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

特許第6436806号
(P6436806)

(45)発行日　平成30年12月12日(2018.12.12)　(24)登録日　平成30年11月22日(2018.11.22)

(51)Int.Cl.　F I
G10L 13/10 (2013.01)　G10L 13/10 113B
G10L 13/06 (2013.01)　G10L 13/06 210

請求項の数 6　(全 21 頁)

(21)出願番号　特願2015-19009(P2015-19009)
(22)出願日　平成27年2月3日(2015.2.3)
(65)公開番号　特開2016-142936(P2016-142936A)
(43)公開日　平成28年8月8日(2016.8.8)
審査請求日　平成29年3月22日(2017.3.22)

(73)特許権者　000233169
株式会社日立超エル・エス・アイ・システムズ
東京都立川市緑町７番地１
(74)代理人　100091096
弁理士　平木　祐輔
(74)代理人　100105463
弁理士　関谷　三男
(74)代理人　100102576
弁理士　渡辺　敏章
(72)発明者　孫　慶華
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内

審査官　上田　雄

最終頁に続く

(54)【発明の名称】音声合成用データ作成方法、及び音声合成用データ作成装置

(57)【特許請求の範囲】
【請求項１】
メモリから各種プログラムを読みだして実行し、音声合成処理で用いられる音声合成用データを作成するプロセッサが、第１種音声のコーパスデータの入力を受け付け、当該コーパスデータに含まれる音声データから第１韻律を抽出するステップと、
前記プロセッサが、予め用意されている第２種音声の韻律モデルを前記コーパスデータに含まれるテキストに対して適用し、前記テキストに対応する第２韻律を予測するステップと、
前記プロセッサが、前記第１韻律と前記第２韻律の差分値を算出するステップと、
前記プロセッサが、前記テキストに含まれる文字のうち、前記差分値が所定の閾値よりも大きい文字を判定するステップと、
前記プロセッサが、前記判定するステップの結果に基づいて、前記第１韻律と前記第２韻律との違いに起因する特徴テキスト部分に対応するデータを抽出するステップと、
を含み、
前記第１種音声は口語調音声であり、前記第２種音声は読み上げ調音声であり、
前記第２種音声の韻律モデルは、読み上げ韻律・音韻予測モデルであり、
前記抽出された特徴テキスト部分に対応するデータは、前記特徴テキスト部分の音声波形データ、韻律・音韻情報、及びテキストデータを含み、
さらに、前記プロセッサが、前記特徴テキスト部分のテキストデータを用いて、与えられるテキストデータにおける口語調表現を抽出するためのルールを生成するステップを含

むことを特徴とする音声合成用データ作成方法。
【請求項２】
請求項１において、
前記プロセッサは、前記特徴テキスト部分のテキストデータに加えて、当該テキストデータが含まれる口語調テキストの前後のコンテキスト情報を用いて前記ルールを生成することを特徴とする音声合成用データ作成方法。
【請求項３】
メモリから各種プログラムを読みだして実行し、音声合成処理で用いられる音声合成用データを作成するプロセッサが、口語調音声のコーパスデータの入力を受け付け、当該コーパスデータに含まれる音声データから口語調韻律データを抽出するステップと、
前記プロセッサが、予め用意されている読み上げ調の韻律モデルを前記コーパスデータに含まれるテキストに対して適用し、前記テキストに対応する読み上げ調韻律データを予測するステップと、
前記プロセッサが、前記口語調韻律データと前記読み上げ調韻律データの差分値を算出するステップと、
前記プロセッサが、前記差分値に基づいて、前記テキストのセグメントに対して、当該セグメントの口語調の程度を示す口語調度を算出し、前記口語調韻律データに付与するステップと、
前記プロセッサが、前記口語調度が付与された前記口語調韻律データを用いて、前記音声合成用データを生成するステップと、
を含むことを特徴とする音声合成用データ作成方法。
【請求項４】
請求項３において、
前記音声合成用データを生成するステップは、前記プロセッサが、前記口語調度が付与された前記口語調韻律データを用いて、入力テキストの口語調度を予測するための統計モデルである口語調度予測モデルを生成することを含むことを特徴とする音声合成用データ作成方法。
【請求項５】
各種プログラムを格納するメモリと、
前記メモリから前記各種プログラムを読みだして実行し、音声合成処理で用いられる音声合成用データを作成するプロセッサと、を有し、
前記プロセッサは、
第１種音声のコーパスデータの入力を受け付け、当該コーパスデータに含まれる音声データから第１韻律を抽出する処理と、
予め用意されている第２種音声の韻律モデルを前記コーパスデータに含まれるテキストに対して適用し、前記テキストに対応する第２韻律を予測する処理と、
前記第１韻律と前記第２韻律の差分を算出する処理と、
前記テキストに含まれる文字のうち、前記差分が所定の閾値よりも大きい文字を判定する処理と、
前記第１韻律と前記第２韻律との違いに起因する特徴テキスト部分を抽出する処理と、
を実行し、
前記第１種音声は口語調音声であり、前記第２種音声は読み上げ調音声であり、
前記第２種音声の韻律モデルは、読み上げ韻律・音韻予測モデルであり、
前記抽出された特徴テキスト部分に対応するデータは、前記特徴テキスト部分の音声波形データ、韻律・音韻情報、及びテキストデータを含み、
前記プロセッサは、さらに、前記特徴テキスト部分のテキストデータを用いて、与えられるテキストデータにおける口語調表現を抽出するためのルールを生成する処理を実行する音声合成用データ作成装置。
【請求項６】

請求項５において、
前記プロセッサは、前記特徴テキスト部分のテキストデータに加えて、当該テキストデータが含まれる口語調テキストの前後のコンテキスト情報を用いて前記ルールを生成することを特徴とする音声合成用データ作成装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声合成用データ作成方法、及び音声合成用データ作成装置に関し、例えば、収録した口語調音声から、読み上げ調との差異（口語調度）を自動的に測定する技術に関するものである。
【背景技術】
【０００２】
テキストを音声に変換して読み上げるテキスト音声合成技術、およびそれを利用したテキスト音声合成システムがある。このような技術、システムの応用先として、例えば、カーナビゲーションでのガイド音声、携帯電話・スマートフォンでのメール読み上げや音声対話インタフェース、視覚障碍者向けのスクリーンリーダー、電子書籍の読み上げ機能などが存在する。
【０００３】
近年、音声合成技術はその重要性が増大している。例えば、カーナビゲーションや携帯電話・スマートフォンの普及に加えて障碍者や高齢者を対象としたユニバーサルデザインの志向によって音声を使ったユーザインターフェースが今まで以上に使われるようになってきている。また、近年では、電子書籍端末の普及が始まり、音声読上げに必須な音声合成技術のニーズも拡大している。
【０００４】
最近、音声合成技術は、カーナビや携帯電話などの音声対話処理で多く用いられるようになってきている。加えて、そのような応用例では、ユーザが会話を自然に進められるように、話し言葉（以降、口語調）での音声を合成したいというニーズが増えてきている。
【０００５】
従来の音声合成技術であっても、合成音声の韻律（抑揚やリズム、強さなど）を工夫することで口語調の音声を生成できるものは存在している。例えば、標準的な発話特性を有する韻律特徴（読み上げ調韻律）に、感情や個性、発話スタイルなどの発話特性の違いに応じて補正するため修正韻律（口語調韻律）を重畳することで、口語調を含む多様な韻律を実現する手法が提案されている（特許文献１参照）。
【０００６】
また、入力されたテキストについて、口語調音声の特徴を担う部分（口語表現部分）とそれ以外の部分を分けて、前者に関しては口語調音声コーパスから得られた韻律や音声を用いて合成し、後者に関しては従来読み上げ調音声合成手法で合成するという技術も考案されている（特許文献２参照）。このような手法では、従来培ってきた読み上げ音声合成技術を最大限に生かしたうえ、自然性が高く、安定した口語調音声合成が実現できると考えられる。
【先行技術文献】
【特許文献】
【０００７】
【特許文献１】特開２００３－３３７５９２号公報
【特許文献２】特開２０１４－０６２９７０号公報
【発明の概要】
【発明が解決しようとする課題】
【０００８】
しかし、特許文献１の技術では、音声合成時に大きく韻律を制御する（変形させる）ため、合成音声の品質が劣化するという課題が指摘されている。

