

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4621936号
(P4621936)

(45) 発行日 平成23年2月2日(2011.2.2)

(24) 登録日 平成22年11月12日(2010.11.12)

(51) Int.Cl. F I
G 1 O L 13/08 (2006.01) G 1 O L 13/08 1 2 7 D
 G 1 O L 13/08 1 3 1 Z

請求項の数 6 (全 33 頁)

<p>(21) 出願番号 特願2005-269624 (P2005-269624)</p> <p>(22) 出願日 平成17年9月16日 (2005. 9. 16)</p> <p>(65) 公開番号 特開2007-79351 (P2007-79351A)</p> <p>(43) 公開日 平成19年3月29日 (2007. 3. 29)</p> <p>審査請求日 平成20年3月26日 (2008. 3. 26)</p> <p>(出願人による申告) 平成17年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p>	<p>(73) 特許権者 393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2</p> <p>(74) 代理人 100115749 弁理士 谷川 英和</p> <p>(72) 発明者 山岸 順一 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 河井 恒 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 平井 俊男 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p style="text-align: right;">最終頁に続く</p>
---	---

(54) 【発明の名称】 音声合成装置、学習データ生成装置、ポーズ予測装置およびプログラム

(57) 【特許請求の範囲】

【請求項1】

情報が記憶される学習データ記憶部と、
 1以上の文の音声を受け付ける音声受付部と、
 前記音声受付部が受け付けた音声の境界を検出し、当該境界を識別する情報である境界情報を、文ごとに取得する境界決定部と、
 前記境界決定部が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第一属性情報取得部と、
 前記音声受付部が受け付けた音声から無音区間を検知し、前記無音区間の長さに基づいて、前記境界のポーズの種類についての情報であるポーズ情報を境界ごとに取得するポーズ情報取得部と、
 前記ポーズ情報において、前記1種以上の属性情報が現われる確率についての情報である属性情報確率情報を算出し、前記学習データ記憶部に記録する属性情報確率情報算出部と、
 前記ポーズ情報取得部が取得した各境界についてのポーズ情報に基づいて、隣り合う2つの境界についての2つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出し、前記学習データ記憶部に記録するポーズ情報確率情報算出部と、
 最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出し、前記学習データ記憶部に記録する最後境界ポーズ情報確率情報算出部を具備する学習データ生成装置。

10

20

【請求項 2】

属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を有する学習データを格納している学習データ格納部と、
 文を受け付ける文受付部と、
 前記文受付部が受け付けた文を分割し、当該文の境界を識別する情報である境界情報を 1 以上取得する境界情報取得部と、
 前記境界情報取得部が取得した 1 以上の境界情報が示す各境界の 1 種以上の属性情報を取得する第二属性情報取得部と、
 前記境界ごとに前記第二属性情報取得部が取得した 1 種以上の各属性情報に対応する 1 以上の属性情報確率情報を前記学習データ格納部から取得する属性情報確率情報取得部と、
 前記境界情報取得部が取得した 1 以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データ格納部から取得するポーズ情報確率情報取得部と、
 前記最後境界ポーズ情報確率情報を前記学習データ格納部から取得する最後境界ポーズ情報確率情報取得部と、
 前記属性情報確率情報取得部が取得した前記境界ごとの 1 以上の属性情報確率情報、前記ポーズ情報確率情報取得部が取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得部が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定部と、
 前記ポーズ情報系列決定部が決定したポーズ情報に対応した無音区間を構成し、前記文受付部が受け付けた文を音声合成出力する音声合成出力部を具備する音声合成装置。

10

20

【請求項 3】

属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を有する学習データを格納している学習データ格納部と、
 文を受け付ける文受付部と、
 前記文受付部が受け付けた文を分割し、当該文の境界を識別する情報である境界情報を 1 以上取得する境界情報取得部と、
 前記境界情報取得部が取得した 1 以上の境界情報が示す各境界の 1 種以上の属性情報を取得する第二属性情報取得部と、
 前記境界ごとに前記第二属性情報取得部が取得した 1 種以上の各属性情報に対応する 1 以上の属性情報確率情報を前記学習データ格納部から取得する属性情報確率情報取得部と、
 前記境界情報取得部が取得した 1 以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データ格納部から取得するポーズ情報確率情報取得部と、
 前記最後境界ポーズ情報確率情報を前記学習データ格納部から取得する最後境界ポーズ情報確率情報取得部と、
 前記属性情報確率情報取得部が取得した前記境界ごとの 1 以上の属性情報確率情報、前記ポーズ情報確率情報取得部が取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得部が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定部と、
 前記ポーズ情報系列決定部が決定したポーズ情報を出力する出力部を具備するポーズ予測装置。

30

40

【請求項 4】

コンピュータに、
 1 以上の文の音声を受け付ける音声受付ステップと、
 前記音声受付ステップで受け付けた音声の境界を検出し、当該境界を識別する情報である境界情報を、文ごとに取得する境界決定ステップと、
 前記境界決定ステップで取得した 1 以上の境界情報が示す各境界の 1 種以上の属性情報を取得する第一属性情報取得ステップと、
 前記音声受付ステップで受け付けた音声から無音区間を検知し、前記無音区間の長さに基づ

50

づいて、前記境界のポーズの種類についての情報であるポーズ情報を境界ごとに取得するポーズ情報取得ステップと、

前記ポーズ情報において、前記1種以上の属性情報が現われる確率についての情報である属性情報確率情報を算出し、記憶部に記録する属性情報確率情報算出ステップと、

前記ポーズ情報取得ステップで取得した各境界についてのポーズ情報に基づいて、隣り合う2つの境界についての2つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出し、記憶部に記録するポーズ情報確率情報算出ステップと、

最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出し、記憶部に記録する最後境界ポーズ情報確率情報算出ステップを実行させるためのプログラム。

10

【請求項5】

コンピュータに、

文を受け付ける文受付ステップと、

前記文受付ステップで受け付けた文を分割し、当該文の境界を識別する情報である境界情報を1以上取得する境界情報取得ステップと、

前記境界情報取得ステップで取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得ステップと、

前記境界ごとに前記第二属性情報取得ステップで取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を、格納している学習データから取得する属性情報確率情報取得ステップと、

20

前記境界情報取得ステップで取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を、前記学習データから取得するポーズ情報確率情報取得ステップと、前記最後境界ポーズ情報確率情報を前記学習データから取得する最後境界ポーズ情報確率情報取得ステップと、

前記属性情報確率情報取得ステップで取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得ステップで取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得ステップで取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定ステップと、

30

前記ポーズ情報系列決定ステップで決定したポーズ情報に対応した無音区間を構成し、前記文受付ステップが受け付けた文を音声合成出力する音声合成出力ステップを実行させるためのプログラム。

【請求項6】

コンピュータに、

文を受け付ける文受付ステップと、

前記文受付ステップで受け付けた文を分割し、当該文の境界を識別する情報である境界情報を1以上取得する境界情報取得ステップと、

前記境界情報取得ステップで取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得ステップと、

40

前記境界ごとに前記第二属性情報取得ステップで取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を、格納している学習データから取得する属性情報確率情報取得ステップと、

前記境界情報取得ステップで取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データから取得するポーズ情報確率情報取得ステップと、前記最後境界ポーズ情報確率情報を前記学習データから取得する最後境界ポーズ情報確率情報取得ステップと、

前記属性情報確率情報取得ステップで取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得ステップで取得したポーズ情報確率情報、および前記最後

50

境界ポーズ情報確率情報取得ステップで取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定ステップと、

前記ポーズ情報系列決定ステップで決定したポーズ情報を出力する出力ステップを実行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、音声合成出力を行う音声合成装置等に関するものである。

【背景技術】

10

【0002】

テキストデータから音声合成し、音声出力する音声合成装置において、入力テキストの構文情報を合成音声の聞き手に的確に伝えるため、音声のポーズを入力テキストの適切な境界に挿入する処理が必要である。また、音声合成装置において、ポーズ挿入の頻度は過剰もしくは過少のどちらであっても合成音声の聞き手の理解を妨げる要因となるため、ポーズ挿入は、一文もしくは連続する複数文内にどの程度の個数や種類のポーズを挿入するかを適切に判断しながら決定する必要がある。

【0003】

しかし、従来のテキストデータから音声合成し、従来のポーズ予測アルゴリズムは単純に境界毎に与えられた説明変数（後述する属性情報と同意義である）を基にポーズの種類を予測する局所的な予測モデルである（例えば、非特許文献1参照）。

20

【0004】

なお、単純に境界毎に与えられた説明変数に基づいて、ポーズの種類を予測することは、パターン認識における単純な「多クラス」の予測問題である。したがって、多クラスの予測モデルである、k-NN、決定木、Naive Bayes、SVMの従来技術が適用可能である。

【0005】

また、関連する先行技術として、文を言語解析し、句境界の係り受け距離や、先行境界の係り受け距離や、着目する句が並列句であるか否か、着目する句が同格句であるか否か、着目する句のモーラ数を取得する言語解析技術が存在する（非特許文献2参照）。

30

【非特許文献1】海木、他1名、「局所的構造によるポーズ挿入規則化の検討」、1996、信学論、J79-D-II、(9)、pp.1455-1463

【非特許文献2】工藤、他1名、「チャンキングの段階適用による日本語係り受け解析」、2002、情処論、43、(6)、1834-1842

【発明の開示】

【発明が解決しようとする課題】

【0006】

しかしながら、従来の音声合成装置等の音声処理システムにおいては、単純に境界毎にポーズの位置、および種類を予測しているため、発話全体（例えば、一文章）で最適になっていない、という課題があった。

40

【課題を解決するための手段】

【0007】

本第一の発明の学習データ生成装置は、情報が記憶される学習データ記憶部と、1以上の文の音声を受け付ける音声受付部と、前記音声受付部が受け付けた音声の境界を検出し、当該境界を識別する情報である境界情報を、文ごとに取得する境界決定部と、前記境界決定部が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第一属性情報取得部と、前記音声受付部が受け付けた音声から無音区間を検知し、前記無音区間の長さに基づいて、前記境界のポーズの種類についての情報であるポーズ情報を境界ごとに取得するポーズ情報取得部と、前記ポーズ情報において、前記1種以上の属性情報が現われる確率についての情報である属性情報確率情報を算出し、前記学習データ記憶部に記

50

録する属性情報確率情報算出部と、前記ポーズ情報取得部が取得した各境界についてのポーズ情報に基づいて、隣り合う2つの境界についての2つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出し、前記学習データ記憶部に記録するポーズ情報確率情報算出部と、最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出し、前記学習データ記憶部に記録する最後境界ポーズ情報確率情報算出部を具備する学習データ生成装置である。

【0008】

かかる構成により、的確にポーズの位置や種類を予測するための学習データが自動的に生成できる。また、受け付けた文に含まれる境界について、ポーズの種類別に、属性情報を取得し、ポーズの種類別および属性情報別に数を数える処理が、学習データを生成するための主たる処理であるので、非常に少ない計算量で、高速に学習データを得ることができる。

10

【0009】

また、本第二の発明の音声合成装置は、属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を有する学習データを格納している学習データ格納部と、文を受け付ける文受付部と、前記文受付部が受け付けた文を分割し、当該文の境界を識別する情報である境界情報を1以上取得する境界情報取得部と、前記境界情報取得部が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得部と、前記境界ごとに前記第二属性情報取得部が取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を前記学習データ格納部から取得する属性情報確率情報取得部と、前記境界情報取得部が取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データ格納部から取得するポーズ情報確率情報取得部と、前記最後境界ポーズ情報確率情報を前記学習データ格納部から取得する最後境界ポーズ情報確率情報取得部と、前記属性情報確率情報取得部が取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得部が取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得部が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定部と、前記ポーズ情報系列決定部が決定したポーズ情報に対応した無音区間を構成し、前記文受付部が受け付けた文を音声合成出力する音声合成出力部を具備する音声合成装置である。

20

30

【0010】

かかる構成により、発話全体（例えば、一文章）で最適なポーズが挿入された音声合成が実現できる。その結果、極めて自然な音声での音声合成出力が可能となる。また、少ない計算量でポーズ予測が可能となるため、例えば、リアルタイムの音声合成処理の前処理として極めて有効である。

【0011】

また、本第三の発明の音声合成装置は、第二の発明に対して、前記属性情報は、着目する境界の係り受け距離、前記着目する境界の直前の境界である先行境界の係り受け距離、読点の有無を示す情報、前記着目する境界の直前の句が並列句であるか否かを示す情報、前記着目する境界の直前の句が同格句であるか否かを示す情報、前記着目する境界の直前の句のモーラ数、前記着目する境界の直後の句のモーラ数のうち、1以上の情報である音声合成装置であることは好適である。

40

【0012】

また、本第四の発明の音声合成装置は、第二、第三いずれかの発明に対して、前記ポーズ情報は、ポーズであることを示すポーズ有り情報、またはポーズでないことを示すポーズ無し情報のいずれかである音声合成装置であることは好適である。

【0013】

また、本第五の発明のポーズ予測装置は、属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を有する学習データを格納している学習データ格納部と、文を受け付ける文受付部と、前記文受付部が受け付けた文を分割し、当該文の境界を識別する

