件を満たしたときに、最終的に得られた各時刻ｋの基本周波数ｙ［ｋ］を、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数の予測結果として、出力部９０により出力する。
【０２６３】
＜基本周波数パターン予測装置の作用＞
【０２６４】
次に、第２の実施の形態に係る基本周波数パターン予測装置１００の作用について説明する。なお、第１の実施の形態と同様の処理については、同一符号を付して詳細な説明を省略する。
【０２６５】
まず、学習サンプルのソース音声の時系列データ及びターゲット音声の時系列データからなるパラレルデータが、基本周波数パターン予測装置１００に入力されると、基本周波数パターン予測装置１００において、図１１に示す学習処理ルーチンが実行される。
【０２６６】
まず、ステップＳ１０１において、入力されたソース音声の時系列データを読み込み、各時刻ｋのスペクトル特徴量ベクトルｃ［ｋ］を抽出する。ステップＳ１０２において、入力されたターゲット音声の時系列データを読み込み、ターゲット音声の各時刻ｋにおける基本周波数y[k]を抽出し、また、各時刻ｋの基本周波数y[k]とその動的成分の結合ベクトルq[k]を抽出する。
【０２６７】
そして、ステップＳ１０３において、上記ステップＳ１０１で抽出された各時刻ｋのスペクトル特徴量ベクトルc[k]と、上記ステップＳ１０２で抽出された各時刻ｋの基本周波数の結合ベクトルｑ［ｋ］とに基づいて、上記式(1) のGMM のパラメータγを学習する。
【０２６８】
ステップＳ２００では、指令系列ｏと、各時刻における状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータθとを初期設定する。また、ターゲット音声の時系列データに基づいて、有声区間、無声区間を特定し、各時刻ｋの基本周波数の不確かさの程度v_n^2［ｋ］を推定する。
【０２６９】
そして、ステップＳ２０１において、上記ステップＳ２００で設定された指令系列oの初期値、または後述するステップＳ２０３で前回更新された指令系列oに基づいて、（k,t）の全ての組み合わせについて、事後確率Ｐ(s_k=t|y,o′,θ′)を更新する。
【０２７０】
ステップＳ２０２では、上記ステップＳ２００で設定された指令系列oの初期値、または後述するステップＳ２０３で前回更新された指令系列oに基づいて、（ｋ、ｌ）の全ての組み合わせについて、上記の式（７１）に従って、補助変数$\lambda_{p,k,l}$、$\lambda_{a,k,l}$、$\lambda_{b,k,l}$を算出して更新する。
【０２７１】
次のステップＳ２０３では、上記ステップＳ１０２で抽出されたされた基本周波数系列ｙと、上記ステップＳ２００で算出された各時刻ｋの不確かさの程度v_n^2［ｋ］と、上記ステップＳ２０１で更新された事後確率Ｐ(s_k=t|y,o′,θ′)と、上記ステップＳ２０２で

更新された補助変数$\lambda_{p,k,l}$、$\lambda_{a,k,l}$、$\lambda_{b,k,l}$とに基づいて、上記式（７２）に従って、非負値である各時刻ｌのフレーズ指令$u_p[l]$及びアクセント指令$u_a[l]$からなる指令系列oとベース成分u_bとを更新する。

【０２７２】

次のステップＳ２０４では、収束条件として、繰り返し回数ｓが、Ｓに到達したか否かを判定し、繰り返し回数ｓがＳに到達していない場合には、収束条件を満足していないと判断して、上記ステップＳ２０２へ戻る。一方、繰り返し回数ｓがＳに到達した場合には、収束条件を満足したと判断し、ステップＳ２０５で、上記ステップＳ２０３で更新された各時刻ｋのフレーズ指令$u_p[k]$及びアクセント指令$u_a[k]$と、上記ステップＳ２０１で更新された事後確率$P(s_k=t|y,o',\theta')$とに基づいて、上記式（７３）、式（７４）、又は式（７５）、式（７６）に従って、各時刻ｋのフレーズ指令の振幅$A^{(p)}[k]$、及び各位置ｎのアクセント指令の振幅$A_a^{(a)}$を更新することにより、各時刻における状態に応じたフレーズ指令の振幅及び各アクセント指令の振幅を表すパラメータθを更新する。