【０００９】
また、特許文献１及び２に開示の何れの手法においても、口語調での音声合成を実現するために、収録した口語調音声コーパスから、口語調音声データの作成が必須となる。例えば、特許文献２では、収録した口語音声から、熟練したラベリング作業者の経験に基づいて、手作業で口語調表現を抽出していた。しかし、この作業は、非常に時間がかかるうえ、作業者ごとに判定基準が異なり、抽出した口語調表現の一貫性を保つことが難しい（例えば、同じ音声に対しても、作業者ごとに、抽出された口語調表現が異なる。）。さらに、特許文献１の音声合成方法を用いる場合、音声のセグメントごとに、口語調表現らしさをより正確に定量的に評価することが望まれる。しかし、作業者の手作業ではこのような定量的な評価はほぼ不可能である。

【００１０】
本発明はこのような状況に鑑みてなされたものであり、口語調音声のように、読み上げ音声（平静音声）の特徴とは異なる韻律や声質の特徴を持つ音声に対して、その特徴部分のデータを自動的に抽出する技術を提供するものである。

【課題を解決するための手段】

【００１１】
上記課題を解決するために、本発明では、まず、予め用意されている第２種音声の韻律モデルをコーパスデータに含まれるテキストに対して適用し、テキストに対応する第２韻律を予測し、当該コーパスに含まれる音声データから抽出された第１韻律と第２韻律の差分値を算出する。次に、コーパスデータのテキストに含まれる文字のうち、差分値が所定の閾値よりも大きい文字を判定する。そして、この判定結果に基づいて、第１韻律と第２韻律との違いに起因する特徴テキスト部分に対応するデータを抽出する。

【００１２】
本発明に関連する更なる特徴は、本明細書の記述、添付図面から明らかになるものである。また、本発明の態様は、要素及び多様な要素の組み合わせ及び以降の詳細な記述と添付される特許請求の範囲の様態により達成され実現される。

【００１３】
本明細書の記述は典型的な例示に過ぎず、本発明の特許請求の範囲又は適用例を如何なる意味に於いても限定するものではないことを理解する必要がある。

【発明の効果】

【００１４】
本発明によれば、口語調音声のように、読み上げ音声（平静音声）の特徴とは異なる韻律や声質の特徴を持つ音声に対して、その特徴部分のデータを自動的に抽出することができるようになる。

【図面の簡単な説明】

【００１５】
【図１】本発明の実施形態による音声合成システムのハードウェア構成例を示す図である。
【図２】本発明の第１の実施形態による音声合成システムの機能ブロックを示す図である。
【図３】本発明の実施形態による読み上げ調韻律・音韻予測部によって予測された韻律特徴量の例を示す図である。
【図４】本発明の実施形態による韻律・音韻特徴抽出部によって抽出された韻律特徴量の例を示す図である。
【図５】本発明の実施形態による口語調度算出部によって算出された韻律差分値の例を示す図である。
【図６】本発明の第１の実施形態による口語調表現部分の自動抽出部によって計算された音節（モーラ）レベルでの口語調度の例を示す図である。
【図７】本発明の第１の実施形態による口語調表現部分の自動抽出部によって計算された形態素レベルでの口語調度の例を示す図である。

【図８】本発明の第１の実施形態による口語調表現部分の自動抽出部によって計算されたアクセント句レベルでの口語調度の例を示す図である。
【図９】本発明の第１の実施形態による口語調表現部分の自動抽出部によって抽出された口語調音声データの例を示す図である。
【図１０】本発明の第１の実施形態による口語調表現部分の自動抽出部によって抽出された口語調韻律データの例を示す図である。
【図１１】本発明の第１の実施形態による口語調表現部分の自動抽出部によって抽出された口語調テキストデータの例を示す図である。
【図１２】本発明の第１の実施形態による口語調表現抽出ルール作成部で用いられるデータの例を示す図である。
【図１３】本発明の第１の実施形態による口語調表現抽出ルール作成部で生成されたルールの例を示す図である。
【図１４】本発明の第１の実施形態による、英語F0パターンによる口語調度の計算を説明する図である。
【図１５】本発明の第１の実施形態による、英語単語レベルで計算した口語調度の例を示す図である。
【図１６】本発明の第１お実施形態による口語調データ作成処理を説明するためのフローチャートである。
【図１７】本発明の第２の実施形態による音声合成システムの機能ブロックを示す図である。
【図１８】本発明の第１の実施形態による口語調表現抽出ルール作成部で用いられるデータ（口語調度が連続値）の例を示す図である。
【図１９】本発明の第２の実施形態による口語調度予測＆韻律混合比決定部で生成した口語調度および口語調混合比の例を示す図である。
【図２０】本発明の第２の実施形態による韻律混合部で生成した韻律の例を示す図である。
【図２１】本発明の第２の実施形態による韻律混合処理の概念を示す図である。
【図２２】本発明の第２の実施形態による口語調データ作成処理を説明するためのフローチャートである。
【発明を実施するための形態】
【００１６】
本発明は、従来手作業で行っていた口語調データ（口語調表現抽出ルール又は口語調度モデル、口語調韻律・音韻モデル、及び口語調音声ＤＢ）の作成を、収録口語調音声コーパスと読み上げ韻律・音韻予測モデルを用いて自動化するものである。具体的には、本発明は、収録した口語調音声から抽出した特徴量と、収録した音声のテキストデータを読み上げモデルで読み上げた場合の特徴量とを比較して口語調表現部分を抽出する。口語の言い回しで特徴ある部分（例えば、「今日は雨かなぁ」の「かなぁ」の部分）以外は、収録音声と読み上げ音声とでは差がないが、特徴部分に関しては特徴量に差異が生じるという性質を利用したものである。この特徴ある部分の情報から口語調表現抽出ルール等、口語調韻律・音韻モデル、口語調音声が作成され、口語調データとして登録される。音声合成装置では、この口語調データを用いて、例えば特許文献２で示された方法で口語調の音声合成データを出力する。
【００１７】
以下、添付図面を参照して本発明の実施形態について説明する。添付図面では、機能的に同じ要素は同じ番号で表示される場合もある。なお、添付図面は本発明の原理に則った具体的な実施形態と実装例を示しているが、これらは本発明の理解のためのものであり、決して本発明を限定的に解釈するために用いられるものではない。
【００１８】
本実施形態では、当業者が本発明を実施するのに十分詳細にその説明がなされているが、他の実装・形態も可能で、本発明の技術的思想の範囲と精神を逸脱することなく構成・

構造の変更や多様な要素の置き換えが可能であることを理解する必要がある。従って、以降の記述をこれに限定して解釈してはならない。

【００１９】

更に、本発明の実施形態は、後述されるように、汎用コンピュータ上で稼動するソフトウェアで実装しても良いし専用ハードウェア又はソフトウェアとハードウェアの組み合わせで実装しても良い。

【００２０】

なお、以後の説明では「テーブル」形式によって本発明の各情報について説明するが、これらの情報は必ずしもテーブルによるデータ構造で表現されていなくても良く、リスト、ＤＢ、キュー等のデータ構造やそれ以外で表現されていても良い。そのため、データ構造に依存しないことを示すために「テーブル」、「リスト」、「ＤＢ」、「キュー」等について単に「情報」と呼ぶことがある。

【００２１】

また、各情報の内容を説明する際に、「識別情報」、「識別子」、「名」、「名前」、「ＩＤ」という表現を用いることが可能であり、これらについてはお互いに置換が可能である。

【００２２】

以下では「プロセッサ」を主語（動作主体）として本発明の実施形態における各処理について説明を行うが、プロセッサはプログラムを実行することで定められた処理をメモリ及び通信ポート（通信制御装置）を用いながら行うため、「プログラム」を主語とした説明としてもよい。プログラムの一部または全ては専用ハードウェアで実現してもよく、また、モジュール化されていても良い。