50

情報である境界情報を1以上取得する境界情報取得部と、前記境界情報取得部が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得部と、前記境界ごとに前記第二属性情報取得部が取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を前記学習データ格納部から取得する属性情報確率情報取得部と、前記境界情報取得部が取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データ格納部から取得するポーズ情報確率情報取得部と、前記最後境界ポーズ情報確率情報を前記学習データ格納部から取得する最後境界ポーズ情報確率情報取得部と、前記属性情報確率情報取得部が取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得部が取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得部が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定部と、前記ポーズ情報系列決定部が決定したポーズ情報を出力する出力部を具備するポーズ予測装置である。

10

かかる構成により、的確にポーズの位置と種類が予測できる。かかるポーズ予測装置は、上述した音声合成装置の他、特に外国人の日本語等の音読の訓練などに利用できる。

【0014】

また、本第六の発明の音声処理システムは、音声合成を行うための学習データを生成する学習データ生成装置と、前記学習データ生成装置が生成した学習データを使用して音声合成を行う音声合成装置を具備する音声処理システムであって、前記学習データ生成装置は、情報が記憶される学習データ記憶部と、1以上の文の音声を受け付ける音声受付部と、前記音声受付部が受け付けた音声の境界を検出し、当該境界を識別する情報である境界情報を、文ごとに取得する境界決定部と、前記境界決定部が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第一属性情報取得部と、前記音声受付部が受け付けた音声から無音区間を検知し、前記無音区間の長さに基づいて、前記境界のポーズの種類についての情報であるポーズ情報を境界ごとに取得するポーズ情報取得部と、前記ポーズ情報において、前記1種以上の属性情報が現われる確率についての情報である属性情報確率情報を算出し、前記学習データ記憶部に記録する属性情報確率情報算出部と、前記ポーズ情報取得部が取得した各境界についてのポーズ情報に基づいて、隣り合う2つの境界についての2つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出し、前記学習データ記憶部に記録するポーズ情報確率情報算出部と、最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出し、前記学習データ記憶部に記録する最後境界ポーズ情報確率情報算出部を具備し、前記音声合成装置は、前記学習データ生成装置から属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を受け付ける学習データ受付部と、前記学習データ受付部が受け付けた学習データを格納する学習データ格納部と、文を受け付ける文受付部と、前記文受付部が受け付けた文を分割し、当該文の境界を識別する情報である境界情報を1以上取得する境界情報取得部と、前記境界情報取得部が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得部と、前記境界ごとに前記第二属性情報取得部が取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を前記学習データ格納部から取得する属性情報確率情報取得部と、前記境界情報取得部が取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データ格納部から取得するポーズ情報確率情報取得部と、前記最後境界ポーズ情報確率情報を前記学習データ格納部から取得する最後境界ポーズ情報確率情報取得部と、前記属性情報確率情報取得部が取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得部が取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得部が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定部と、前記ポーズ情報系列決定部が決定したポーズ情報に対応した無音区間を構成し、前記文受付部が受け付けた文を音声合成出力

20

30

40

50

する音声合成出力部を具備する音声処理システムである。

かかる構成により、学習データを生成し、かつ当該学習データを利用して、的確なポーズの位置や種類の予測が可能となる。そのために、極めて自然な音声合成出力ができる。

【発明の効果】

【0015】

本発明による音声処理システムによれば、発話全体で最適なポーズの位置や種類が決定でき、そのためにより自然な音声合成出力が可能となる。

【発明を実施するための最良の形態】

【0016】

以下、音声処理システム等の実施形態について図面を参照して説明する。なお、実施の形態において同じ符号を付した構成要素は同様の動作を行うので、再度の説明を省略する
10
場合がある。

(実施の形態1)

図1は、本実施の形態における音声処理システムのブロック図である。

音声処理システムは、学習データ生成装置11、音声合成装置12を具備する。

学習データ生成装置11は、音声合成装置12が音声合成処理するための学習データを生成する。

【0017】

学習データ生成装置11は、学習データ記憶部1101、音声受付部1102、境界決定部1103、第一属性情報取得部1104、ポーズ情報取得部1105、属性情報確率
20
情報算出部1106、ポーズ情報確率情報算出部1107、最後境界ポーズ情報確率情報算出部1108を具備する。

【0018】

音声合成装置12は、学習データ受付部1201、学習データ格納部1202、文受付部1203、境界情報取得部1204、第二属性情報取得部1205、属性情報確率情報取得部1206、ポーズ情報確率情報取得部1207、最後境界ポーズ情報確率情報取得部1208、ポーズ情報系列決定部1209、音声合成出力部1210を具備する。

【0019】

学習データ記憶部1101には、情報が記憶される。情報とは、デジタルデータであれば何でも良いが、ここでは後述する学習データが記憶される。学習データ記憶部1101
30
は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。なお、学習データ記憶部1101が揮発性の記録媒体である場合、通常、学習データは、学習データ生成装置11の外部に存在する装置(学習データ生成装置11とは別の装置)の記録媒体に蓄積される。つまり、かかる場合、学習データ生成装置11は、生成した学習データを外部の装置に送信する。

【0020】

音声受付部1102は、文の音声を受け付ける。音声受付部1102は、通常、複数文、文ごとに音声を受け付ける。文の音声とは、例えば、アナウンサーなどの人が、1文ずつ文を朗読した場合に発声される音声である。音声の入力手段は、マイク165やDVD
40
やハードディスクや磁気テープ等の記録媒体からの読み出し手段等、何でも良い。音声受付部1102は、マイク165等の入力手段のデバイスドライバーや等で実現され得る。

【0021】

境界決定部1103は、音声受付部1102が受け付けた音声の境界を検出し、当該境界を識別する情報である境界情報を、文ごとに取得する。詳細には、境界決定部1103は、文の情報(文字のコード列であり、例えば、アナウンサーなどの人が読む文の情報)を保持している。そして、境界決定部1103は、保持している文を読み出し、文中の境界を決定する。そして、境界決定部1103は、境界を識別する境界情報を、文ごとに取得する。なお、境界は、文節、または句、または2以上の所定長の句などである。文を句や文節等に区切る技術は公知技術であるので、詳細な説明は省略する。また、「音声の境界の検出」とは、音声に対応する文の境界の検出でも良い。また、境界決定部1103は
50

、受け付けた音声を音声認識し、文字コード列に変換してから、境界を検出して良い。境界決定部1103は、通常、MPUやメモリ等から実現され得る。なお、音声認識処理は公知技術であるので詳細な説明を省略する。境界決定部1103の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。なお、境界決定部1103は、音声認識処理のために、音響モデルや言語モデルを有する場合がある。

【0022】

第一属性情報取得部1104は、境界決定部1103が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する。第一属性情報取得部1104は、通常、境界決定部1103が保持している文の情報(文字コード列)を利用して、属性情報を取得する。属性情報は、例えば、着目する境界の係り受け距離である。また、属性情報は、例えば、前記着目する境界の直前の境界である先行境界の係り受け距離である。また、属性情報は、例えば、読点の有無を示す情報である。また、属性情報は、例えば、前記着目する境界の直前の句が並列句であるか否かを示す情報である。また、属性情報は、例えば、前記着目する境界の直前の句が同格句であるか否かを示す情報である。また、属性情報は、例えば、前記着目する境界の直前の句のモーラ数である。また、属性情報は、例えば、前記着目する境界の直後の句のモーラ数のうち、1以上の情報である。なお、境界の係り受け距離とは、境界の直前の句や文節等が直接修飾する句や文節等までの句数や文節数等である。先行境界の係り受け距離とは、境界の一つ前の境界の直前の句や文節等が直接修飾する句や文節等までの句数や文節数等である。また、「・・・例えば、A市、B市、C市などでは、・・・」という文中の「A市」、「B市」、「C市」は、並列句である。また、「隣のラーメン屋の店長さん、山田さんは明朗で、・・・」の「店長さん」、「山田さん」は同格句である。モーラ数は、拍数のことである。なお、境界の係り受け距離を取得する技術、並列句であるか否かを示す情報を取得する技術、同格句であるか否かを示す情報を取得する技術は、従来技術(非特許文献2)であるので、詳細な説明は省略する。また、読点の有無を示す情報とは、境界が読点を有するか否かを示す情報であり、かかる情報の取得は、簡単な文字のパターンマッチング(従来技術)で可能である。さらに、モーラ数の取得も、従来技術により可能であるので、詳細な説明は省略する。第一属性情報取得部1104は、通常、MPUやメモリ等から実現され得る。第一属性情報取得部1104の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

【0023】

ポーズ情報取得部1105は、音声受付部1102が受け付けた音声から無音区間を検知し、無音区間の長さに基づいて、境界のポーズの種類についての情報であるポーズ情報を境界ごとに取得する。ポーズ情報は、例えば、ポーズであることを示すポーズ有り情報(例えば「1」)、またはポーズでないことを示すポーズ無し情報(例えば、「0」)のいずれかである。また、ポーズ情報は、例えば、無音区間の長さに応じて、ロングポーズ、ショートポーズ、ポーズ無しの3種類の情報を採りえる、としても良い。なお、ポーズ情報取得部1105は、受け付けた音声と文(テキストコード)との整合をとる。この整合をとった結果を、ポーズ情報取得部1105が利用し、文の境界のうちのポーズの位置を決定する。また、受け付けた音声から無音区間を検出し、無音区間の長さを取得することは、例えば、HMMの音響モデルを用いて可能であり、公知の技術であるので、詳細な説明を省略する。さらに、音声と文(テキストコード)との整合をとる技術も公知技術である。ポーズ情報取得部1105は、音声を音素に分割し、音素の集合が示す文字とテキストコードを比較し、対応付け、かつ取得した無音区間長により境界のポーズの種類を決定する。ポーズ情報取得部1105は、通常、MPUやメモリ等から実現され得る。ポーズ情報取得部1105の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

【0024】

属性情報確率情報算出部 1106 は、ポーズ情報（ポーズ情報の種類ごとに）において、1 種以上の属性情報が現われる確率についての情報である属性情報確率情報を、属性情報の採りえる値ごとに算出し、学習データ記憶部 1101 に記録する。以下、ポーズ情報がポーズ有り情報、ポーズ無し情報の 2 種類の値のみとり得る場合について説明する。また、属性情報を「読点の有無についての情報」である、とする。かかる場合、属性情報確率情報算出部 1106 は、 $P(\text{読点有り} | \text{ポーズ有り})$ 、 $P(\text{読点無し} | \text{ポーズ有り})$ 、 $P(\text{読点有り} | \text{ポーズ無し})$ 、 $P(\text{読点無し} | \text{ポーズ無し})$ の 4 つの確率値を算出する。具体的には、属性情報確率情報算出部 1106 は、境界決定部 1103 が取得した全境界情報が示す全境界の中で、ポーズ情報取得部 1105 が取得したポーズ情報が「ポーズ有り」の個数（ PX ）を算出する。次に、属性情報確率情報算出部 1106 は、「ポーズ有り」の境界の属性値が「読点有り」の個数（ SX ）を算出する。そして、属性情報確率情報算出部 1106 は、「 SX / PX 」を算出し、「 $P(\text{読点有り} | \text{ポーズ有り})$ 」を得る。なお、「 $P(\text{読点有り} | \text{ポーズ有り})$ 」は、境界がポーズである場合に、読点がある確率値である。属性情報確率情報算出部 1106 は、ポーズ情報が「ポーズ有り」の個数、ポーズ情報が「ポーズ無し」の個数、「読点有り」の個数、「読点無し」の個数を算出することにより、 $P(\text{読点無し} | \text{ポーズ有り})$ 、 $P(\text{読点有り} | \text{ポーズ無し})$ 、 $P(\text{読点無し} | \text{ポーズ無し})$ も取得できる。なお、属性情報確率情報は、通常、確率値（0 から 1 までの値）であるが、その他の情報でも良い。その他の情報とは、例えば、確率値を N 倍した値である。

10

【0025】

20

なお、属性情報確率情報算出部 1106 は、属性情報の値が四則演算の対象とならない場合（この場合の属性情報を「質的データ」、属性情報の値を「質的な値」ということとする。）、例えば、最尤推定（ML 推定）を用いて、属性情報確率情報を算出する。具体的には、下記の数式 1 により属性情報確率情報（ P_{ML} ）を算出する。なお、質的な値を、適宜「離散値」ともいうこととする。また、質的な値とは、読点の有無（例えば、読点有りの場合は「1」、読点なしの場合は「0」）、並列句であるか否か（例えば、並列句である場合「1」、並列句で無い場合「0」）、文のスタイル（例えば、肯定文の場合「1」、否定文の場合「2」、疑問文の場合「3」）などの属性情報の値である。

【数 1】

$$P_{ML}(x_{ki}=v_k | o_i=c) = \frac{\sum_n^N \gamma_{nv_kc}}{\sum_{n, v_k}^{N, V_k} \gamma_{nv_kc}}$$

$$P_{ML}(o_{i-1}=c' | o_i=c) = \frac{\sum_n^N \gamma_{nc'c}}{\sum_{n, c'}^{N, C} \gamma_{nc'c}}$$

$$P_{ML}(o_i=c) = \frac{\sum_n^N \gamma_{nc}}{\sum_{n, c}^{N, C} \gamma_{nc}}$$

30

【0026】

40

数式 1 において、受け付ける文が N 個の文で構成され、 x_{ki} が離散値であるとする。そして、 n 番目の文で $x_{ki}^{(n)} = v_k$ 、かつ $o_i^{(n)} = c$ となる回数を γ_{nv_kc} 、 $o_{i-1}^{(n)} = c'$ かつ $o_i^{(n)} = c$ となる回数を $\gamma_{nc'c}$ 、 $o_i^{(n)} = c$ となる回数を γ_{nc} とすると最尤推定量（ P_{ML} ）は、数式 1 により算出できる。