【０２７３】

ステップＳ２０６において、予め定められた収束判定条件を満たしたか否かを判定し、収束判定条件を満たしていない場合には、上記ステップＳ２０１へ戻る。一方、収束判定条件を満たした場合には、ステップＳ２０７において、

【０２７４】

上記ステップＳ２０３で最終的に更新された指令系列ｏに基づいて、Viterbi アルゴリズムにより、状態系列ｓを推定する。また、推定された状態系列ｓから、状態遷移確率φを求める。

【０２７５】

そして、ステップＳ２０８において、上記ステップＳ１０３で学習されたパラメータγ、上記ステップＳ２０５で最終的に得られたパラメータθと、上記ステップＳ１０６で得られた状態遷移確率φとを、パラメータ記憶部４０に格納する。

【０２７６】

次に、予測対象のソース音声の時系列データが、基本周波数パターン予測装置１００に入力されると、基本周波数パターン予測装置１００において、図１２に示す基本周波数パターン予測処理ルーチンが実行される。

【０２７７】

まず、ステップＳ１２１において、入力された予測対象のソース音声の時系列データを読み込み、各時刻ｋのスペクトル特徴量ベクトルｃ［ｋ］を抽出する。ステップＳ１２２において、パラメータ記憶部４０に記憶されたパラメータγと、上記ステップＳ１２１で抽出された各時刻のスペクトル特徴量ベクトルc[k]とに基づいて、各時刻の基本周波数ｙ［ｋ］を初期設定する。

【０２７８】

そして、ステップＳ１２３では、パラメータ記憶部４０に記憶されたパラメータγと、上記ステップＳ１２１で抽出された各時刻のスペクトル特徴量ベクトルc[k]とに基づいて、各時刻ｋの正規分布のインデックス$\hat{m}_k$を推定する。

【０２７９】

ステップＳ２２１では、上記ステップＳ１２２で初期設定された、または後述するステップＳ１２５で前回更新された各時刻ｋの基本周波数ｙ［ｋ］を固定して、上記ステップＳ２０１～ステップＳ２０６と同様に、事後確率Ｐ（ｏ，θ｜ｙ）を局所最大化する指令系列ｏを推定する。

【０２８０】

そして、ステップＳ２２２において、上記ステップＳ２２１で推定された指令系列ｏと、上記ステップＳ１２３で推定された各時刻の正規分布のインデックス$\hat{m}_k$とに基づいて、上記式（１０１）に従って、各時刻ｋの基本周波数ｙ［ｋ］を更新する。

【０２８１】

ステップＳ１２６では、予め定められた収束判定条件を満たしたか否かを判定し、収束判定条件を満たしていない場合には、ステップＳ２２１へ戻る。一方、収束判定条件を満たした場合には、ステップＳ１２７において、上記ステップＳ２２２で最終的に得られた各時刻ｋの基本周波数ｙ［ｋ］を、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数の予測結果として、出力部９０により出力し、基本周波数パターン予測処理ルーチンを終了する。

【０２８２】

以上説明したように、第２の実施の形態に係る基本周波数パターン予測装置によれば、ソース音声の各時刻のスペクトル特徴量ベクトルと、ターゲット音声の各時刻の基本周波数との間の関係をモデル化したＧＭＭである第１確率分布Ｐ(q[k]，c[k]｜γ)のパラメータγを学習し、基本周波数パターン生成過程をモデル化した第２確率分布Ｐ（ｙ，ｏ｜θ，φ）のパラメータθ、φを学習し、予測対象のソース音声の時系列データから抽出される各時刻のスペクトル特徴量ベクトルから、第１確率分布Ｐ(q[k]，c[k]｜γ)と第２確率分布Ｐ（ｙ，ｏ｜θ，φ）との積を用いて表される規準を大きくするように、予測対象のソース音声に対応するターゲット音声の各時刻の基本周波数を予測することにより、F_0パターンの物理的な生成過程の制約を考慮しながらスペクトル特徴量系列に対応する最適なF_0パターンを推定することができる。