【００２３】

（１）第１の実施形態

＜ハードウェア構成＞

図１は、本発明の実施形態による音声合成システムのハードウェア構成例を示す図である。音声合成システム１は、各種プログラムを実行するプロセッサ（ＣＰＵ：Central Processing Unit）１０１と、各種プログラムを格納するメモリ１０２と、各種データを格納する記憶装置１０３と、出力装置１０４と、入力装置１０５と、必要に応じて通信デバイス１０６と、を有している。

【００２４】

メモリ１０２は、少なくとも口語調データ（口語調表現抽出ルール、口語調韻律・音韻モデル、及び口語調音声ＤＢを含む）を作成する口語調データ作成プログラム１０２１と、読み上げ調データ（読み上げ韻律・音韻予測モデル、及び読み上げ調音声ＤＢを含む）を作成する読み上げ調データ作成プログラム１０２２と、音声合成処理を実行する音声合成プログラム１０２３と、を格納する。

【００２５】

記憶装置１０３は、読み上げ調データを作成するために用いられる、収録した読み上げ調音声データ１０３１と、収録した読み上げ調音声データから抽出された特徴量を学習して得られる読み上げ韻律・音韻予測モデル１０３２と、収録した読み上げ調音声データ１０３１から生成された読み上げ調音声ＤＢ１０３３と、を格納する。また、記憶装置１０３は、口語調データを作成する元データである収録した口語調音声コーパスデータ１０３４と、口語調音声コーパスデータ１０３４から抽出された口語調抽出データ１０３５と、口語調抽出データ１０３５に基づいて生成された口語調表現抽出ルール１０３６と、口語調抽出データ１０３５の特徴を学習して得られる口語調韻律・音韻モデル１０３７と、口語調抽出データ１０３５から生成される口語調音声ＤＢ１０３８と、を格納している。

【００２６】

出力装置１０４は、音声合成処理して得られる音声を出力するデバイスである。例えば、スピーカが該当する。また、出力装置１０４は、口語調データ作成処理によって生成されるデータを画面上に表示したりしても良い。

【００２７】
入力装置１０５は、テキストや音声を入力するためのデバイスである。例えば、テキスト入力であればキーボード、マウス、入力すべきテキストデータを取り込んで自動的に入力するソフトウェア等が該当し、音声入力であれば、マイクや入力すべき音声データを取り込んで自動的に入力するソフトウェア等が該当する。
【００２８】
通信デバイス１０６は、必要に応じて設けられるデバイスであり、処理に必要なデータを受信したり、処理済のデータを他の装置に送信したりする場合に用いられる。
【００２９】
なお、図１に示される音声合成システム１は、読み上げ調データ作成処理１０と、口語調データ作成処理１１と、音声合成処理１２と、を実行するシステムとして構成されているが、本発明の特徴である口語調データ作成処理１１のみを実行する口語調データ作成装置として構成しても良い。この場合、メモリ１０２に格納されるプログラムは口語調データ作成プロプログラムのみということになる。また、各処理を別のコンピュータで実行させるようなシステムを音声合成システムとして構成しても良い。この場合、読み上げ調データ作成処理１０を実行するコンピュータ（読み上げデータ作成装置）、口語調データ作成処理１１を実行するコンピュータ（口語調データ作成装置）、音声合成処理１２を実行するコンピュータ（音声合成装置）がそれぞれ独立に存在していても良い（それぞれがネットワークに接続され、遠隔的に設置されていても良い）。
【００３０】
コンピュータが口語調データ作成装置を構成する場合には、メモリ１０２は、口語調データ作成プログラム１０２１を格納する。また、この場合、記憶装置１０３は、収録した口語調音声コーパスデータ１０３４、口語調抽出データ１０３５、口語調表現抽出ルール１０３６、口語調韻律・音韻モデル１０３７、及び口語調音声ＤＢ１０３８を格納することとなる。
【００３１】
＜機能ブロックと処理内容＞
図２は、本発明の第１の実施形態による音声合成システム１の機能ブロック図である。音声合成システム１は、読み上げ調データ作成処理１０と、口語調データ作成処理（口語調の音声合成用データを作成する処理）１１と、口語調データを用いた音声合成処理１２を実行する。
【００３２】
読み上げ調データ作成処理１０は、韻律・音韻抽出部、音声ＤＢ作成部、韻律・音韻特徴自動学習部などで構成されるが、これらによる処理は一般的な音声合成データ作成に用いられる処理であり、本発明の特徴となるものではないので、以降、その説明は原則省略する。
【００３３】
（ｉ）口語調データ作成処理
口語調データ作成処理１１は、収録した口語調音声と、読み上げ韻律・音韻予測モデルを入力すると、口語調音声から口語調表現部分を自動抽出し、口語調音声合成に必要な口語調表現抽出ルール、口語調韻律・音韻モデル、口語調音声ＤＢ（データベースの略）を生成する処理である。この口語調データ作成処理１１を実現するために、韻律・音韻特徴抽出部２０２、韻律・音韻自動学習部２０６、音声ＤＢ作成部２０７という通常の音声データ作成装置が持つ処理単位を備える。これに加えて、本発明に特徴的な、読み上げ調韻律・音韻予測部２０１、口語調度算出部２０３、口語調表現部分の自動抽出部２０４が設けられ、さらに、入力テキストから口語調テキスト表現を自動検出できる口語調表現抽出ルールを生成する口語調表現抽出ルール作成部２０５が設けられる。
【００３４】
音声合成処理１２は、口語調データ作成処理によって生成された口語調表現抽出ルール１０３６、口語調韻律・音韻モデル１０３７、及び口語調音声ＤＢ１０３８を用いて、入

力されたテキストデータを処理し、合成音声を生成する処理である。当該音声合成処理は、特許文献２に開示された内容と同様であるので、以降、詳細な説明は原則省略する。

【００３５】

以上のように、読み上げ調データ作成処理１０と音声合成処理１２は、通常の音声合成システムにも存在する処理単位であり、口語調データ作成処理１１が本発明の特徴となる処理に相当する。従って、以下では口語調データ作成処理１１を中心に説明することとする。

【００３６】

なお、以降の説明において、収録した音声として「お願いします」や「今日は雨かなぁ」などの単文を用いているが、複数の文を結合した長い文書の収録も可能である。また、図２では入力テキストを漢字かな文としているが、もちろん、英語や中国語などの外国語でも構わない。その場合は、内部処理もその外国語に対応したプログラム・データ（たとえば、収録した口語調音声、収録した読み上げ調音声）を用いなければならないことは言うまでもない。

【００３７】

読み上げ調韻律・音韻予測部２０１は、収録した口語調音声コーパスデータ１０３４から、収録した口語調音声の発話テキストを読み出し、それに対して読み上げ韻律・音韻予測モデル１０３２を適用し、読み上げ調の韻律特徴量および音韻特徴量を予測する。つまり、ここでは、発話者がこのテキストに対して、読み上げ調スタイルで発話した場合は、韻律・音韻特徴がどのようなものであるかが分かる。ただし、韻律特徴量は、発話速度を表す特徴量（例えば、音素継続長、音節継続長など）、声の高さを表す特徴量（例えば、基本周波数の時間変化パターン（Ｆ０パターン）など）、音の大きさを表す特徴量（例えば、短時間平均パワーなど）等である。音韻特徴量は、声道形状を表す特徴量（例えば、ケプストラム、ＬＰＣ係数など）が考えられる。また、これらの情報をすべて用いる必要がないが、口語調の特徴に最も寄与する基本周波数を用いることが好ましい。ただし、以降、本明細書では、音韻特徴量についての説明を省略し、単に韻律特徴量と記載した場合でも、韻律特徴量と音韻特徴量と両方を意味するものとする。また、韻律特徴量についても、理解しやすいＦ０パターンを中心に説明を行うこととする。例えば、収録した口語調音声「今日は雨かなぁ」のテキストに対して予測された韻律特徴は図３に示されるようなものとなる。なお、読み上げ韻律・音韻予測モデル１０３２は、口語調音声と同じ話者の読み上げ調音声から学習したものを用いることが望ましいが、別の話者から学習したモデルを口語調音声話者に適用したものでも良い。