【0027】

なお、数式 1 において、 V_k は k 番目の属性情報（適宜「説明変数」とも言う）の種類数の総数、 C はポーズの種類数の総数（ポーズの有無の予測ならば $C = 2$ ）を示す。例えば、 k 番目の属性情報が読点の有無の情報ならば、 $V_k = 2$ （読点あり / なしの 2 種類）であり、最尤推定量は文の中の読点のある境界の個数もしくは読点の無い境界の個数をポーズの有無別に数え上げれば得られることを示している。

50

【 0 0 2 8 】

ただし、属性情報確率情報算出部 1 1 0 6 は、MAP 推定により属性情報確率情報を算出しても良い。少量の学習データに該当サンプルが無い場合、最尤推定では確率が零となるため、よりロバストな MAP 推定を導入することは好適である。事前分布はディリクレ分布であると仮定すると、モデルパラメータの MAP 定量は、数式 2 により算出できる。

【 数 2 】

$$P_{MAP}(x_{ki}=v_k | o_i=c) = \frac{\sum_n^N \gamma_{nv_kc} + \alpha - 1}{\sum_{n, v_k}^{N, V_k} \gamma_{nv_kc} + V_k(\alpha - 1)}$$

$$P_{MAP}(o_{i-1}=c' | o_i=c) = \frac{\sum_n^N \gamma_{nc'c} + \alpha - 1}{\sum_{n, c'}^{N, C} \gamma_{nc'c} + C(\alpha - 1)}$$

$$P_{MAP}(o_i=c) = \frac{\sum_n^N \gamma_{nc} + \alpha - 1}{\sum_{n, c}^{N, C} \gamma_{nc} + C(\alpha - 1)}$$

10

数式 2 において、 α は正の定数のハイパーパラメータであり、例えば、 $\alpha = 2$ (ラプラススムージング) である。

【 0 0 2 9 】

また、属性情報確率情報算出部 1 1 0 6 は、属性情報の値が、四則演算が可能な値の場合(この場合の属性情報を「量的データ」、属性情報の値を「量的な値」ということとする。)、例えば、密度推定を行う。属性情報確率情報算出部 1 1 0 6 は、正規分布を用いることも可能であるが、 x_k が正規分布の条件を満たしていない場合も存在する。そこで、属性情報確率情報算出部 1 1 0 6 は、好ましくは、カーネル密度推定(「核密度推定」とも言う。)を利用する。なお、カーネル密度推定は公知技術であるので、詳細な説明は省略する。属性情報確率情報算出部 1 1 0 6 は、カーネル密度推定を用いて、数式 3 のように属性情報確率情報を算出する。なお、量的な値を、適宜「連続値」ともいうこととする。また、量的な値とは、係り受け距離(例えば、「1」「2」「3」・・・(自然数))、話速(例えば、5 m o r a / s e c (実数))などの属性情報の値である。

20

【 数 3 】

$$P(x_k=x | o=c) = \frac{1}{Nh} \sum_{n=1}^N K\left[\frac{x-x^{(n)}}{h}\right]$$

数式 3 において、 h は b a n d w i d t h パラメータである。また、カーネル関数(「核密度関数」とも言う。)には、例えば、標準正規分布「 $K(x) = N(x; 0, 1)$ 」を利用する。

【 0 0 3 0 】

属性情報確率情報算出部 1 1 0 6 は、通常、MPU やメモリ等から実現され得る。属性情報確率情報算出部 1 1 0 6 の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアは ROM 等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

40

【 0 0 3 1 】

ポーズ情報確率情報算出部 1 1 0 7 は、ポーズ情報取得部 1 1 0 5 が取得した各境界についてのポーズ情報に基づいて、隣り合う 2 つの境界についての 2 つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出し、学習データ記憶部 1 1 0 1 に記録する。ポーズ情報確率情報算出部 1 1 0 7 は、 $P(o_{i-1} | o_i)$ を算出する。 $P(o_{i-1} | o_i)$ は、ポーズ情報確率情報である。ポーズ情報確率情報は、例えば、 i 番目の境界のポーズ情報(2 種類または 3 種類)の場合の、 $(i-1)$ 番目の境界のポーズ情報(2 種類または 3 種類)の確率値である。ポーズ情報が「ポーズ有り」と「ポーズ無し」の 2 種類の場合、ポーズ情報確率情報は、 $P((i-1)$ 番目の境界がポーズ有

50

り | i 番目の境界がポーズ有り)、 $P((i-1)$ 番目の境界がポーズ無し | i 番目の境界がポーズ有り)、 $P((i-1)$ 番目の境界がポーズ有り | i 番目の境界がポーズ無し)、 $P((i-1)$ 番目の境界がポーズ無し | i 番目の境界がポーズ無し) の 4 種類である。「 $P((i-1)$ 番目の境界がポーズ有り | i 番目の境界がポーズ有り)」は、 i 番目の境界がポーズ有りの場合の、 $(i-1)$ 番目の境界がポーズ有りの確率値である。なお、ポーズ情報確率情報は、必ずしも「0 から 1」の値であるとは限らず、いわゆる確率値(0 から 1 の値)を N 倍して情報等でも良い。また、ポーズ情報確率情報算出部 1107 は、例えば、ポーズ情報取得部 1105 が取得した各境界についてのポーズ情報が「ポーズ有り」か「ポーズ無し」かをチェックし、 $P((i-1)$ 番目の境界がポーズ有り | i 番目の境界がポーズ有り)等を算出する。ポーズ情報確率情報算出部 1107 は、通常、1 以上の文の全境界についてのポーズ情報を、処理対象とする。ポーズ情報確率情報算出部 1107 は、通常、MPU やメモリ等から実現され得る。ポーズ情報確率情報算出部 1107 の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアは ROM 等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

10

【0032】

最後境界ポーズ情報確率情報算出部 1108 は、文中の境界のうち、文の最後に存在する境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出し、学習データ記憶部 1101 に記録する。ポーズ情報が「ポーズ有り」、「ポーズ無し」の 2 種類の情報である場合、最後境界ポーズ情報確率情報算出部 1108 は、1 以上の文の最後尾の境界のポーズ情報が「ポーズ有り」の数を算出する。また、最後境界ポーズ情報確率情報算出部 1108 は、1 以上の文の最後尾の境界のポーズ情報が「ポーズ無し」の数を算出する。そして、 $P(\text{最後尾がポーズ有り})$ 、 $P(\text{最後尾がポーズ無し})$ を算出する。 $P(\text{最後尾がポーズ有り})$ は、最後の境界が「ポーズ有り」の確率値である。最後境界ポーズ情報確率情報算出部 1108 は、通常、MPU やメモリ等から実現され得る。最後境界ポーズ情報確率情報算出部 1108 の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアは ROM 等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

20

【0033】

学習データ受付部 1201 は、学習データ生成装置 11 が生成した属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を受け付ける。属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を、ここでは、学習データという。ここで、「受け付け」とは、受信でも良いし、記録媒体を介して取得することも含む。学習データ受付部 1201 は、無線または有線の通信手段等で実現され得る。

30

【0034】

学習データ格納部 1202 は、学習データ受付部 1201 が受け付けた学習データを格納している。学習データ格納部 1202 が学習データを格納する処理を行うと考えるても良いし、学習データ受付部 1201 が学習データを格納する処理を行うと考えるても良い。学習データ格納部 1202 は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0035】

文受付部 1203 は、文を受け付ける。文とは、通常、文を構成する文字コード列である。文の入力手段は、テンキーやキーボードやマウスやメニュー画面によるもの等、何でも良い。文受付部 1203 は、テンキーやキーボード等の入力手段のデバイスドライバ等で実現され得る。

40

【0036】

境界情報取得部 1204 は、文受付部 1203 が受け付けた文(文字コード列)を分割し、当該文の境界を識別する情報である境界情報を 1 以上取得する。文字コード列である文を、文節や句などに分割し、境界情報を取得する技術は公知技術(形態素解析技術)であるので、詳細な説明を省略する。境界情報取得部 1204 は、通常、MPU やメモリ等から実現され得る。境界情報取得部 1204 の処理手順は、通常、ソフトウェアで実現さ

50

れ、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

【0037】

第二属性情報取得部1205は、境界情報取得部1204が取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する。属性情報の例は、上述した通りである。第二属性情報取得部1205の処理は、第一属性情報取得部1104の処理と同様である。第二属性情報取得部1205は、通常、MPUやメモリ等から実現され得る。第二属性情報取得部1205の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

10

【0038】

属性情報確率情報取得部1206は、境界ごとに第二属性情報取得部1205が取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を学習データ格納部1202から検索し、取得する。例えば、着目する境界が「読点有り」の場合、属性情報確率情報取得部1206は、P（読点有り | ポーズ有り）、およびP（読点有り | ポーズ無し）を取得する。また、例えば、着目する境界が「読点無し」の場合、属性情報確率情報取得部1206は、P（読点無し | ポーズ有り）、およびP（読点無し | ポーズ無し）を取得する。

属性情報確率情報取得部1206は、通常、MPUやメモリ等から実現され得る。属性情報確率情報取得部1206の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

20

【0039】

ポーズ情報確率情報取得部1207は、各境界についてのポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を学習データ格納部1202から取得する。例えば、i番目の境界が「ポーズ有り」の場合、ポーズ情報確率情報取得部1207は、P（(i-1)番目の境界がポーズ有り | i番目の境界がポーズ有り）、P（(i-1)番目の境界がポーズ無し | i番目の境界がポーズ有り）を、「i番目の境界が「ポーズ有り」」をキーとして、検索し、取得する。ポーズ情報確率情報取得部1207は、通常、MPUやメモリ等から実現され得る。ポーズ情報確率情報取得部1207の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

30

【0040】

最後境界ポーズ情報確率情報取得部1208は、最後境界ポーズ情報確率情報を学習データ格納部1202から取得する。最後境界ポーズ情報確率情報取得部1208は、通常、MPUやメモリ等から実現され得る。最後境界ポーズ情報確率情報取得部1208の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

【0041】

ポーズ情報系列決定部1209は、属性情報確率情報取得部1206が取得した境界ごとの1以上の属性情報確率情報、ポーズ情報確率情報取得部1207が取得したポーズ情報確率情報、および最後境界ポーズ情報確率情報取得部1208が取得した最後境界ポーズ情報確率情報を用いて、所定の演算を行い、演算結果が最も良好なポーズ情報の系列を決定する。ここで演算とは、1以上の属性情報確率情報、ポーズ情報確率情報、最後境界ポーズ情報確率情報をパラメータとする演算処理（例えば、「f（1以上の属性情報確率情報，ポーズ情報確率情報，最後境界ポーズ情報確率情報）」）である。この演算処理は、上記確率情報の積が最大となるポーズ情報の系列を決定する処理である。この決定する処理は、動的計画法（DP）により行う。なお、ポーズ情報が「ポーズ有り」、「ポーズ無し」の2種類の場合、ポーズ情報の系列とは、各境界が「ポーズ有り」または「ポーズ

40

50

無し」のいずれかであることを示す情報列である。

なお、ポーズ情報系列決定部 1209 は、具体的には、例えば、数式 4 によりポーズ情報の系列を算出する。

【数 4】

$$O^* = (o_1^*, \dots, o_I^*) = \underset{O}{\operatorname{argmax}} P(O | X)$$

【0042】

数式 4 において、一文の総境界数を I 、各境界の説明変数の系列を $X = (x_1, \dots, x_I)$ とし、 X が所与のもとでの一文のポーズ系列 $O = (o_1, \dots, o_I)$ の確率 $P(O | X)$ を最大にする O^* を求めることによりポーズの位置（個数）を決定することを示している。

10

【0043】

ただし、各境界の説明変数の集合 x_i は K 個の説明変数 $x_i = (x_{1i}, \dots, x_{Ki})$ から構成されるとする。数式 4 を厳密に解くことは困難であるため、ここでは以下のようなモデル化により、効率的なポーズ予測を実現する。まず、説明変数および各境界には一次マルコフ過程（図 18 参照）があるとする。ここで o_0 は文頭のサイレンスを表す。これを図 19 のモデルによりモデル化する。 x_i と o_{i-1} が独立と仮定されていることを考慮すると数式 4 は、数式 5 のように変形できる。

【数 5】

20

$$\begin{aligned} \underset{O}{\operatorname{argmax}} P(O | X) &= \underset{O}{\operatorname{argmax}} \prod_{i=1}^I P_i(o_i | x_i, o_{i-1}) \\ &= \underset{O}{\operatorname{argmax}} \prod_{i=1}^I \frac{P_i(x_i, o_{i-1} | o_i) P_i(o_i)}{P_i(x_i) P_i(o_{i-1})} \\ &\approx \underset{O}{\operatorname{argmax}} \prod_{i=1}^I \frac{\prod_{k=1}^K P(x_{ki} | o_i) P(o_{i-1} | o_i) P(o_i)}{P(x_i) P(o_{i-1})} \\ &= \underset{O}{\operatorname{argmax}} \prod_{i=1}^I \prod_{k=1}^K P(x_{ki} | o_i) P(o_{i-1} | o_i) P(o_i) \end{aligned}$$