【０２８３】

＜実験＞

【０２８４】

図１３に示すF_0パターンの音声データに対し、上述した従来手法である関連技術１のスペクトル特徴量系列からのF_0 パターン予測方法と、本発明の第１の実施の形態に係る手法とによりスペクトル特徴量系列からF_0 パターンの予測を行う実験を行った。図１４に、両手法により予測されたF_0パターンを示す。図１４では、実線が、従来手法による音声特徴量系列からのF_0 パターンの予測結果の例を示し、点線が、第１の実施の形態に係る手法による音声特徴量系列からのF_0 パターンの予測結果の例を示す。

【０２８５】

図１３のF_0パターンとの近さがF_0パターンの良さの指標になる。そこで、それぞれの手法で得られたF_0パターンと、図１３のF_0パターンとのコサイン距離（1 に近いほど近いこと意味する）を測ったところ、従来手法が0.55、第１の実施の形態に係る手法が0.59 であった。このことから、本発明の第１の実施の形態の手法の、従来手法に対する優位性が示された。

【０２８６】

なお、本発明は、上述した実施形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

【０２８７】

例えば、上述の基本周波数パターン予測装置は、内部にコンピュータシステムを有しているが、「コンピュータシステム」は、ＷＷＷシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

【０２８８】

また、本願明細書中において、プログラムが予めインストールされている実施形態として説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。

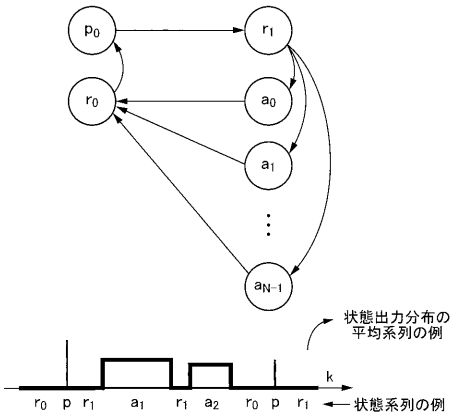
【符号の説明】

【０２８９】

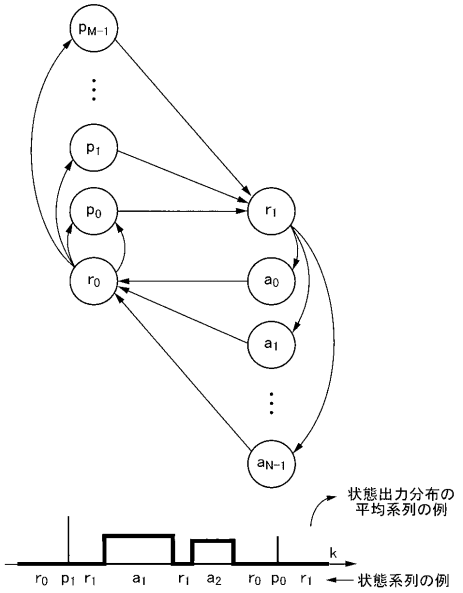
１０　入力部
２０　演算部
３０　学習部
３２　特徴量抽出部
３４　基本周波数系列抽出部

３６　第１モデルパラメータ学習部
３８　第２モデルパラメータ学習部
４０　パラメータ記憶部
５０　変換処理部
５２　特徴量抽出部
５４　基本周波数系列予測部
５６　正規分布系列予測部
５８　状態系列推定部
６０　基本周波数系列更新部
６２　収束判定部
９０　出力部
１００　基本周波数パターン予測装置
２５６　正規分布系列予測部
２５８　指令系列推定部
２６０　基本周波数系列更新部