【００３８】

韻律・音韻特徴抽出部２０２は、収録した収録した口語調音声コーパスデータ１０３４から収録音声データを読み出し、その音声の韻律・音韻特徴量を抽出する。つまり、発話者がこのテキストに対して、実際に口語調スタイルで発話した場合の韻律・音韻特徴がどのようなものであるかが分かる。だたし、収録した口語調音声には、事前に音素セグメンテーション情報が、自動および手動で付与されているものとする。なお、抽出する特徴量は、読み上げ調韻律・音韻予測部で予測された特徴量と同じである。例えば、収録した口語調音声「今日は雨かなぁ」の音声波形に対して、抽出した韻律特徴量は、図４に示されるようなものとなる。

【００３９】

口語調度算出部２０３は、読み上げ調韻律・音韻予測部２０１で予測された特徴量と、韻律・音韻特徴抽出部２０２で抽出した収録口語調音声の特徴量とを比較し、口語調への寄与度（口語調度）を計算する。例えば、口語調算出部２０３は、単純に音素ごとに韻律特徴量の差分を取り、下記式１を用いて音素ごとの口語調度を計算する。図５は、継続長係数＝０．３；高さ係数＝０．５；強さ係数＝０．２の場合、計算された口語調度を示している。

【００４０】

口語調度＝｜継続長係数＊継続長差分｜＋｜高さ係数＊高さ差分｜

＋｜強さ係数＊強さ差分｜　・・・・・　（式１）

ここで、“｜Ａ｜”は“Ａ”の絶対値を示すものとする。

【００４１】

口語調表現部分の自動抽出部２０４は、口語調度算出部２０３で計算された口語調度を用いて、音声を構成する各セグメントについて、セグメントの口語調度を計算し、口語調度が所定の閾値以上を示すセグメントを口語調表現部分として自動抽出する。抽出された口語調表現部分は、口語調抽出データ１０３５に格納される。口語調表現のセグメント単位は、合成時に用いる韻律モデルにも依存するが、日本語であれば音節単位、形態素単位、アクセント句単位などが適切だと考えられる。例えば、セグメント単位が音節（モーラ）の場合、口語調特徴が母音のみに現れることを仮定すると、音節口語調度は式２のようになる。つまり、音節に含まれる母音の口語調度が音節口語調度として与えられる。式２に従うと、音節口語調度は、図６のようになる。閾値が“２０”の場合は、７番目の“ナ”と８番目の“ァ”が抽出される。閾値が“１０”の場合には、６番目の“カ”も口語調データとして、抽出されることになる。なお、例えば、閾値は経験値で定められる値であり、予め決めておく。

【００４２】

音節口語調度　＝　母音口語調度　・・・・・　（式２）

【００４３】

また、例えば、セグメント単位が形態素であるとする場合、口語調度は式３のように表される。つまり、音節に含まれる音節の口語調度の平均値が形態素口語調度として与えられる。式３に従うと、形態素口語調度は、図７のようになる。閾値が“２５”の場合は、４番目の“かなぁ”が口語調データとして、抽出される。

【００４４】

形態素口語調度　＝　音節平均口語調度　・・・・・　（式３）

【００４５】

さらに、セグメント単位がアクセント句の場合、口語調度は式４のように表される。つまり、アクセント句に含まれる形態素の形態素口語調度のうち、最大値が口語調度として与えられる。式４に従うと、アクセント句口語調度は図８のようになる。閾値が“２０”の場合は、２番目の“雨かなぁ”が口語調データとして、抽出される。

【００４６】

アクセント句口語調度　＝　形態素最大口語調度　・・・・・　（式４）

【００４７】

音声ＤＢ作成部２０７は、口語調表現部分の自動抽出部２０４によって抽出された口語調抽出データの音声波形を蓄積し、音声合成に用いる口語調音声ＤＢ１０３８を作成する。口語調音声ＤＢ１０３８は、音声合成装置による音声合成処理に適合する所定のフォーマットで作成される。例えば、「今日は雨かなぁ」から抽出した口語調表現部分の音声波形は、図９のようになる。

【００４８】

韻律音韻特徴自動学習部２０６は、口語調表現部分の自動抽出部２０４から抽出された口語調抽出データの韻律・音韻情報（図１０参照）を用いて、音声合成に用いる口語調韻律・音韻モデル１０３７を作成する。口語調韻律・音韻モデル１０３７は、音声合成装置による音声合成処理に適合する所定のフォーマットで作成される。口語調韻律・音韻モデル１０３７は、コンテキストから韻律・音韻情報を推定する統計モデルでも良いし、口語調のデータとして抽出され肉声の韻律・音韻情報をそのまま蓄積したモデルでも良い。例えば、「今日は雨かなぁ」から抽出した口語調表現部分の韻律（Ｆ０パターン）は、図１０のようになる。

【００４９】

口語調表現抽出ルール作成部２０５は、口語調表現部分の自動抽出部２０４によって抽出された口語調抽出データのテキスト（図１１）を用いて、口語調表現抽出ルールを作成する。最も簡単な口語調表現抽出ルールは、「“かなぁ”という文字列がマッチした場合

、その部分を口語調表現とする。」のように、文字列表現のみを用いた文字列マッチングルールである。ただし、このようなルール作成手法では、例えば「お願いします」の口語調音声に対して、“します”の部分を口語調表現として抽出されたとすると、「します”という文字列がマッチした場合、その部分を口語調表現とする。」というルールを作成されてしまう。このルールは明らかに不適切である。従って、作成したルールには、前後のコンテキスト情報を考慮した方が良いと考えられる。例えば、「お願いします」「今日は雨かなぁ」の文に対して、口語調テキストを形態素単位（アクセント句単位など、形態素より大きい言語単位でも良い）に分解し、それぞれコンテキストと口語調度（“Ｙｅｓ”と“Ｎｏ”の２値）を付与すると、図１２のようになる。このデータに対して、機械学習手法を用いて、口語調表現抽出ルールを自動作成することができる。例えば、２分岐決定木を自動構築した場合、図１３のようになる。もちろん、ニューラルネットワーク、スーパーベクトルマシンなどのカテゴリを推測する手法を用いても良い。図１３は、「お願いします」「今日は雨かなぁ」の文に対して、口語調テキストを形態素単位（アクセント句単位など、形態素より大きい言語単位でも良い）に分解し、それぞれコンテキストと口語調度（口語調表現部分抽出部で口語調抽出に用いる口語調度の連続値）を付与した場合の２分岐決定木（図１２を基に学習したツリー）を示している。このデータに対して、機械学習手法を用いて、口語調表現度予測モデルを自動作成することができる。例えば、重回帰解析などの連続値を推測する統計手法を用いることができる。そして、合成時に文を構成する各形態素について、口語調度を予測し、ある閾値を超えた形態素を「口語調表現」とし、一方、予測した口語調度がその閾値より小さい形態素を「口語調表現でない」とする。

【００５０】

以上のように、読み上げ調韻律・音韻予測部２０１、韻律・音韻特徴抽出部２０２、口語調度算出部２０３、及び口語調表現部分の自動抽出部２０４については、日本語「今日は雨かなぁ」を適用した場合を例に説明したが、英語や中国語などの外国語でも構わない。例えば、口語調音声が英語「0h, It's raining.」である場合、図１４で示すように、Ｆ０観測値（収録した口語調音声からのＦ０値）とＦ０予測値（読み上げ調音声からのＦ０値）が得られたとする。英語の場合は、口語調表現のセグメント単位は、音素や音節より、単語や韻律語を用いたほうが良い。例えば、セグメント単位が単語の場合、口語調度は式５のようになる。ただし、この式は一例であり、上記式１を用いても構わない。式５に従うと、各単語の口語調度が図１５のようになる。閾値を５０と設定した場合、一番最初の“0h”が、口語調表現として抽出される。