30

【0044】

数式 5 において、 O と独立な項は省略している。このモデル化では境界 $i-1$ におけるポーズの確率は後続の境界 o_i に依存するため、ポーズ情報系列決定部 1209 は、動的計画法（DP）を用いて確率最大となるポーズの系列を決定する。

また、数式 5 において、「 $P(x_{ki} | o_i)$ 」は、属性情報確率情報を示す。また、「 $P(o_{i-1} | o_i)$ 」は、ポーズ情報確率情報を示す。「 $P(o_i)$ 」は、最後境界ポーズ情報確率情報を示す。

【0045】

なお、ポーズ情報系列決定部 1209 は、通常、MPU やメモリ等から実現され得る。ポーズ情報系列決定部 1209 の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアは ROM 等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。なお、DP を用いた演算処理は、公知技術であるので、詳細な説明を省略する。

40

【0046】

音声合成出力部 1210 は、ポーズ情報系列決定部 1209 が決定したポーズ情報に対応した無音区間を構成し、文受付部 1203 が受け付けた文を音声合成出力する。文（文字コード列）を音声合成し、音声出力する技術は公知技術である。また、ポーズの有無についての情報であるポーズ情報を参照し、「ポーズ有り」の場合はポーズを挿入（無音を挿入）し、音声合成する技術も公知技術であるので、詳細な説明を省略する。また、「無

50

音区間を構成する」とは、例えば、文（文字コード列）に、無音を識別するコード（情報）を挿入することである。また、「無音区間を構成する」とは、音声出力の際に、無音の音声を示す情報を挿入することでも良い。音声合成出力部 1210 は、スピーカ等出力デバイスを含むと考えても含まないと考えても良い。音声合成出力部 1210 は、出力デバイスのドライバーソフトまたは、出力デバイスのドライバーソフトと出力デバイス等で実現され得る。

次に、音声処理システムの動作について説明する。まず、学習データ生成装置 11 の動作について、図 2 から図 5 のフローチャートを用いて説明する。

（ステップ S201）音声受付部 1102 は、文の音声を受け付けたか否かを判断する。文の音声を受け付ければステップ S202 に行き、文の音声を受け付けなければステップ S211 に行く。

10

（ステップ S202）境界決定部 1103 は、受け付けた文の音声に対応する文（文字コード列）を読み出す。

【0047】

（ステップ S203）境界決定部 1103 は、ステップ S202 で読み出した文を言語解析し、文節等の境界を得る。具体的には、例えば、境界決定部 1103 は、境界を識別する境界情報（例えば、文節の先頭のオフセットや、文節の先頭の文頭からの文字数等）を得る。また、ここで、境界決定部 1103 は、境界の数も取得する、とする。

（ステップ S204）ポーズ情報取得部 1105 は、ステップ S201 で受け付けた音声に対して、音声処理し、音声と文（ステップ S202 で読み出した文）とのアライメントをとる。

20

（ステップ S205）ポーズ情報取得部 1105 は、カウンタ i に 1 を代入する。

（ステップ S206）ポーズ情報取得部 1105 は、変数 I に、ステップ S203 で得た境界の数を代入する。

（ステップ S207）ポーズ情報取得部 1105 は、「 $i > I$ 」を満たすか否かを判断する。「 $i > I$ 」を満たせばステップ S201 に戻り、「 $i > I$ 」を満たさなければステップ S208 に行く。

【0048】

（ステップ S208）ポーズ情報取得部 1105 は、 i 番目の境界のポーズ情報を取得する。ポーズ情報取得部 1105 は、例えば、 i 番目の境界に対応する音声が無音区間を有すればポーズ有りとして判断し、 i 番目の境界に対応する音声が無音区間を有さない、または所定長未満の無音区間しか有さなければポーズ無しとして判断する。また、ポーズ情報取得部 1105 は、無音区間の長さにより、ロングポーズ、ショートポーズ、ポーズ無しの 3 段階に区分して、ポーズ情報を取得（決定）しても良い。

30

（ステップ S209）第一属性情報取得部 1104 は、 i 番目の境界の 1 種類以上の属性情報を取得する。

（ステップ S210）第一属性情報取得部 1104 は、カウンタ i を 1、インクリメントする。そして、ステップ S207 に戻る。

【0049】

（ステップ S211）音声受付部 1102 は、音声の受付を終了するか否かを判断する。例えば、ユーザの指示（例えば、所定のボタンが押下されるなど）により、音声の受付が終了した、と決定される。また、音声の受付の終了は、例えば、所定時間以上、音声を受け付けなかった場合に、音声の受付が終了した、と決定される。

40

【0050】

（ステップ S212）属性情報確率情報算出部 1106 は、ポーズ情報において、1 種以上の属性情報が示す確率についての情報である属性情報確率情報を算出する。属性情報確率情報を算出する処理の詳細については、図 3 のフローチャートを用いて説明する。

【0051】

（ステップ S213）ポーズ情報確率情報算出部 1107 は、ポーズ情報取得部 1105 が取得した各境界についてのポーズ情報に基づいて、隣り合う 2 つの境界についての 2

50

つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出する。ポーズ情報確率情報を算出する処理の詳細については、図4のフローチャートを用いて説明する。

【0052】

(ステップS214) 最後境界ポーズ情報確率情報算出部1108は、最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出する。最後境界ポーズ情報確率情報を算出する処理の詳細については、図5のフローチャートを用いて説明する。

【0053】

(ステップS215) ステップS212、ステップS213、ステップS214で算出した学習データ(属性情報確率情報、ポーズ情報確率情報、最後境界ポーズ情報確率情報)を、学習データ記憶部1101に記録する。そして、処理を終了する。

なお、図2のフローチャートにおいて、ステップS212、ステップS213、ステップS214の処理順序は問わない。

次に、ステップS212における属性情報確率情報を算出する処理の詳細について、図3のフローチャートを用いて説明する。

(ステップS301) 属性情報確率情報算出部1106は、カウンタ*i*に1を代入する。

【0054】

(ステップS302) 属性情報確率情報算出部1106は、*i*番目の種類の属性情報が存在するか否かを判断する。*i*番目の種類の属性情報が存在すればステップS303に行き、*i*番目の種類の属性情報が存在しなければ上位関数にリターンする。属性情報の種類は、例えば、以下の7種類である。つまり、1)着目する境界の係り受け距離、2)着目する境界の直前の境界である先行境界の係り受け距離、3)読点の有無を示す情報、4)着目する境界の直前の句が並列句であるか否かを示す情報、5)着目する境界の直前の句が同格句であるか否かを示す情報、6)着目する境界の直前の句のモーラ数、7)着目する境界の直後の句のモーラ数、である。

(ステップS303) 属性情報確率情報算出部1106は、カウンタ*j*に1を代入する。

【0055】

(ステップS304) 属性情報確率情報算出部1106は、*j*番目のポーズ情報(ポーズ情報の候補値)が存在するか否かを判断する。ポーズ情報が「ポーズ有り」「ポーズ無し」しか取り得ない場合、*j*は「2」までインクリメントされる。*j*番目のポーズ情報の候補値が存在すればステップS305に行き、*j*番目のポーズ情報の候補値が存在しなければステップS310に行く。なお、ポーズ情報の候補値または/およびポーズ情報の候補値の数は、予め属性情報確率情報算出部1106が保持している、とする。

(ステップS305) 属性情報確率情報算出部1106は、カウンタ*k*に1を代入する。

【0056】

(ステップS306) 属性情報確率情報算出部1106は、*i*番目の種類の属性情報の*k*番目の属性値(*k*番目の属性情報の値の候補値)が存在するか否かを判断する。存在すればステップS307に行き、存在しなければステップS309に行く。

【0057】

(ステップS307) 属性情報確率情報算出部1106は、 $P(k\text{番目の属性情報の値} | j\text{番目のポーズ情報の値})$ を算出する。好ましくは、属性情報確率情報算出部1106は、属性情報が質的データである場合、上述したML推定やMAP推定により $P(k\text{番目の属性情報の値} | j\text{番目のポーズ情報の値})$ を算出する。また、属性情報確率情報算出部1106は、属性情報が量的データである場合、上述した密度推定により $P(k\text{番目の属性情報の値} | j\text{番目のポーズ情報の値})$ を算出する。密度推定は、例えば、上述したカーネル密度推定である。なお、属性情報確率情報算出部1106は、属性情報ごとに、質的

10

20

30

40

50

データであるか、量的データであるかを識別する情報を保持している、とする。

(ステップS308) 属性情報確率情報算出部1106は、カウンタkを1、インクリメントする。そして、ステップS306に戻る。

(ステップS309) 属性情報確率情報算出部1106は、カウンタjを1、インクリメントする。そして、ステップS304に戻る。

(ステップS310) 属性情報確率情報算出部1106は、カウンタiを1、インクリメントする。そして、ステップS302に戻る。

次に、ステップS213における、ポーズ情報確率情報を算出する処理の詳細について、図4のフローチャートを用いて説明する。

(ステップS401) ポーズ情報確率情報算出部1107は、カウンタiに1を代入する。 10

【0058】

(ステップS402) ポーズ情報確率情報算出部1107は、i番目の種類のポーズ情報(ポーズ情報の候補値)が存在するか否かを判断する。i番目の種類のポーズ情報が存在すればステップS403に行き、存在しなければ上位関数にリターンする。なお、ポーズ情報の候補値の例として、「ポーズ有り」および「ポーズ無し」がある。

(ステップS403) ポーズ情報確率情報算出部1107は、カウンタjに1を代入する。

【0059】

(ステップS404) ポーズ情報確率情報算出部1107は、j番目の種類のポーズ情報(ポーズ情報の候補値)が存在するか否かを判断する。j番目の種類のポーズ情報が存在すればステップS405に行き、存在しなければステップS407に行く。 20

【0060】

(ステップS405) ポーズ情報確率情報算出部1107は、P(直前の境界がj番目のポーズ情報の値|一の境界がi番目のポーズ情報の値)を算出する。ここで、「直前の境界」とは、一の境界の直前の境界である。つまり、ポーズ情報確率情報算出部1107は、一の境界がi番目のポーズ情報の境界の数(x1とする)を算出する。そして、ポーズ情報確率情報算出部1107は、一の境界がi番目のポーズ情報であり、その直前の境界がj番目のポーズ情報である境界の数(x2とする)を算出する。そして、ポーズ情報確率情報算出部1107は、(x2/x1)を算出する。この(x2/x1)が、P(直前の境界がj番目のポーズ情報|一の境界がi番目のポーズ情報)である。 30

(ステップS406) ポーズ情報確率情報算出部1107は、カウンタjを1、インクリメントする。そして、ステップS404に戻る。

(ステップS407) ポーズ情報確率情報算出部1107は、カウンタiを1、インクリメントする。そして、ステップS402に戻る。

次に、ステップS214における最後境界ポーズ情報確率情報を算出する処理の詳細について、図5のフローチャートを用いて説明する。

【0061】

(ステップS501) 最後境界ポーズ情報確率情報算出部1108は、初期化の処理を行う。初期化とは、カウンタiに1を代入し、変数Nに文の数を代入することである。ここで、文の数とは、音声受付部1102が受け付けた文の数である。文の数は、音声受付部1102が取得しても良いし、境界決定部1103が読み出した文の数から取得しても良い。 40

【0062】

(ステップS502) 最後境界ポーズ情報確率情報算出部1108は、i番目の種類のポーズ情報(ポーズ情報の候補値)が存在するか否かを判断する。i番目の種類のポーズ情報(候補値)が存在すればステップS503に行き、i番目の種類のポーズ情報(候補値)が存在しなければ上位関数にリターンする。

(ステップS503) 最後境界ポーズ情報確率情報算出部1108は、文の最後の境界のポーズ情報がj番目の種類のポーズ情報である数(xとする)を算出する。 50

(ステップS504) 最後境界ポーズ情報確率情報算出部1108は、「 x/N 」を算出する。「 x/N 」が、 i 番目の種類のポーズ情報(候補値)に対応する最後境界ポーズ情報確率情報である。

(ステップS505) 最後境界ポーズ情報確率情報算出部1108は、カウンタ i を1、インクリメントし、ステップS502に戻る。

【0063】

次に、音声合成装置12の動作について、図6、図7のフローチャートを用いて説明する。なお、図6のフローチャートにおいて、音声合成装置12の学習データ受付部1201は、学習データ生成装置11から属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を受け付け、学習データ格納部1202に学習データが格納されている、とする。音声合成装置12は、例えば、記録媒体を經由して、学習データ生成装置11が生成した学習データを受け付けても良い。

10

(ステップS601) 文受付部1203は、文を受け付けたか否かを判断する。文を受け付ければステップS602に行き、文を受け付けなければステップS601に戻る。

(ステップS602) 境界情報取得部1204は、ステップS601で受け付けた文を分割し(境界を取得し)、当該文の境界を識別する情報である境界情報を1以上取得する。

(ステップS603) 第二属性情報取得部1205は、カウンタ i に1を代入する。

【0064】

(ステップS604) 第二属性情報取得部1205は、受け付けた文中に、 i 番目の境界が存在するか否かを判断する。 i 番目の境界が存在するか否かは、例えば、ステップS602で取得した境界情報のうち、 i 番目の境界情報が存在するか否かにより判断する。 i 番目の境界が存在すればステップS605に行き、 i 番目の境界が存在しなければステップS608に行く。