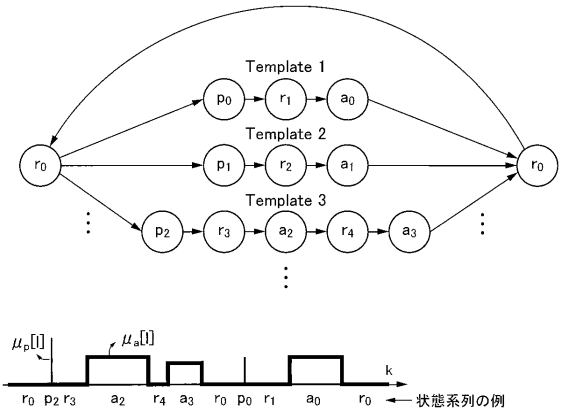
【図１】



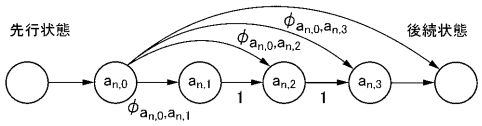
【図２】



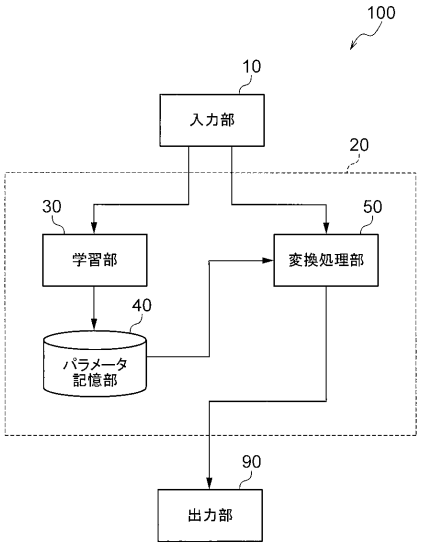
【図３】


Template 1
Template 2
Template 3
状態系列の例

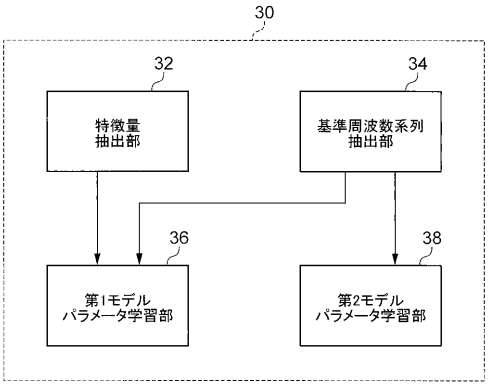
【図４】


先行状態
後続状態

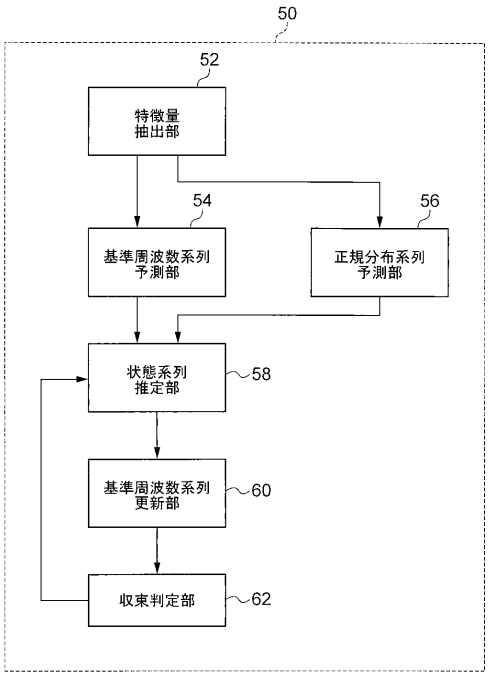
【図５】


100
入力部
学習部
変換処理部
パラメータ記憶部
出力部

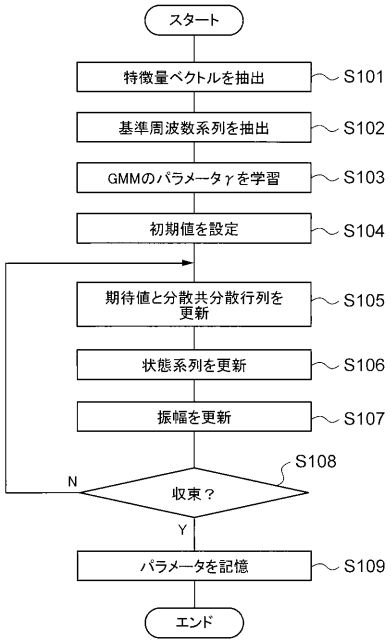
【図６】


特徴量抽出部
基準周波数系列抽出部
第1モデルパラメータ学習部
第2モデルパラメータ学習部