【００５１】

単語口語調度＝｜予測した単語最大Ｆ０値　－　観測した単語最大Ｆ０値｜　・・・（式５）

ここで、“｜Ａ｜”は“Ａ”の絶対値を示している。

【００５２】

（ii）音声合成処理

音声合成処理では、まずテキスト入力部に音声合成すべきテキスト（例えば、かな漢字文）がユーザによって入力され、テキスト解析部で解析される。

【００５３】

口語調表現自動抽出部は、テキスト解析部で解析されたコンテキスト情報と口語調データ作成処理１１の口語調表現抽出ルール作成部２０５で作成された口語調表現抽出ルール１０３６を用いて、入力テキストを「口語調表現」部分と「口語調表現でない」部分に分割する。「口語調表現」部分は、口語調部分の韻律・音韻作成部に出力し、「口語調表現でない」部分は、読み上げ部分の韻律・音韻予測部に出力する。ただし、入力テキストに必ず「口語調表現」部分と「口語調表現でない」部分と両方存在すると限らないので、入力テキストが必ず分割されると限らない。

【００５４】

例えば、テキスト「今日は晴れかなぁ」が入力された場合、図１３で示した口語調表現

抽出ルールに従い、口語調表現が抽出される。この例では、形態素「かなぁ」が「口語調表現」として抽出され、残りの形態素が「口語調表現でない」と判断される。そのため、入力テキスト「今日は晴れかなぁ」は、口語調表現でない部分の「今日は晴れ」と口語調表現部分の「かなぁ」と分割される。また、例えば、テキスト「掃除します」が入力された場合、図１３で示した口語調表現抽出ルールには「掃除」「します」の両方とも口語調表現として登録されていないため、口語調表現は抽出されず、文分割は行われない。

【００５５】

韻律生成部は、口語調部分の韻律・音韻作成部で生成された韻律・音韻特徴量と読み上げ部分の韻律・音韻予測部で生成された韻律・音韻特徴量を合併し、文全体の韻律・音韻特徴量ターゲットを生成する。

【００５６】

そして、波形生成部は、読み上げ調音声ＤＢ１０３３を参照して、口語調ではない部分のテキストについて声質を考慮した処理を実行し、読み上げ調部分のテキストについて音声波形を生成する。また、口語調音声生成部は、口語調音声ＤＢ１０３８を参照して、口語調部分のテキストについて音声波形を生成する。

【００５７】

波形接続部は、口語調部分の音声波形と口語調ではない部分（読み上げ調部分）の音声波形を接続し、音声出力部は、最終的な合成音声を出力する。

【００５８】

＜口語調データ作成処理のフローチャート＞

図１６は、本発明の第１の実施形態による口語調データ作成処理を説明するためのフローチャートである。

【００５９】

（ｉ）ステップ１６０１

プロセッサ１０１は、収録した口語調音声コーパスデータ１０３４の入力を受け付ける。当該データには、収録音声データとそれに対応するテキストデータ（発話テキスト）がセットとなっている。

【００６０】

（ｉｉ）ステップ１６０２

プロセッサ１０１は、収録した収録した口語調音声コーパスデータ１０３４の収録音声データから、その音声の韻律・音韻特徴量を抽出する。詳細については上述した通りである。

【００６１】

（ｉｉｉ）ステップ１６０３

プロセッサ１０１は、収録した口語調音声コーパスデータ１０３４の発話テキストに対して読み上げ韻律・音韻予測モデル１０３２を適用し、読み上げ調の韻律特徴量および音韻特徴量を予測する。つまり、ここでは、発話者がこのテキストに対して、読み上げ調スタイルで発話した場合は、韻律・音韻特徴がどのようなものであるかが分かる。詳細は上述した通りである。

【００６２】

（ｉｖ）ステップ１６０４

プロセッサ１０１は、ステップ１６０２で抽出した収録口語調音声の特徴量と、ステップ１６０３で予測した韻律・音韻特徴量とを比較し、口語調への寄与度（口語調度）を計算する。

【００６３】

（ｖ）ステップ１６０５

プロセッサ１０１は、ステップ１６０４で得られた口語調度を用いて、音声を構成する各セグメントについて、セグメントの口語調度を計算し、口語調度が所定の閾値以上を示すセグメントを口語調表現部分として自動抽出する。抽出された口語調表現部分は、口語調抽出データ１０３５に格納される。詳細は上述した通りである。

【００６４】
（vi）ステップ１６０６
プロセッサ１０１は、ステップ１６０５で得られた口語調抽出データの音声波形を蓄積し、音声合成に用いる口語調音声ＤＢ１０３８を作成する
【００６５】
（vii）ステップ１６０７
プロセッサ１０１は、口語調抽出データの韻律・音韻情報（図１０参照）を用いて、音声合成に用いる口語調韻律・音韻モデル１０３７を作成する。詳細は上述した通りである。
【００６６】
（viii）ステップ１６０８
プロセッサ１０１は、ステップ１６０５で得られた口語調抽出データのテキスト（図１１）を用いて、口語調表現抽出ルール１０３６を作成する。詳細は上述した通りである。
【００６７】
（２）第２の実施形態
第２の実施形態は、特許文献１のような音声合成装置に用いる口語調音声合成用データを作成することを想定したものである。ハードウェア構成は第１の実施形態と同様であるので、説明は省略する。ただし、記憶装置１０３は、口語調抽出データ１０３５の代わりに口語調度付き口語調音声データ１７０２、口語調表現抽出ルール１０３６の代わりに口語調度予測モデル１７０４を格納する。
【００６８】
＜機能ブロックと処理内容＞
図１７は、本発明の第２の実施形態による音声合成システムの機能ブロックを示す図である。第２の実施形態では、従来手作業による音声の口語調度ラベリングに代わって、収録した口語調音声にセグメントごとに、口語調度の定量的な評価を実現し、入力テキストの各セグメントに対する口語調度を予測する。この予測した口語調度によって、口語調音声から学習した韻律・音韻モデルと読み上げ調音声から学習した読み上げ調韻律・音韻モデルと、セグメント毎の混合割合を計算し、文全体の韻律・音韻特徴の予測を行う。第１の実施形態とは異なり、入力テキストを分割することがないので、分割された口語調表現部分と口語調表現でない部分と接続するときの不連続感を低減できると考えられる。
【００６９】
以下では、第１の実施形態とは異なる部分のみ説明することとする。
（i）口語調データ作成処理
口語調度付与部１７０１は、口語調度算出部２０３で算出された韻律特徴の差分情報を用いて、収録した口語調音声の各セグメントに口語調度を付与し、口語調度付き口語調音声データ１７０２を生成する。ここで、セグメントの単位は、音素、音節、形態素、アクセント句、フレーズ、文などが考えられるが、口語調音声の特徴を担う最小単位として、形態素を用いたことが好ましい。各セグメントの口語調度の計算については、口語調度算出部２０３で算出された韻律特徴の差分情報から求められるが、その具体例については、第１の実施形態で説明したので、ここでは詳細については省略する。「お願いします」「今日は雨かなぁ」の文に対して、口語調テキストを形態素単位に分解し、それぞれコンテキストと口語調度を付与すると、図１８のようになる。
【００７０】
口語調度予測モデル学習部１７０３は、口語調度付与部１７０１が生成した口語調度付き口語調音声データ１７０２を用いて、口語調度を予測する統計モデル（口語調度予測モデル）１７０４を生成する。第１の実施形態では、入力文（テキスト）を「口語調」の部分と「口語調でない」の部分と分割するためのルールを作成しているが、第２の実施形態では、入力文を構成するすべてのセグメントについて、口語調度を予測するための統計モデルを作成することになる。
【００７１】