20

(ステップS605) 第二属性情報取得部1205は、 i 番目の境界に対応する全属性情報を取得する。全属性情報とは、例えば、上述した7種類の属性情報である。

(ステップS606) 第二属性情報取得部1205は、ステップS605で取得した全属性情報を、メモリ等に一時格納する。

(ステップS607) 第二属性情報取得部1205は、カウンタ i を1、インクリメントする。そして、ステップS604に戻る。

30

(ステップS608) 属性情報確率情報取得部1206は、属性情報確率情報を取得する。属性情報確率情報を取得する処理の詳細は、図7のフローチャートを用いて説明する。

【0065】

(ステップS609) ポーズ情報確率情報取得部1207は、各境界についてのポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を学習データ格納部1202から読み出す。ポーズ情報が「ポーズ有り」「ポーズ無し」の場合、ポーズ情報確率情報取得部1207が取得するポーズ情報確率情報は、4種類である。4種類とは、 P (直前の境界がポーズ有り | 一の境界がポーズ有り)、 P (直前の境界がポーズ無し | 一の境界がポーズ有り)、 P (直前の境界がポーズ有り | 一の境界がポーズ無し)、 P (直前の境界がポーズ無し | 一の境界がポーズ無し)の4種類である。

40

【0066】

(ステップS610) 最後境界ポーズ情報確率情報取得部1208は、最後境界ポーズ情報確率情報を学習データ格納部1202から読み出す。ポーズ情報が「ポーズ有り」「ポーズ無し」の場合、最後境界ポーズ情報確率情報は、 P (最後の境界のポーズ情報がポーズ有り)、および P (最後の境界のポーズ情報がポーズ無し)の2種類である。

【0067】

(ステップS611) ポーズ情報系列決定部1209は、属性情報確率情報取得部1206が取得した境界ごとの1以上の属性情報確率情報、ポーズ情報確率情報取得部120

50

7 が取得したポーズ情報確率情報、および最後境界ポーズ情報確率情報取得部 1208 が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定する。ポーズ情報の系列のデータ構造は問わない。ポーズ情報の系列は、例えば、ポーズ有りの境界を識別する 1 以上の境界情報でも良い。

【0068】

(ステップ S612) 音声合成出力部 1210 は、ポーズ情報系列決定部 1209 が決定したポーズ情報に対応した無音区間を構成し、文受付部 1203 が受け付けた文を音声合成出力する。そして、処理を終了する。

なお、図 6 のフローチャートにおいて、受け付けた文中に、境界が 1 以上あるものとする。

10

次に、ステップ S608 の属性情報確率情報を取得する処理の詳細について、図 7 のフローチャートを用いて説明する。

(ステップ S701) 属性情報確率情報取得部 1206 は、カウンタ i に 1 を代入する。

(ステップ S702) 属性情報確率情報取得部 1206 は、 i 番目の境界が存在するかどうかを判断する。 i 番目の境界が存在すればステップ S703 に行き、 i 番目の境界が存在しなければ上位関数にリターンする。

(ステップ S703) 属性情報確率情報取得部 1206 は、カウンタ j に 1 を代入する。

【0069】

20

(ステップ S704) 属性情報確率情報取得部 1206 は、 j 番目の種類の属性情報が存在するか否かを判断する。 j 番目の種類の属性情報が存在すればステップ S705 に行き、 j 番目の種類の属性情報が存在しなければステップ S712 に行く。

(ステップ S705) 属性情報確率情報取得部 1206 は、カウンタ k に 1 を代入する。

【0070】

(ステップ S706) 属性情報確率情報取得部 1206 は、 k 番目のポーズ情報の候補値が存在するか否かを判断する。 k 番目のポーズ情報の候補値が存在すればステップ S707 に行き、 k 番目のポーズ情報の候補値が存在しなければステップ S711 に行く。

(ステップ S707) 属性情報確率情報取得部 1206 は、 i 番目の境界の j 番目の種類の属性情報の値を取得する。かかる値は、ステップ S605 で取得されている。

30

(ステップ S708) 属性情報確率情報取得部 1206 は、 P (属性情報の値 | k 番目のポーズ情報の値) を、学習データ格納部 1202 から検索し、取得する。

(ステップ S709) 属性情報確率情報取得部 1206 は、ステップ S708 で取得した確率情報 (P (属性情報の値 | k 番目のポーズ情報の値)) を、メモリ等に一時格納する。

(ステップ S710) 属性情報確率情報取得部 1206 は、カウンタ k を 1、インクリメントし、ステップ S706 に戻る。

(ステップ S711) 属性情報確率情報取得部 1206 は、カウンタ j を 1、インクリメントし、ステップ S704 に戻る。

40

(ステップ S712) 属性情報確率情報取得部 1206 は、カウンタ i を 1、インクリメントし、ステップ S702 に戻る。

以下、本実施の形態における音声処理システムの具体的な動作について説明する。

まず、学習データ生成装置 11 の動作について説明する。

【0071】

今、図 8 に示す (A) から (D) の 4 つの文を、人が読み上げる、とする。文の区切りは、読み上げる人が、例えば、学習データ生成装置 11 が有するボタン (図示しない) を押下することにより指示しても良いし、複数の文を、連続して読み上げた後、人手により、文の区切りの情報を挿入しても良いし、所定時間長以上の無音区間を挿入することにより、学習データ生成装置 11 が文の区切りであると、自動判定しても良い。なお、本自動

50

判定は、無音区間長を取得できれば良いので、公知技術により可能である。

【0072】

また、本具体例において、ポーズ情報は、「ポーズ有り」「ポーズ無し」の2種類の情報が採りえる、とする。また、文の境界は、文節である、とする。そして、属性情報は、「読点の有無を示す情報」（適宜「読点の有無」という。）、および「並列句であるか否かを示す情報」（適宜「並列句であるか否か」という。）の2種類である、とする。

そして、学習データ生成装置11の音声受付部1102は、人が読み上げた音声を、文ごとに受け付ける。

【0073】

次に、境界決定部1103は、図8に示す4つの文の情報（文字コード列）を読み出す。なお、4つの文の情報は、予め学習データ生成装置11の記録媒体に格納されている。そして、境界決定部1103は、各文を文節に区切り、各文節を示す境界情報を取得する。境界情報は、例えば、文の先頭からのバイト数である。このバイト数の情報は、文の先頭から文節の先頭のオフセットである。

10

そして、境界決定部1103は、図9に示すような境界を得る。図9において(1)から(16)が境界である。

【0074】

次に、ポーズ情報取得部1105は、人が読み上げて、受け付けた音声を音声処理し、無音区間長を検知し、所定時間以上の無音区間の箇所をポーズがあった境界であると判断する。そして、ポーズ情報取得部1105は、図10に示す「」の境界をポーズ有りの境界であると判断する。そして、ポーズ情報取得部1105は、各境界に対して、「ポーズ有り」の場合には「1」、ポーズ無しの場合には「0」を書き込む。そして、ポーズ情報取得部1105は、図11に示す第一境界情報管理表を得る。第一境界情報管理表は、「ID」「ポーズ情報」を有するレコードを複数有する。「ID」は、境界（レコード）を識別する情報であり、表におけるレコード管理のために存在する。「ポーズ情報」は、境界がポーズ有るか、ポーズ無しかを示す情報である。

20

【0075】

次に、第一属性情報取得部1104は、各境界の属性情報を取得する。属性情報は、ここでは、「読点の有無を示す情報」、および「並列句であるか否かを示す情報」の2種類である。そして、第一属性情報取得部1104は、図12の第二境界情報管理表を得る。第二境界情報管理表は、「ID」「ポーズ情報」「読点の有無」「並列句であるか否か」の属性値を有するレコードを複数有する。「読点の有無」の属性値は、境界に読点がある場合に「1」、境界に読点がない場合に「0」となる。「並列句であるか否か」の属性値は、境界の前後の句が並列句である場合に「1」、境界の前後の句が並列句でない場合に「0」となる。なお、本音声合成装置12において、第一境界情報管理表と第二境界情報管理表は、一の表であることは好適である。

30

【0076】

次に、属性情報確率情報算出部1106は、まず、各ポーズ情報において、属性情報「読点の有無」が示す確率についての情報である属性情報確率情報を算出する。つまり、属性情報確率情報算出部1106は、図12の表の情報から、(1)P（現在の境界に読点あり | 現在の境界にポーズあり）、(2)P（現在の境界に読点なし | 現在の境界にポーズあり）、(3)P（現在の境界に読点あり | 現在の境界にポーズなし）、(4)P（現在の境界に読点なし | 現在の境界にポーズなし）を算出する。なお、「P（現在の境界に読点あり | 現在の境界にポーズあり）」とは、現在の境界（着目している境界）が「ポーズ有り」である場合の、現在の境界が「読点有り」である確率である。図12において、「ポーズ有り」（ポーズ情報が「1」）の境界は、(1)(4)(7)(8)(11)(13)(15)の7つである。また、上記7つの境界の中で、属性「読点の有無」の属性値が「1」（つまり、「読点有り」）のものは、(1)(7)(11)(15)の4つである。したがって、属性情報確率情報算出部1106は、「P（現在の境界に読点あり | 現在の境界にポーズあり）= 4 / 7」と算出する。同様に、図12の表の情報から、属性

40

50

情報確率情報算出部 1106 は、「 $P(\text{現在の境界に読点なし} | \text{現在の境界にポーズあり}) = 3/7$ 」、「 $P(\text{現在の境界に読点あり} | \text{現在の境界にポーズなし}) = 1/9$ 」、「 $P(\text{現在の境界に読点なし} | \text{現在の境界にポーズなし}) = 8/9$ 」と算出する。

【0077】

次に、属性情報確率情報算出部 1106 は、各ポーズ情報において、属性情報「並列句であるか否か」が示す確率についての情報である属性情報確率情報を算出する。つまり、属性情報確率情報算出部 1106 は、図 12 の表の情報から、(1) $P(\text{現在の境界は並列句である} | \text{現在の境界にポーズあり})$ 、(2) $P(\text{現在の境界は並列句でない} | \text{現在の境界にポーズあり})$ 、(3) $P(\text{現在の境界は並列句である} | \text{現在の境界にポーズなし})$ 、(4) $P(\text{現在の境界は並列句でない} | \text{現在の境界にポーズなし})$ を算出する。図 12 において、「ポーズ有り」(ポーズ情報が「1」)の境界(IDが(1)(4)(7)(8)(11)(13)(15))の中で、属性「並列句であるか否か」の属性値が「1」(つまり、「並列句である」)のものは、(13)のみの 1 つであるので、属性情報確率情報算出部 1106 は、「 $P(\text{現在の境界は並列句である} | \text{現在の境界にポーズあり}) = 1/7$ 」と算出する。同様に、属性情報確率情報算出部 1106 は、「 $P(\text{現在の境界は並列句でない} | \text{現在の境界にポーズあり}) = 6/7$ 」、「 $P(\text{現在の境界は並列句である} | \text{現在の境界にポーズなし}) = 2/9$ 」、「 $P(\text{現在の境界は並列句でない} | \text{現在の境界にポーズなし}) = 7/9$ 」と算出する。

【0078】

次に、ポーズ情報確率情報算出部 1107 は、ポーズ情報取得部 1105 が取得した各境界についてのポーズ情報に基づいて、隣り合う 2 つの境界についての 2 つのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を算出する。具体的には、ポーズ情報確率情報算出部 1107 は、図 12 の「ポーズ情報」の属性値を読み込み、現在の境界の「ポーズ情報」の属性値が「1」である「ポーズ有り」を(1)(4)(7)(8)(11)(13)(15)の 7 つであることを取得する。そして、ポーズ情報確率情報算出部 1107 は、7 つの境界のうち、直前の境界が「ポーズ有り」である数を「3」(「ID」が(1)(8)(13))と取得する。なお、文の先頭は、サイレンスであり、常にポーズと見なす、とする。そして、ポーズ情報確率情報算出部 1107 は、「 $P(\text{一つ前の境界にポーズあり} | \text{現在の境界にポーズあり}) = 3/7$ 」を得る。同様に、ポーズ情報確率情報算出部 1107 は、「 $P(\text{一つ前の境界にポーズあり} | \text{現在の境界にポーズあり}) = 4/7$ 」、「 $P(\text{一つ前の境界にポーズあり} | \text{現在の境界にポーズなし}) = 7/9$ 」、「 $P(\text{一つ前の境界にポーズなし} | \text{現在の境界にポーズなし}) = 2/9$ 」を得る。

【0079】

次に、最後境界ポーズ情報確率情報算出部 1108 は、最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出する。具体的には、最後境界ポーズ情報確率情報算出部 1108 は、「 $P(\text{文の最後の境界がポーズである}) = 1/4$ 」、「 $P(\text{文の最後の境界がポーズでない}) = 3/4$ 」を得る。つまり、最後境界ポーズ情報確率情報算出部 1108 は、4 つの文の最後の境界(「ID」が(2)(8)(12)(16))の属性値「ポーズ情報」を読み出し、「ID = 8」の境界のみポーズ情報が「1」であることを取得する。最後境界ポーズ情報確率情報算出部 1108 は、「ID = 2、12、16」の境界のポーズ情報が「0」であることを取得する。そして、最後境界ポーズ情報確率情報算出部 1108 は、「 $P(\text{文の最後の境界がポーズである}) = 1/4$ 」を得る。また、最後境界ポーズ情報確率情報算出部 1108 は、「 $P(\text{文の最後の境界がポーズでない}) = 3/4$ 」を得る。