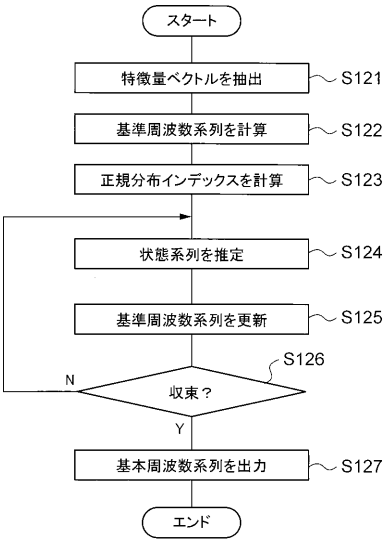
【図７】


特徴量抽出部
基準周波数系列予測部
正規分布系列予測部
状態系列推定部
基準周波数系列更新部
収束判定部

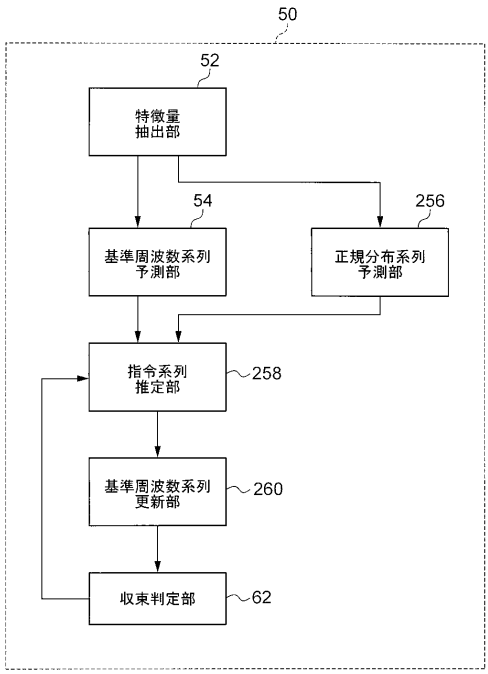
【図８】


スタート
特徴量ベクトルを抽出
S101
基準周波数系列を抽出
S102
GMMのパラメータγを学習
S103
初期値を設定
S104
期待値と分散共分散行列を更新
S105
状態系列を更新
S106
振幅を更新
S107
S108
N
収束？
Y
パラメータを記憶
S109
エンド

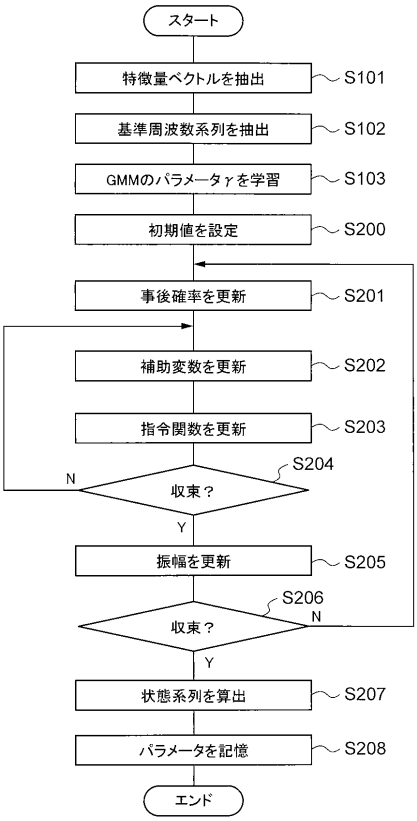
【図９】


スタート
特徴量ベクトルを抽出
S121
基準周波数系列を計算
S122
正規分布インデックスを計算
S123
状態系列を推定
S124
基準周波数系列を更新
S125
S126
N
収束？
Y
基本周波数系列を出力
S127
エンド