（ii）音声合成処理
第２の実施形態では、テキスト解析部が入力テキストを解析した後、口語調度予測＆韻律混合比決定部が、口語調度予測モデル１７０４を用いて、テキスト文を構成する各セグメントについて、口語調度を予測する。さらに、口語調度予測＆韻律混合比決定部は、この予測した口語調度に基づいて、口語調韻律と読み上げ調韻律の混合比率を計算する。例えば、「今日は晴れかなぁ」というテキスト文が入力された場合、すべての形態素について口語調を予測した結果は、図１９のようになる。ここで、口語調混合比を式６のように定義した場合（口語調下限値＝０，口語調上限値＝５０とする）、口語調混合比は、図１９に示される値となる。
【００７２】
口語調混合比＝ＭＩＮ（１００％，（口語調度－口語調下限値）／（口語調上限値－口語調下限値））
・・・・・　（式６）
ここで、ＭＩＮ（Ａ，Ｂ）は、ＡとＢとの間で小さい方の値を選ぶことを意味するものとする。
【００７３】
読み上げ調韻律・音韻予測部２０１は、読み上げ韻律・音韻予測モデル１０３２を参照し、入力テキストの読み上げ調の音声データを予測する。口語調韻律・音韻作成部は、口語調韻律・音韻モデル１０３７を参照し、入力テキストの口語調の音声データを作成する。
【００７４】
韻律混合部は、口語調度予測＆韻律混合比決定部で生成された口語調混合比を用いて、口語調の音声データと読み上げ調の音声データの韻律混合処理を実施する。例えば、入力テキスト「今日は晴れかなぁ」に対して、図２０に示されるように口語調韻律と読み上げ韻律が予測された場合、式７を用いて、韻律生成を行う。
【００７５】
韻律＝口語調韻律＊口語調混合比＋読み上げ調韻律＊（１-口語調混合比）　・・・・・　（式７）
【００７６】
図２１は、当該韻律混合処理の概念を示す図である。図２１に示されるように、「今日」については読み上げ調音声データが９０％、口語調音声データが１０％用いられる。「は」についてはそれぞれ８４％、１６％用いられ、「晴れ」についてはそれぞれ６６％、３４％用いられる。そして、「かな」については読み上げ調音声データが４％、口語調音声データが９６％用いられて、混合韻律が生成される。
【００７７】
最後に、音声生成部は、韻律混合部で生成した韻律をターゲットとして音声を生成し、音声出力部がこれを出力する。
【００７８】
＜口語調データ作成処理のフローチャート＞
図２２は、本発明の第２の実施形態による口語調データ作成処理を説明するためのフローチャートである。
【００７９】
（i）ステップ２２０１
プロセッサ１０１は、収録した口語調音声コーパスデータ１０３４の入力を受け付ける。当該データには、収録音声データとそれに対応するテキストデータ（発話テキスト）がセットとなっている。
【００８０】
（ii）ステップ２２０２
プロセッサ１０１は、収録した収録した口語調音声コーパスデータ１０３４の収録音声データから、その音声の韻律・音韻特徴量を抽出する。詳細については上述した通りであ

る。
【００８１】
（iii）ステップ２２０３
プロセッサ１０１は、収録した口語調音声コーパスデータ１０３４の発話テキストに対して読み上げ韻律・音韻予測モデル１０３２を適用し、読み上げ調の韻律特徴量および音韻特徴量を予測する。つまり、ここでは、発話者がこのテキストに対して、読み上げ調スタイルで発話した場合は、韻律・音韻特徴がどのようなものであるかが分かる。詳細は上述した通りである。
【００８２】
（iv）ステップ２２０４
プロセッサ１０１は、ステップ２２０２で抽出した収録口語調音声の特徴量と、ステップ２２０３で予測した韻律・音韻特徴量とを比較し、口語調への寄与度（口語調度）を計算する。
【００８３】
（v）ステップ２２０５
プロセッサ１０１は、ステップ２２０４で算出された口語調度（韻律特徴の差分情報）を用いて、収録した口語調音声の各セグメントに口語調度を付与し、口語調度付き口語調音声データ１７０２を生成する。詳細は上述した通りである。
【００８４】
（vi）ステップ２２０６
プロセッサ１０１は、ステップ２２０５で得られた口語調度付き口語調音声データ１７０２の音声波形を蓄積し、音声合成に用いる口語調音声ＤＢ１０３８を作成する。
【００８５】
（vii）ステップ２２０７
プロセッサ１０１は、口語調度付き口語調音声データ１７０２の韻律・音韻情報（図１０参照）を用いて、音声合成に用いる口語調韻律・音韻モデル１０３７を作成する。詳細は上述した通りである。
【００８６】
（viii）ステップ２２０８
プロセッサ１０１は、ステップ２２０４で得られた口語調度付き口語調音声データ１７０２を用いて、口語調予測モデル（口語調度予測モデル）１７０４を生成する。詳細は上述した通りである。
【００８７】
（３）まとめ
（i）第１の実施形態では、口語調音声データから韻律特徴量を抽出し、一方、当該口語調音声データに対応するテキストデータに対して読み上げ韻律・音韻予測モデルを適用して読み上げ調の韻律特徴量を予測する。次に、これらの韻律特徴量の差分を取り、差分値が所定の閾値（経験から設定される値）よりも大きい箇所を口語調の特徴部分（音声合成に用いる口語調データ）として抽出する。これらの処理は、収録した口語調音声コーパスと読み上げ韻律・音韻予測モデルを与えれば自動的に実行される。このように、口語調音声を始めとする韻律や声質の変化が大きい発話スタイルの合成音声から、その特徴を担う部分（口語調音声の場合は、口語調表現部分）を自動的に抽出するので、作業コストを抑えることができるうえ、異なった作業者による基準の不統一を改善できる。
【００８８】
第１の実施形態では、口語調表現抽出ルールが生成される。このルールは、口語調の特徴部分のテキストデータを用いて、与えられるテキストデータにおける口語調表現を抽出するためのルールとして生成される。この場合、特徴部分のテキストデータに加えて、当該テキストデータが含まれる口語調テキストの前後のコンテキスト情報を用いて当該ルールを生成するようにしても良い。このようなルールを作成することにより、このルールに従って生成された合成音声を、より自然で安定的な口語調音声とすることができるように

なる。

【００８９】

第２の実施形態では、口語調韻律データと読み上げ調韻律データの差分値に基づいて、テキストのセグメントに対して、当該セグメントの口語調の程度を示す口語調度を算出し、これを口語調音声データに付与する。そして、この口語調度が付与された口語調音声データを用いて、音声合成用データが生成される。第２の実施形態による音声合成用データは、音声合成すべき入力テキストの口語調度を予測するための統計モデル（口語調度予測モデル）となっている。第２の実施形態によっても上述の第１の実施形態と同様の技術的効果を期待することができる。

【００９０】

（ii）本発明は、実施形態の機能を実現するソフトウェアのプログラムコードによっても実現できる。この場合、プログラムコードを記録した記憶媒体をシステム或は装置に提供し、そのシステム或は装置のコンピュータ（又はＣＰＵやＭＰＵ）が記憶媒体に格納されたプログラムコードを読み出す。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコード自体、及びそれを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク、ＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ、ハードディスク、光ディスク、光磁気ディスク、ＣＤ－Ｒ、磁気テープ、不揮発性のメモリカード、ＲＯＭなどが用いられる。

【００９１】

また、プログラムコードの指示に基づき、コンピュータ上で稼動しているＯＳ（オペレーティングシステム）などが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。さらに、記憶媒体から読み出されたプログラムコードが、コンピュータ上のメモリに書きこまれた後、そのプログラムコードの指示に基づき、コンピュータのＣＰＵなどが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。

【００９２】

さらに、実施の形態の機能を実現するソフトウェアのプログラムコードを、ネットワークを介して配信することにより、それをシステム又は装置のハードディスクやメモリ等の記憶手段又はＣＤ－ＲＷ、ＣＤ－Ｒ等の記憶媒体に格納し、使用時にそのシステム又は装置のコンピュータ（又はＣＰＵやＭＰＵ）が当該記憶手段や当該記憶媒体に格納されたプログラムコードを読み出して実行するようにしても良い。

【００９３】

ここで述べたプロセス及び技術は本質的に如何なる特定の装置に関連することはなく、コンポーネントの如何なる相応しい組み合わせによっても実装できる。更に、汎用目的の多様なタイプのデバイスがここで記述内容に従って使用可能である。ここで述べた方法のステップを実行するのに、専用の装置を構築するのも有益である。また、実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成することもできる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態にわたる構成要素を適宜組み合わせてもよい。本発明は、具体例に関連して記述したが、これらは、すべての観点において限定の為ではなく説明のためである。本分野にスキルのある者であれば、本発明を実施するのに相応しいハードウェア、ソフトウェア、及びファームウエアの多数の組み合わせがあることを理解できるものと考えられる。例えば、記述したソフトウェアは、アセンブラ、Ｃ／Ｃ＋＋、ｐｅｒｌ、Ｓｈｅｌｌ、ＰＨＰ、Ｊａｖａ（登録商標）等の広範囲のプログラム又はスクリプト言語で実装できる。