そして、学習データ生成装置 11 は、学習データ記憶部 1101 に学習データを記録する。記録された学習データの例を図 13 に示す。

次に、音声合成装置 12 の具体的な動作について説明する。

まず、ユーザは、文「14日、中日と阪神の試合が行われた。」を入力する。

次に、文受付部 1203 は、文「14日、中日と阪神の試合が行われた。」を受け付ける。

10

20

30

40

50

【0080】

次に、境界情報取得部1204は、受け付けた文「14日、中日と阪神の試合が行われた。」を文節に分割し、当該文の境界を識別する情報である境界情報を1以上取得する。その結果、境界情報取得部1204は、「14日、|中日と|阪神の|試合が|行われた。」を得る。なお、ここで、「|」は、文節の境界を示す。文「14日、中日と阪神の試合が行われた。」における4つの境界に対して(1)から(4)のIDを付与する、とする(図14参照)。

【0081】

次に、第二属性情報取得部1205は、図14の4つの境界の読点の有無、並列句であるか否かを検出し、図15に示す属性情報管理表を得る。本属性情報管理表は、「ID」「読点の有無」「並列句であるか否か」の属性値を有するレコードを1以上保持している。なお、「読点の有無」「並列句であるか否か」の属性値を得る方法は、学習データ生成装置11と同様であり、従来技術(非特許文献2参照)により可能である。なお、「中日と」「阪神の」は、並列句である。

10

【0082】

次に、属性情報確率情報取得部1206は、属性情報確率情報を取得する。具体的には、「ID=1」の境界は、「読点有り」「並列句でない」であるので、属性情報確率情報取得部1206は、「ID=1」の境界に対して、ポーズ情報が「ポーズ有り」の場合は、「 $P(\text{現在の境界に読点あり}|\text{現在の境界にポーズあり}) = 4/7$ 」および「 $P(\text{現在の境界は並列句でない}|\text{現在の境界にポーズあり}) = 6/7$ 」を学習データ(図13参照)から検索する。そして、属性情報確率情報取得部1206は、「 $(4/7) \times (6/7)$ 」を得る。また、属性情報確率情報取得部1206は、「ID=1」の境界に対して、ポーズ情報が「ポーズ無し」の場合は、「 $P(\text{現在の境界に読点あり}|\text{現在の境界にポーズなし}) = 1/9$ 」および「 $P(\text{現在の境界は並列句でない}|\text{現在の境界にポーズなし}) = 7/9$ 」を学習データから検索する。そして、属性情報確率情報取得部1206は、「 $(1/9) \times (7/9)$ 」を得る。

20

【0083】

同様に、「ID=2」の境界は、「読点無し」「並列句である」であるので、属性情報確率情報取得部1206は、ポーズ情報が「ポーズ有り」の場合は、「 $P(\text{現在の境界に読点なし}|\text{現在の境界にポーズあり}) = 3/7$ 」および「 $P(\text{現在の境界は並列句である}|\text{現在の境界にポーズなし}) = 1/7$ 」を学習データ(図13参照)から検索する。そして、属性情報確率情報取得部1206は、「 $(3/7) \times (1/7)$ 」を得る。また、属性情報確率情報取得部1206は、「ID=2」の境界に対して、ポーズ情報が「ポーズ無し」の場合は、「 $P(\text{現在の境界に読点なし}|\text{現在の境界にポーズなし}) = 8/9$ 」および「 $P(\text{現在の境界は並列句でない}|\text{現在の境界にポーズなし}) = 2/9$ 」を学習データから検索する。そして、属性情報確率情報取得部1206は、「 $(8/9) \times (2/9)$ 」を得る。同様に、属性情報確率情報取得部1206は、「ID=3」の境界、「ID=4」の境界に対応する属性情報確率情報を算出する。そして、属性情報確率情報取得部1206は、図16に示す属性情報確率情報管理表を得る。

30

【0084】

次に、ポーズ情報確率情報取得部1207は、各境界についてのポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を学習データ格納部1202から読み出す。読み出したポーズ情報確率情報は、「 $P(\text{一つ前の境界にポーズあり}|\text{現在の境界にポーズあり}) = 3/7$ 」「 $P(\text{一つ前の境界にポーズなし}|\text{現在の境界にポーズあり}) = 4/7$ 」「 $P(\text{一つ前の境界にポーズあり}|\text{現在の境界にポーズなし}) = 7/9$ 」「 $P(\text{一つ前の境界にポーズなし}|\text{現在の境界にポーズなし}) = 2/9$ 」である。なお、ポーズ情報確率情報は、状態の遷移確率とも言える。

40

【0085】

次に、最後境界ポーズ情報確率情報取得部1208は、最後境界ポーズ情報確率情報を

50

学習データ格納部 1202 から読み出す。読み出した最後境界ポーズ情報確率情報は、「 $P(\text{文の最後の境界がポーズである}) = 1/4$ 」「 $P(\text{文の最後の境界がポーズでない}) = 3/4$ 」である。なお、最後境界ポーズ情報確率情報は、DPでバックトラックする場合の初期確率と言える。

【0086】

次に、ポーズ情報系列決定部 1209 は、属性情報確率情報取得部 1206 が取得した境界ごとの 1 以上の属性情報確率情報、ポーズ情報確率情報取得部 1207 が取得したポーズ情報確率情報、および最後境界ポーズ情報確率情報取得部 1208 が取得した最後境界ポーズ情報確率情報を用いて、確率が最大（積が最大）となるポーズ情報の系列を動的計画法により決定する。具体的には、図 17 に示すように、動的計画法（DP）を用いて、DP 累積スコアをもとに、文章の前からバックトラックする事で、最適な系列を決定する。ここでは、太線の矢印がポーズ情報の系列である。つまり、本発明にかかる音声合成装置において、「 o_0 （ポーズ有り）、 o_1 （ポーズ有り）、 o_2 （ポーズ無し）、 o_3 （ポーズ有り）、 o_4 （ポーズ無し）」であった。なお、上述した従来技術（ k -NN、決定木、Naive Bayes、SVM などを利用した技術）では、「 o_0 （ポーズ有り）、 o_1 （ポーズ有り）、 o_2 （ポーズ無し）、 o_3 （ポーズ無し）、 o_4 （ポーズ無し）」となる。つまり、本発明において、前の状態（前の境界がポーズ有りか無しか）を考慮することにより、より好適なポーズ挿入が行えた。

【0087】

そして、ポーズ情報系列決定部 1209 は、文節「14日、」と「中日と」の間、および文節「阪神の」と「試合が」の間にポーズを挿入することを得る。ここで、ポーズ情報の系列は、ポーズの位置を示す情報（例えば、「中日と」の「中」の先頭からのバイト数、および「試合が」の先頭からのバイト数など）だけでも良いし、図 17 に示す状態遷移図を表現するための情報でも良い。

そして、音声合成出力部 1210 は、ポーズ情報系列決定部 1209 が決定したポーズ情報に対応した無音区間を構成し、文受付部 1203 が受け付けた文を音声合成出力する。

【0088】

なお、以下、本音声処理システムが有するポーズ予測装置を用いて、ポーズの位置を予測した場合の評価について述べる。本ポーズ予測装置は、属性情報確率情報とポーズ情報確率情報と最後境界ポーズ情報確率情報を有する学習データを格納している学習データ格納部と、文を受け付ける文受付部と、前記文受付部が受け付けた文を分割し、当該文の境界を識別する情報である境界情報を 1 以上取得する境界情報取得部と、前記境界情報取得部が取得した 1 以上の境界情報が示す各境界の 1 種以上の属性情報を取得する第二属性情報取得部と、前記境界ごとに前記第二属性情報取得部が取得した 1 種以上の各属性情報に対応する 1 以上の属性情報確率情報を前記学習データ格納部から取得する属性情報確率情報取得部と、前記境界情報取得部が取得した 1 以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データ格納部から取得するポーズ情報確率情報取得部と、前記最後境界ポーズ情報確率情報を前記学習データ格納部から取得する最後境界ポーズ情報確率情報取得部と、前記属性情報確率情報取得部が取得した前記境界ごとの 1 以上の属性情報確率情報、前記ポーズ情報確率情報取得部が取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得部が取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定部と、前記ポーズ情報系列決定部が決定したポーズ情報を出力する出力部を具備する。

また、本ポーズ予測装置の評価において、単純に境界毎に与えられた説明変数に基づいて、ポーズの位置を予測する従来技術との比較を行った。

【0089】

本評価実験では文中の各境界におけるポーズの有無を一文単位で予測した。これらのポーズは文節境界に入ると仮定し、文節境界の属性情報（説明変数） x_i には、1) 当該境

10

20

30

40

50

界の係り受け距離（量的データ）、2）先行境界の係り受け距離（量的データ）、3）読点の有無（質的データ）、4）並列句であるか否か（質的データ）、5）同格句であるか否か（質的データ）、6）前の句のモーラ数（量的データ）、7）次の句のモーラ数（量的データ）を使用した。係り受け解析は、非特許文献2の従来技術を用いた。また、学習に用いた音声コーパスは男女2名（以下のF009（男性）、M007（女性））のナレータで構成され、各話者のそれぞれ51896文、28083文の音声データにラベリングしたものを入力の文として用いた。ポーズのラベリングは音声ファイルに対しHMMを用いて行った。学習に用いた音声コーパスに含まれるポーズの長さを平均モーラ長で規格化し、頻度分布として表したものを図20に示す。図20より、双峰分布となっていることが確認できる。

10

【0090】

比較対象の従来技術は、Naive Bayes（正規分布）、Naive Bayes（カーネル密度推定）、決定木の3種類である。本ポーズ予測装置の評価結果を図21に示す。図21は、両話者の全データに対し「10-fold cross validation（「交差分割法」とも言う。）」で評価を行った結果で、表中の「ポーズあり」はポーズの予測精度を示している。また、図21中の（a）-（d）はそれぞれ正規分布を利用したNaive Bayes、カーネル密度推定を利用したNaive Bayes、本ポーズ予測装置、決定木による識別結果の再現率・適合率・F値を示している。この結果より、カーネル密度推定の利用、および文全体の確率をマルコフ過程により考慮することはポーズ挿入の予測精度を高めることができる。

20

以上、本実施の形態によれば、発話全体で最適なポーズの位置や種類が決定でき、そのためにより自然な音声合成出力が可能となる。

【0091】

また、本実施の形態の学習データ生成装置によれば、受け付けた文に含まれる境界について、ポーズの有無別に、属性情報を取得し、ポーズの種類別および属性情報別に数を数える処理が、学習データを生成するための主たる処理であるので、非常に少ない計算量で、高速に学習データを取得することができる。

また、本実施の形態の音声合成装置によれば、少ない計算量でポーズ予測が可能となるため、例えば、リアルタイムの音声合成処理の前処理として極めて有効である。

なお、本実施の形態によれば、音声合成装置が有する上述のポーズ予測装置は、音声合成以外の処理にも適用可能である。音声合成以外の処理とは、例えば、入力した文の発音練習や、発音評価などである。

30

また、本実施の形態によれば、主として、7つの属性情報を説明したが、7つ以外の属性情報を用いても良い。

【0092】

さらに、本実施の形態における処理は、ソフトウェアで実現しても良い。そして、このソフトウェアをソフトウェアダウンロード等により配布しても良い。また、このソフトウェアをCD-ROMなどの記録媒体に記録して流布しても良い。なお、このことは、本明細書における他の実施の形態においても該当する。なお、本実施の形態における学習データ生成装置を実現するソフトウェアは、以下のようなプログラムである。つまり、このプログラムは、コンピュータに、1以上の文の音声を受け付ける音声受付ステップと、前記音声受付ステップで受け付けた音声の境界を検出し、当該境界を識別する情報である境界情報を、文ごとに取得する境界決定ステップと、前記境界決定ステップで取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第一属性情報取得ステップと、前記音声受付ステップで受け付けた音声から無音区間を検知し、前記無音区間の長さに基づいて、前記境界のポーズの種類についての情報であるポーズ情報を境界ごとに取得するポーズ情報取得ステップと、前記ポーズ情報において、前記1種以上の属性情報が示す確率についての情報である属性情報確率情報を算出し、記憶部に記録する属性情報確率情報算出ステップと、前記ポーズ情報取得ステップで取得した各境界についてのポーズ情報に基づいて、隣り合う2つの境界についての2つのポーズ情報が連続する確率についての情

40

50

報であるポーズ情報確率情報を算出し、記憶部に記録するポーズ情報確率情報算出ステップと、最後の境界における各ポーズ情報の確率についての情報である最後境界ポーズ情報確率情報を算出し、記憶部に記録する最後境界ポーズ情報確率情報算出ステップを実行させるためのプログラム、である。

【 0 0 9 3 】

また、音声合成装置を実現するプログラムは、コンピュータに、文を受け付ける文受付ステップと、前記文受付ステップで受け付けた文を分割し、当該文の境界を識別する情報である境界情報を1以上取得する境界情報取得ステップと、前記境界情報取得ステップで取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得ステップと、前記境界ごとに前記第二属性情報取得ステップで取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を、格納している学習データから取得する属性情報確率情報取得ステップと、前記境界情報取得ステップで取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を、前記学習データから取得するポーズ情報確率情報取得ステップと、前記最後境界ポーズ情報確率情報を前記学習データから取得する最後境界ポーズ情報確率情報取得ステップと、前記属性情報確率情報取得ステップで取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得ステップで取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得ステップで取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定ステップと、前記ポーズ情報系列決定ステップで決定したポーズ情報に対応した無音区間を構成し、前記文受付ステップが受け付けた文を音声合成出力する音声合成出力ステップを実行させるためのプログラム、である。