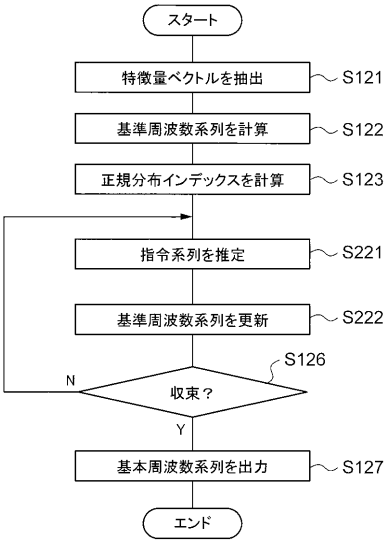
【図１０】


50
52
特徴量抽出部
54
256
基準周波数系列予測部
正規分布系列予測部
指令系列推定部
258
基準周波数系列更新部
260
収束判定部
62

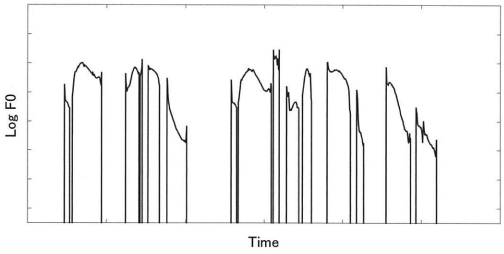
【図１１】


スタート
特徴量ベクトルを抽出
S101
基準周波数系列を抽出
S102
GMMのパラメータγを学習
S103
初期値を設定
S200
事後確率を更新
S201
補助変数を更新
S202
指令関数を更新
S203
S204
N
収束？
Y
振幅を更新
S205
S206
収束？
N
Y
状態系列を算出
S207
パラメータを記憶
S208
エンド

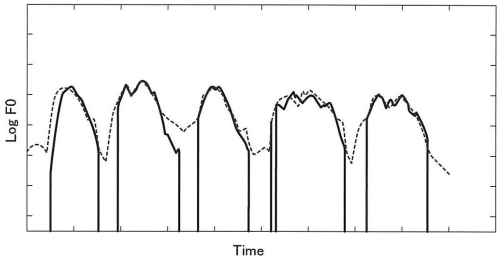
【図１２】


スタート
特徴量ベクトルを抽出
S121
基準周波数系列を計算
S122
正規分布インデックスを計算
S123
指令系列を推定
S221
基準周波数系列を更新
S222
S126
N
収束？
Y
基本周波数系列を出力
S127
エンド

【図１３】


Log F0
Time

【図１４】


Log F0
Time

フロントページの続き

(72)発明者　戸田　智基
奈良県生駒市高山町８９１６－５　国立大学法人奈良先端科学技術大学院大学内
(72)発明者　中村　哲
奈良県生駒市高山町８９１６－５　国立大学法人奈良先端科学技術大学院大学内

審査官　上田　雄

(56)参考文献　特開２０１５－０４１０８１（ＪＰ，Ａ）
国際公開第２０１０／１３７３８５（WO，Ａ1）
特開２０１３－１７１１９６（ＪＰ，Ａ）
特開２０１５－０４１００４（ＪＰ，Ａ）
特開２０１４－１３４７３０（ＪＰ，Ａ）
国際公開第２０１３／１３２９５９（WO，Ａ1）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２５／９０
Ｇ１０Ｌ　１３／００－１３／１０