【００９４】

さらに、上述の実施形態において、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示しているとは限らない。全ての構成が相互に接続されていても良い。

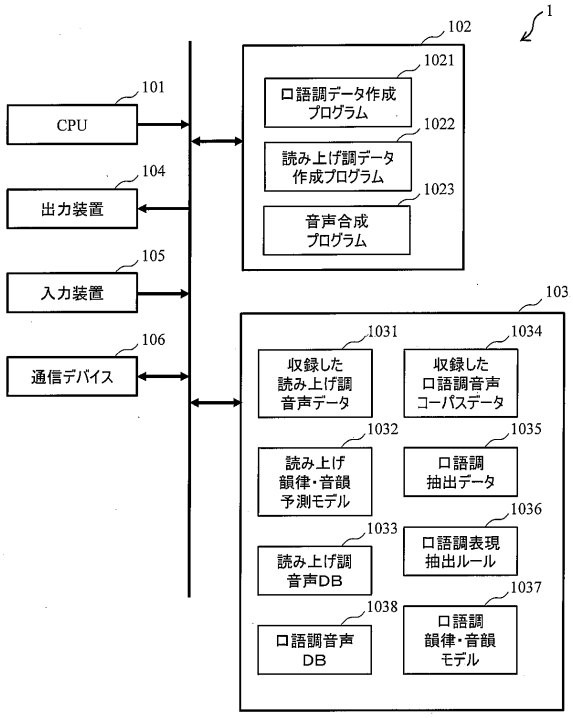
【符号の説明】

【００９５】

１　音声合成システム
１０　読み上げ調データ作成処理
１１　口語調データ作成処理
１２　音声合成処理
１０１　ＣＰＵ
１０２　メモリ
１０３　記憶装置
１０４　出力装置
１０５　入力装置
１０６　通信デバイス

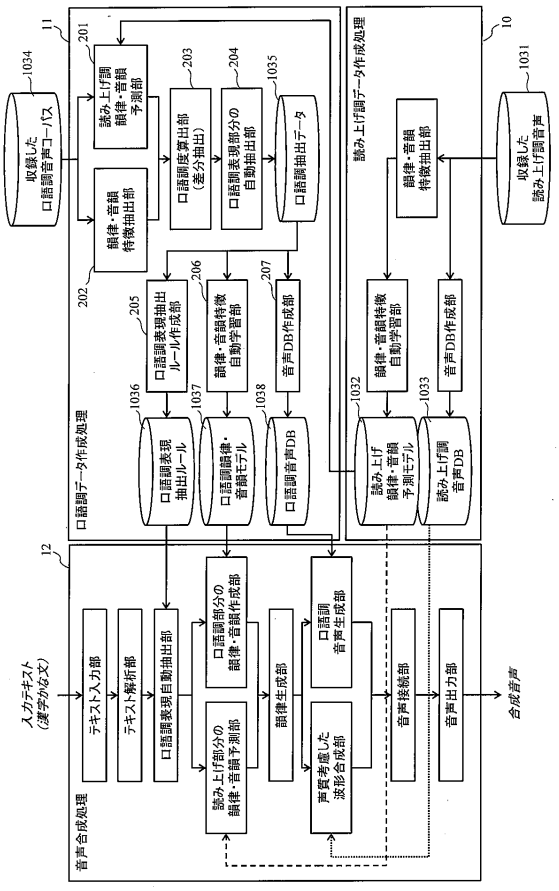
【図１】

図１



【図２】

図２



【図３】

図３

ID	音素	長さ(ms)	高さ(Hz)	強さ(dB)
1	KY	25	232	-40
2	O	83	248	-20
3	O	75	274	-20
4	W	32	262	-30
5	A	75	239	-25
6	A	62	230	-25
7	M	21	239	-30
8	E	55	263	-20
9	K	27	252	-35
10	A	68	237	-30
11	N	33	225	-35
12	A	50	200	-40
13	A	30	180	-40

【図４】

図４

ID	音素	長さ(ms)	高さ(Hz)	強さ(dB)
1	KY	25	232	-40
2	O	83	248	-20
3	O	75	274	-20
4	W	32	262	-30
5	A	75	239	-25
6	A	62	230	-25
7	M	21	239	-30
8	E	55	263	-20
9	K	27	252	-35
10	A	68	207	-25
11	N	33	215	-25
12	A	74	248	-20
13	A	89	259	-25

【図５】

図５

ID	音素	長さ差分 (ms)	高さ差分 (Hz)	強さ差分 (dB)	口語調度
1	KY	0	0	0	0
2	O	0	0	0	0
3	O	0	0	0	0
4	W	0	0	0	0
5	A	0	0	0	0
6	A	0	0	0	0
7	M	0	0	0	0
8	E	0	0	0	0
9	K	0	0	0	0
10	A	0	-30	5	14
11	N	0	-10	10	3
12	A	24	48	20	35.2
13	A	59	79	15	60.2

【図６】

図６

ID	音節 (モーラ)	口語調度
1	キョ	0
2	ウ	0
3	ワ	0
4	ア	0
5	メ	0
6	カ	14
7	ナ	35.2
8	ァ	60.2

【図７】

図７

ID	形態素	口語調度
1	今日	0
2	は	0
3	雨	0
4	かなぁ	36.4

【図８】

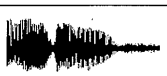
図８

ID	アクセント句	口語調度
1	今日 は	0
2	雨かなぁ	36.4

【図９】

図９

抽出した口語調音声データ

ID	文字列	口語調音声DB
1	かなぁ	

【図１０】

図１０

抽出した口語調韻律データ

ID	文字列	口語調韻律データ(F0)
1	かなぁ	

【図１１】

図１１

抽出した口語調テキスト

ID	文字列	所属文書
1	かなぁ	今日は雨かなぁ

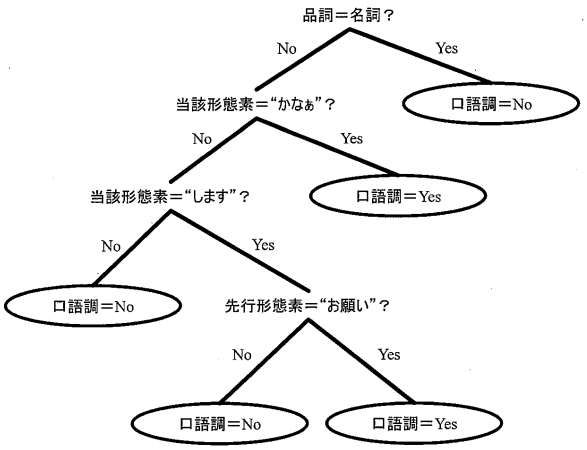
【図１２】

図１２

ID	表記	口語調	コンテキスト
1	今日	No	先行形態素＝なし; 当該形態素＝“今日”; 後続形態素＝“は”; 品詞＝名詞; ・・・
2	は	No	先行形態素＝“今日”; 当該形態素＝“は”; 後続形態素＝“雨”; 品詞＝助詞; ・・・
3	雨	No	先行形態素＝“は”; 当該形態素＝“雨”; 後続形態素＝“かなぁ”; 品詞＝名詞; ・・・
4	かなぁ	Yes	先行形態素＝“雨”; 当該形態素＝“かなぁ”; 後続形態素＝なし; 品詞＝助詞; ・・・
5	お願い	No	先行形態素＝なし; 当該形態素＝“お願い”; 後続形態素＝“します”; 品詞＝名詞; ・・・
6	します	Yes	先行形態素＝“お願い”; 当該形態素＝“します”; 後続形態素＝なし; 品詞＝動詞; ・・・