【 0 0 9 4 】

また、ポーズ予測装置を実現するプログラムは、コンピュータに、文を受け付ける文受付ステップと、前記文受付ステップで受け付けた文を分割し、当該文の境界を識別する情報である境界情報を1以上取得する境界情報取得ステップと、前記境界情報取得ステップで取得した1以上の境界情報が示す各境界の1種以上の属性情報を取得する第二属性情報取得ステップと、前記境界ごとに前記第二属性情報取得ステップで取得した1種以上の各属性情報に対応する1以上の属性情報確率情報を、格納している学習データから取得する属性情報確率情報取得ステップと、前記境界情報取得ステップで取得した1以上の境界情報が示す各境界について、ポーズ情報と当該各境界の前の境界についてのポーズ情報が連続する確率についての情報であるポーズ情報確率情報を前記学習データから取得するポーズ情報確率情報取得ステップと、前記最後境界ポーズ情報確率情報を前記学習データから取得する最後境界ポーズ情報確率情報取得ステップと、前記属性情報確率情報取得ステップで取得した前記境界ごとの1以上の属性情報確率情報、前記ポーズ情報確率情報取得ステップで取得したポーズ情報確率情報、および前記最後境界ポーズ情報確率情報取得ステップで取得した最後境界ポーズ情報確率情報を用いて、確率が最大となるポーズ情報の系列を動的計画法により決定するポーズ情報系列決定ステップと、前記ポーズ情報系列決定ステップで決定したポーズ情報を出力する出力ステップを実行させるためのプログラム、である。なお、出力ステップにおける出力とは、ディスプレイへの表示、プリンタへの印字、音出力、外部の装置への送信、記録媒体への蓄積等を含む概念である。

【 0 0 9 5 】

また、上記各実施の形態において、各処理（各機能）は、単一の装置（システム）によって集中処理されることによって実現されてもよく、あるいは、複数の装置によって分散処理されることによって実現されてもよい。

【 0 0 9 6 】

また、図22は、本明細書で述べたプログラムを実行して、上述した種々の実施の形態の情報処理装置を実現するコンピュータの外観を示す。上述の実施の形態は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムで実現され得る。図22

10

20

30

40

50

は、このコンピュータシステム160の概観図であり、図23は、システム160のブロック図である。

【0097】

図11において、コンピュータシステム160は、FD(Flexible Disk)ドライブ、CD-ROM(Compact Disk Read Only Memory)ドライブを含むコンピュータ161と、キーボード162と、マウス163と、モニタ164と、マイク165と、スピーカー166とを含む。

【0098】

図12において、コンピュータ161は、FDドライブ1611、CD-ROMドライブ1612に加えて、CPU(Central Processing Unit)1613と、CPU1613、CD-ROMドライブ1612及びFDドライブ1611に接続されたバス1614と、ブートアッププログラム等のプログラムを記憶するためのROM(Read-Only Memory)1615と、CPU1613に接続され、アプリケーションプログラムの命令を一時的に記憶するとともに一時記憶空間を提供するためのRAM(Random Access Memory)1616と、アプリケーションプログラム、システムプログラム、及びデータを記憶するためのハードディスク1617とを含む。ここでは、図示しないが、コンピュータ161は、さらに、LANへの接続を提供するネットワークカードを含んでも良い。

【0099】

コンピュータシステム160に、上述した実施の形態の情報処理装置の機能を実行させるプログラムは、CD-ROM1701、またはFD1702に記憶されて、CD-ROMドライブ1612またはFDドライブ1611に挿入され、さらにハードディスク1617に転送されても良い。これに代えて、プログラムは、図示しないネットワークを介してコンピュータ161に送信され、ハードディスク1617に記憶されても良い。プログラムは実行の際にRAM1616にロードされる。プログラムは、CD-ROM1701、FD1702またはネットワークから直接、ロードされても良い。

【0100】

プログラムは、コンピュータ161に、上述した実施の形態の情報処理装置の機能を実行させるオペレーティングシステム(OS)、またはサードパーティープログラム等は、必ずしも含まなくても良い。プログラムは、制御された態様で適切な機能(モジュール)を呼び出し、所望の結果が得られるようにする命令の部分のみを含んでいれば良い。コンピュータシステム160がどのように動作するかは周知であり、詳細な説明は省略する。

また、上記プログラムを実行するコンピュータは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、あるいは分散処理を行ってもよい。

本発明は、以上の実施の形態に限定されることなく、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

【産業上の利用可能性】

【0101】

以上のように、本発明にかかる音声処理システムは、自然な音声合成出力が可能となる、という効果を有し、音声合成システム等として有用である。

【図面の簡単な説明】

【0102】

【図1】実施の形態1における音声処理システムのブロック図

【図2】同学習データ生成装置の動作について説明するフローチャート

【図3】同属性情報確率情報算出処理について説明するフローチャート

【図4】同ポーズ情報確率情報算出処理について説明するフローチャート

【図5】同最後境界ポーズ情報確率情報算出処理について説明するフローチャート

【図6】同音声合成装置の動作について説明するフローチャート

【図7】同属性情報確率情報取得処理について説明するフローチャート

【図8】同受け付ける文を示す図

10

20

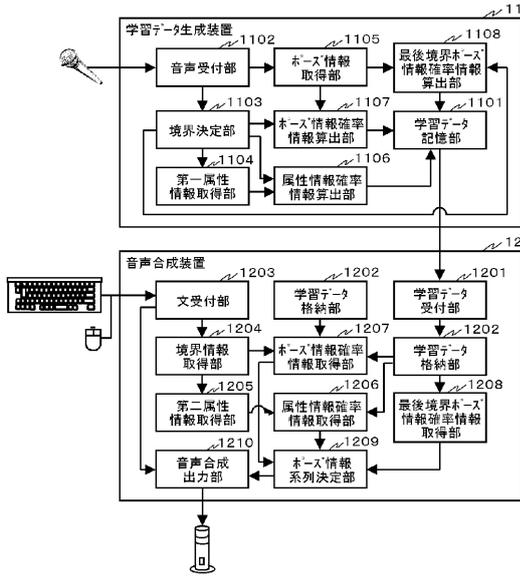
30

40

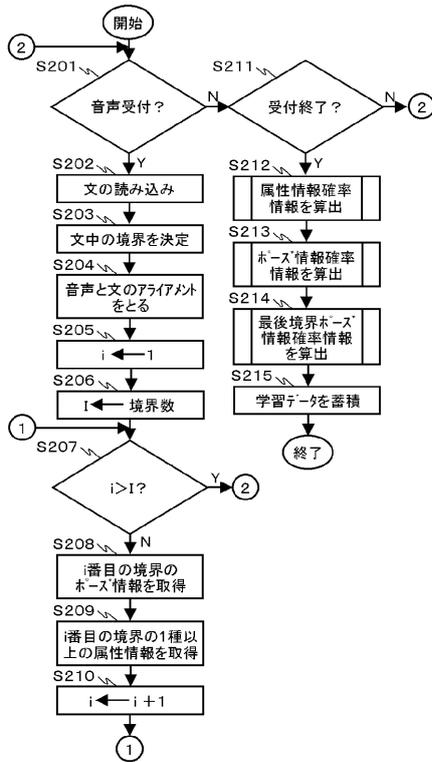
50

- 【図 9】同文を境界に区切った場合の例を示す図
- 【図 10】同ポーズ情報取得部が取得したポーズの位置を説明する図
- 【図 11】同第一境界情報管理表を示す図
- 【図 12】同第二境界情報管理表を示す図
- 【図 13】同記録された学習データの例を示す図
- 【図 14】同入力された文と境界に区分された例を示す図
- 【図 15】同属性情報管理表を示す図
- 【図 16】同属性情報確率情報管理表を示す図
- 【図 17】同動的計画法(DP)を用いて最適な系列を決定する処理を説明する図
- 【図 18】同一次マルコフ過程のモデルを示す図 10
- 【図 19】同本音声処理システムで用いるモデルを説明する図
- 【図 20】同ポーズ長の頻度分布を示す図
- 【図 21】同本ポーズ予測装置の処理の評価結果を示す図
- 【図 22】同音声装置システムを構成する装置を示す図
- 【図 23】同装置のブロック図
- 【符号の説明】
- 【0103】
 - 11 学習データ生成装置
 - 12 音声合成装置
 - 1101 学習データ記憶部 20
 - 1102 音声受付部
 - 1103 境界決定部
 - 1104 第一属性情報取得部
 - 1105 ポーズ情報取得部
 - 1106 属性情報確率情報算出部
 - 1107 ポーズ情報確率情報算出部
 - 1108 最後境界ポーズ情報確率情報算出部
 - 1201 学習データ受付部
 - 1202 学習データ格納部
 - 1203 文受付部 30
 - 1204 境界情報取得部
 - 1205 第二属性情報取得部
 - 1206 属性情報確率情報取得部
 - 1207 ポーズ情報確率情報取得部
 - 1208 最後境界ポーズ情報確率情報取得部
 - 1209 ポーズ情報系列決定部
 - 1210 音声合成出力部

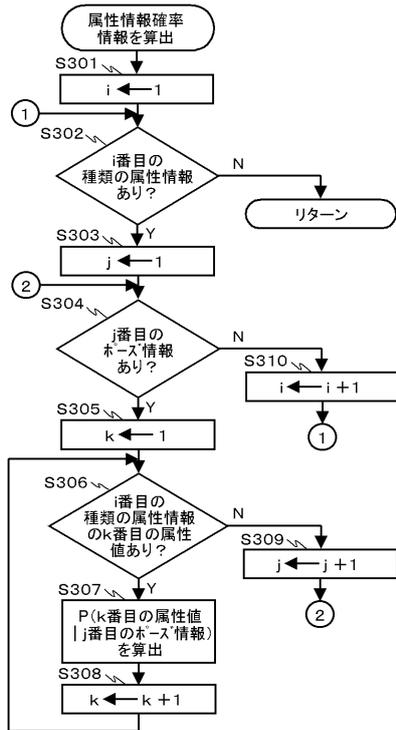
【図1】



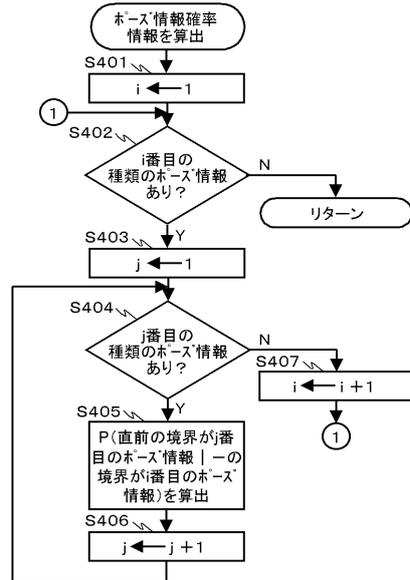
【図2】



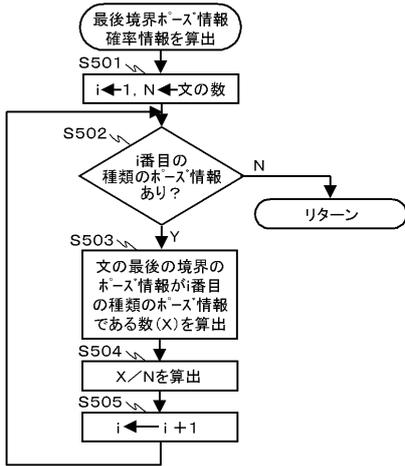
【図3】



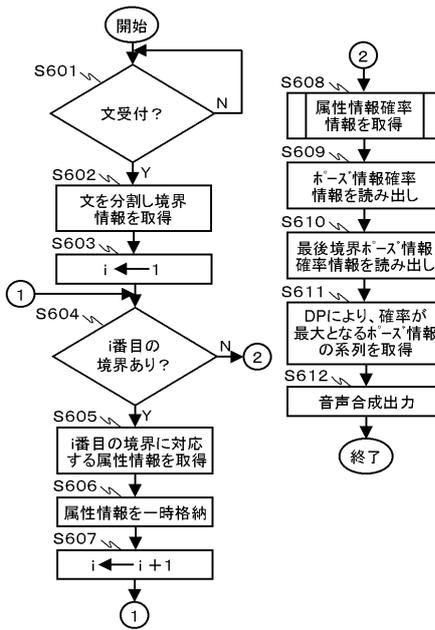
【図4】



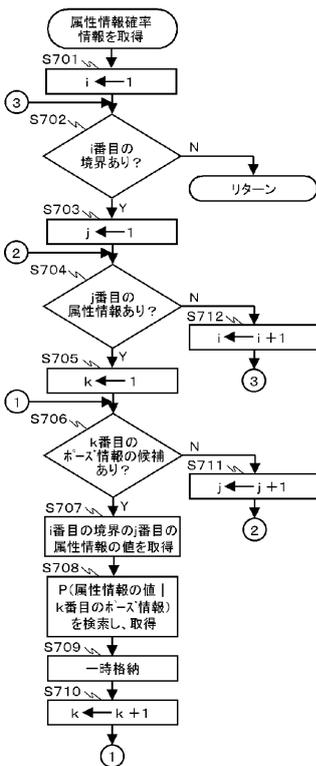
【図5】



【図6】



【図7】



【図8】

(A) 今日は、よい天気です。
 (B) 明日は、全国で晴れるところが多く、お出かけ日和となるでしょう。
 (C) 京都と奈良と大阪は、関西の観光名所です。
 (D) A社とB社は隣だが、歩くと遠い。

【図9】

(1) (2)
 (A) 今日は、 | よい | 天気です。
 (3) (4) (5) (6)
 (B) 明日は、 | 全国で | 晴れる | ところが | 多く、
 (7) (8)
 | お出かけ日和と | なるでしょう。
 (9) (10) (11) (12)
 (C) 京都と | 奈良と | 大阪は、 | 関西の | 観光名所です。
 (13) (14) (15) (16)
 (D) A社と | B社は | 隣だが、 | 歩くと | 遠い。

【図 10】

(A)今日は、⁽¹⁾ | ⁽²⁾ | よい | 天気です。
○

(B)明日は、⁽³⁾ | 全国で | ⁽⁴⁾ | 晴れる | ⁽⁵⁾ | ところが | ⁽⁶⁾ | 多く、
○

⁽⁷⁾ | お出かけ日和と | ⁽⁸⁾ | なるでしょう。
○

(C)京都と | ⁽⁹⁾ | 奈良と | ⁽¹⁰⁾ | 大阪は、 | ⁽¹¹⁾ | 関西の | ⁽¹²⁾ | 観光名所です。
○

(D)A社と | ⁽¹³⁾ | B社は | ⁽¹⁴⁾ | 隣だが、 | ⁽¹⁵⁾ | 歩くと | ⁽¹⁶⁾ | 遠い。
○

【図 11】

ID	ポーズ情報
1	1
2	0
3	0
4	1
5	0
6	0
7	1
8	1
9	0
10	0
11	1
12	0
13	1
14	0
15	1
16	0

【図 12】

ID	ポーズ情報	読点の有無	並列句であるか否か
1	1	1	0
2	0	0	0
3	0	1	0
4	1	0	0
5	0	0	0
6	0	0	0
7	1	1	0
8	1	0	0
9	0	0	1
10	0	0	1
11	1	1	0
12	0	0	0
13	1	0	1
14	0	0	0
15	1	1	0
16	0	0	0

【図 13】

$P(x_i | o_i)$: 「読点の有無」の属性情報確率情報
 $P(\text{現在の境界に読点あり} | \text{現在の境界にポーズあり})=4/7$
 $P(\text{現在の境界に読点なし} | \text{現在の境界にポーズあり})=3/7$
 $P(\text{現在の境界に読点あり} | \text{現在の境界にポーズなし})=1/9$
 $P(\text{現在の境界に読点なし} | \text{現在の境界にポーズなし})=8/9$

$P(x_i | o_i)$: 「並列句であるか否か」の属性情報確率情報
 $P(\text{現在の境界は並列句である} | \text{現在の境界にポーズあり})=1/7$
 $P(\text{現在の境界は並列句でない} | \text{現在の境界にポーズあり})=6/7$
 $P(\text{現在の境界は並列句である} | \text{現在の境界にポーズなし})=2/9$
 $P(\text{現在の境界は並列句でない} | \text{現在の境界にポーズなし})=7/9$

$P(o_{i-1} | o_i)$: ポーズ情報確率情報
 $P(\text{一つ前の境界にポーズあり} | \text{現在の境界にポーズあり})=3/7$
 $P(\text{一つ前の境界にポーズなし} | \text{現在の境界にポーズあり})=4/7$
 $P(\text{一つ前の境界にポーズあり} | \text{現在の境界にポーズなし})=7/9$
 $P(\text{一つ前の境界にポーズなし} | \text{現在の境界にポーズなし})=2/9$

$P(o_i)$: 最後境界ポーズ情報確率情報
 $P(\text{文の最後の境界がポーズである})=1/4$
 $P(\text{文の最後の境界がポーズでない})=3/4$

【図 14】

(1) | (2) | (3) | (4)
 14日、 | 中日と | 阪神の | 試合が | 行われた。

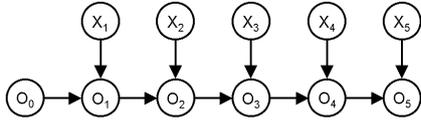
【図 15】

ID	読点の有無	並列句であるか否か
1	1	0
2	0	1
3	0	0
4	0	0

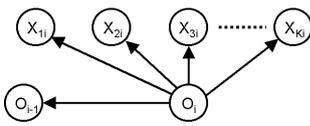
【図16】

ID	読点の有無	並列句であるか否か	属性情報確率情報	
			ホース有り	ホース無し
1	1	0	$\frac{4}{7} \times \frac{6}{7}$	$\frac{1}{9} \times \frac{7}{9}$
2	0	1	$\frac{3}{7} \times \frac{1}{7}$	$\frac{8}{9} \times \frac{2}{9}$
3	0	0	$\frac{3}{7} \times \frac{6}{7}$	$\frac{8}{9} \times \frac{7}{9}$
4	0	0	$\frac{3}{7} \times \frac{6}{7}$	$\frac{8}{9} \times \frac{7}{9}$

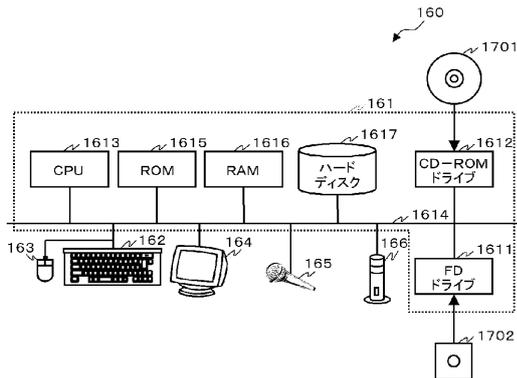
【図18】



【図19】



【図23】



【図21】

話者	種別	再現率	適合率	F値
F009	ホース有り	0.59	0.782	0.673
M007	ホース有り	0.528	0.733	0.613

(a)

話者	種別	再現率	適合率	F値
F009	ホース有り	0.652	0.834	0.732
M007	ホース有り	0.6	0.787	0.681

(b)

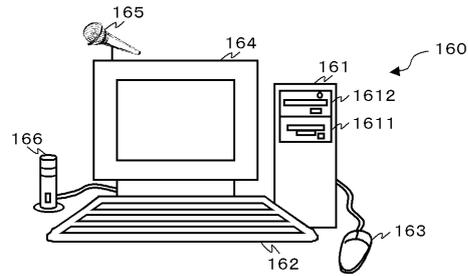
話者	種別	再現率	適合率	F値
F009	ホース有り	0.784	0.733	0.757
M007	ホース有り	0.749	0.729	0.735

(c)

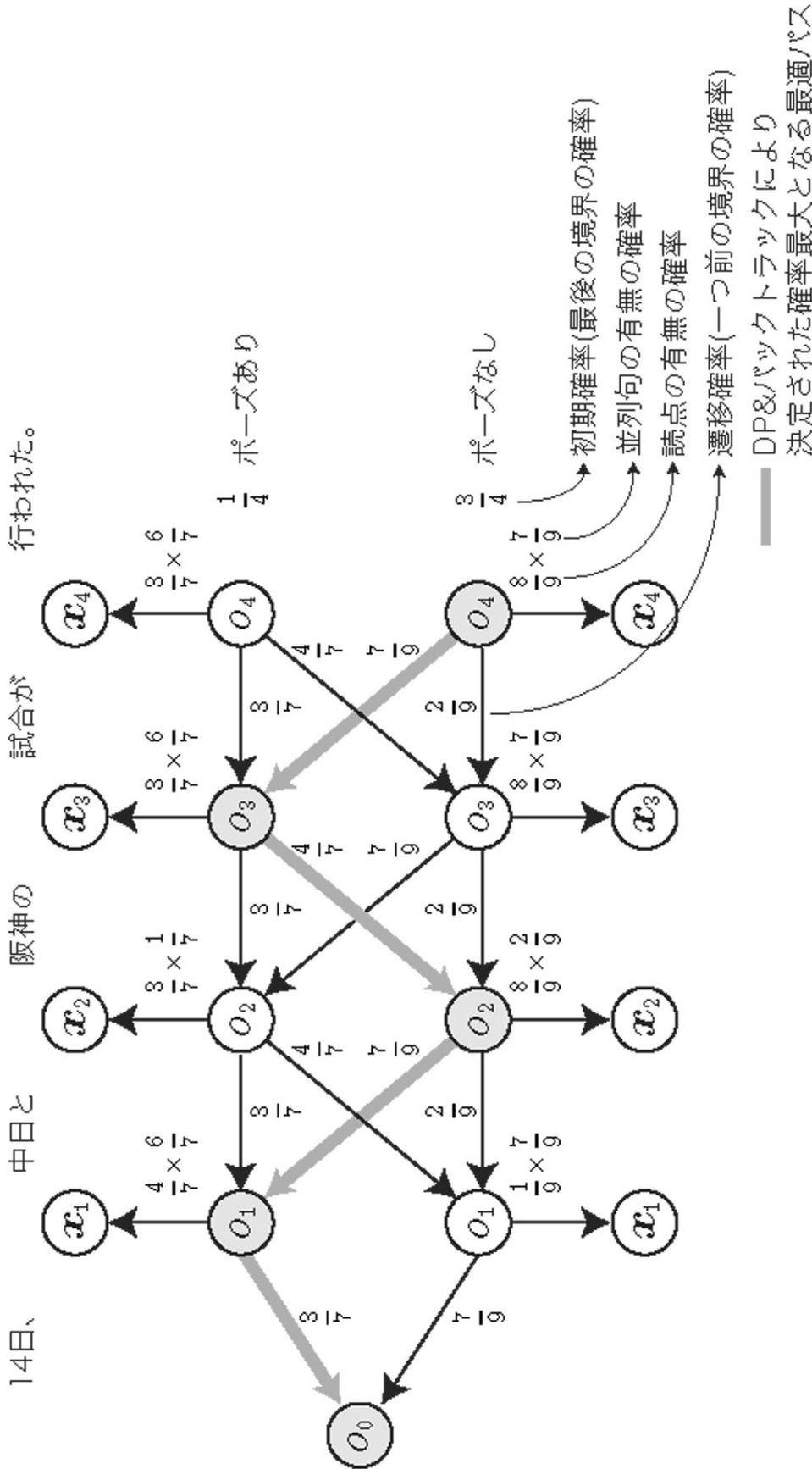
話者	種別	再現率	適合率	F値
F009	ホース有り	0.704	0.823	0.759
M007	ホース有り	0.693	0.779	0.733

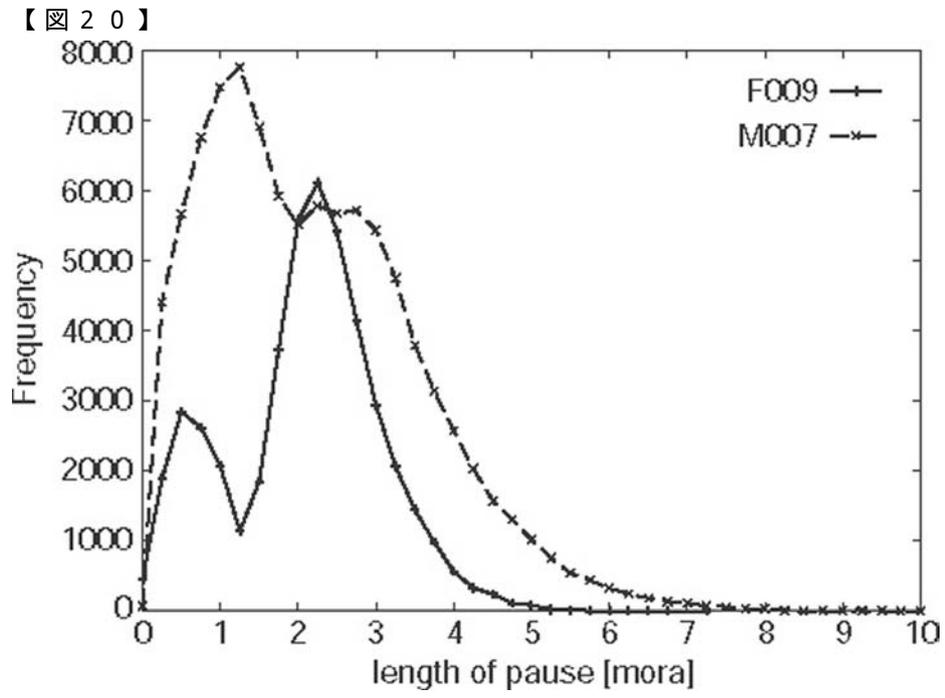
(d)

【図22】



【 図 17 】





フロントページの続き

審査官 山下 剛史

- (56)参考文献 特開2005-031259(JP,A)
特開2003-084787(JP,A)
特開2003-015680(JP,A)
特開2002-311982(JP,A)
特開2001-075584(JP,A)
特開2002-287785(JP,A)
藤尾茂他, "確率文脈自由文法を用いた韻律句境界とポーズ位置の予測", 電子情報通信学会論文誌, Vol.J80-D-II, No.1(1997-01), pp.18-25
海老原充他, "音声合成におけるネットワークモデルによるポーズ位置予測", 電子情報通信学会技術研究報告, Vol.96, No.566, SP96-133(1997-03), pp.45-50

- (58)調査した分野(Int.Cl., DB名)
G10L 13/00 - 15/28