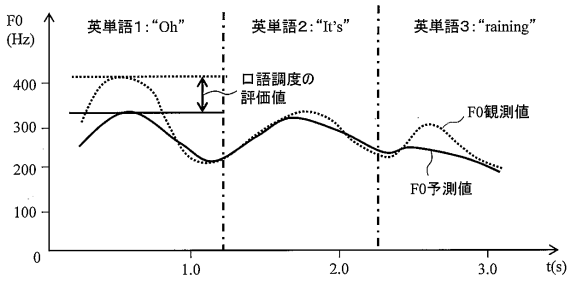
【図１３】

図１３



【図１４】

図１４



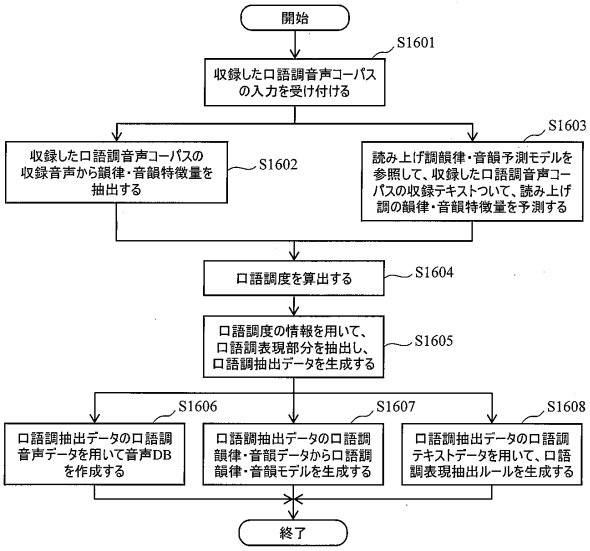
【図１５】

図１５

ID	単語	観測した最大F0値	予測したd最大F0値	口語調度
1	Oh	405	325	80
2	It’s	335	330	5
3	Raining	325	285	40

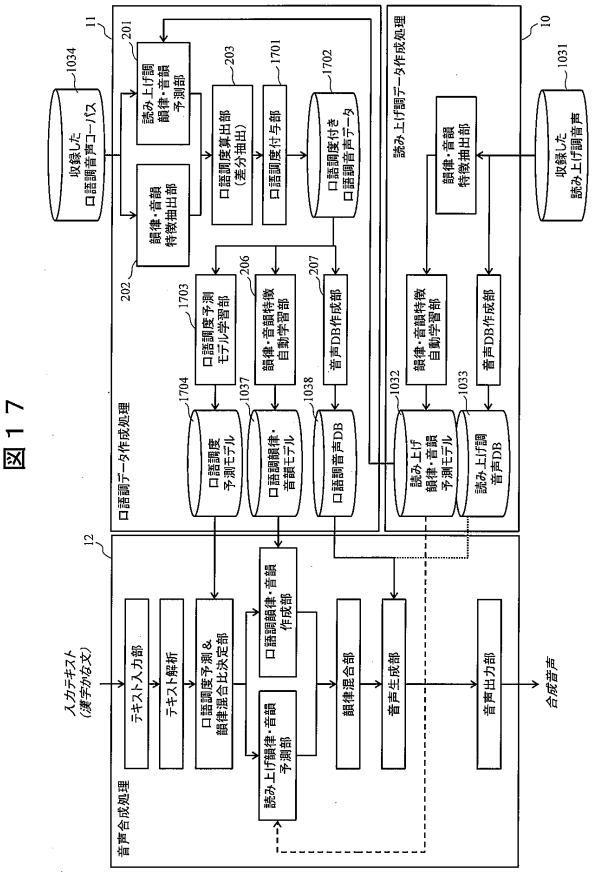
【図１６】

図１６



【図１７】

図１７



【図１８】

図１８

ID	表記	口語調度	コンテキスト
1	今日	0	先行形態素＝なし; 当該形態素＝“今日”; 後続形態素＝“は”; 品詞＝名詞; ・・・
2	は	0	先行形態素＝“今日”; 当該形態素＝“は”; 後続形態素＝“雨”; 品詞＝助詞; ・・・
3	雨	0	先行形態素＝“は”; 当該形態素＝“雨”; 後続形態素＝“かなぁ”; 品詞＝名詞; ・・・
4	かなぁ	36.4	先行形態素＝“雨”; 当該形態素＝“かなぁ”; 後続形態素＝なし; 品詞＝助詞; ・・・
5	お願い	8.3	先行形態素＝なし; 当該形態素＝“お願い”; 後続形態素＝“します”; 品詞＝名詞; ・・・
6	します	23.6	先行形態素＝“お願い”; 当該形態素＝“します”; 後続形態素＝なし; 品詞＝動詞; ・・・

【図１９】

図１９

ID	表記	口語調度	口語調混合比
1	今日	5	10%
2	は	8	16%
3	晴れ	17	34%
4	かなぁ	48	96%

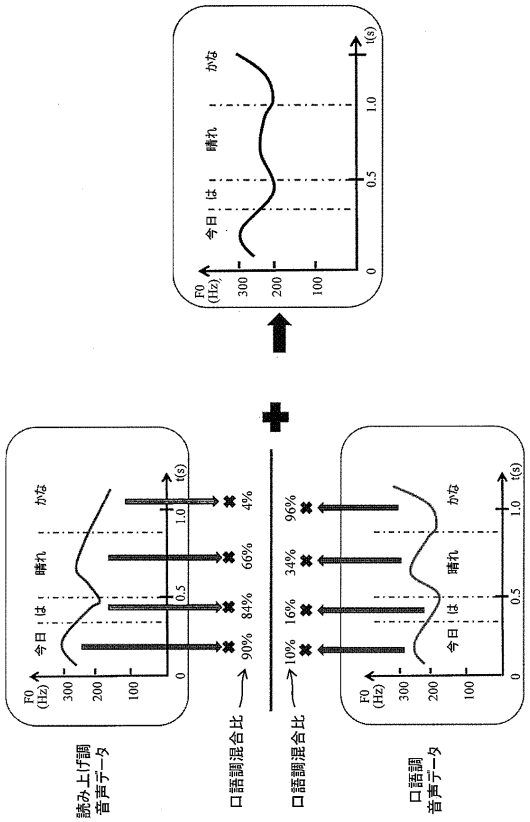
【図２０】

図２０

ID	音素	口語調混合比	予測された口語調韻律（F0）	予測された読み上げ調韻律（F0）	生成された韻律（F0）
1	KY	10%	232	232	232
2	O	10%	248	248	248
3	O	10%	274	274	274
4	W	16%	262	262	262
5	A	16%	239	239	239
6	H	34%	235	235	235
7	A	34%	230	230	230
8	R	34%	239	239	239
9	E	34%	263	263	263
10	K	96%	252	252	252
11	A	96%	207	237	208
12	N	96%	215	225	215
13	A	96%	248	200	246
14	A	96%	259	180	256

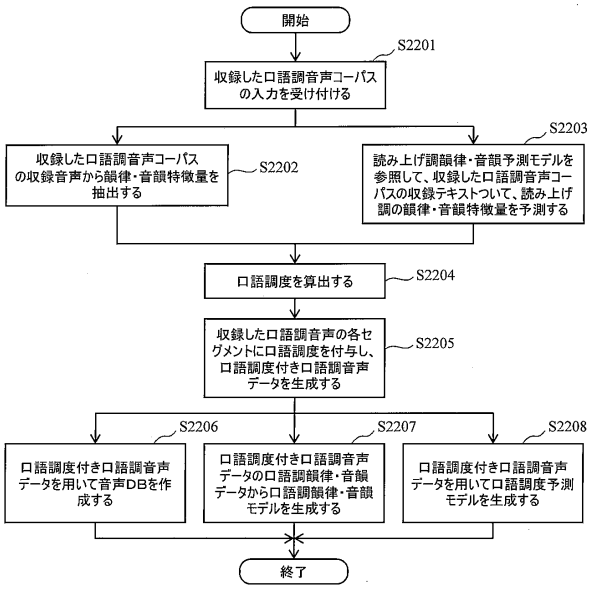
【図２１】

図２１



【図２２】

図２２



フロントページの続き

(56)参考文献 特開２０１３－２４２５１５（ＪＰ，Ａ）
特開２０１３－０１５６９３（ＪＰ，Ａ）
特開２００３－３３７５９２（ＪＰ，Ａ）
特開２０１４－０６２９７０（ＪＰ，Ａ）
特開２０１２－１９８２７７（ＪＰ，Ａ）
特開２００３－３０２９９２（ＪＰ，Ａ）
特開２００４－２２６５０５（ＪＰ，Ａ）
特開２０１４－１６３９７８（ＪＰ，Ａ）

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