

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4585759号
(P4585759)

(45) 発行日 平成22年11月24日(2010.11.24)

(24) 登録日 平成22年9月10日(2010.9.10)

(51) Int.Cl. F I
G 1 O L 15/22 (2006.01)
 G 1 O L 15/22 3 O O U
 G 1 O L 15/22 4 6 O D

請求項の数 4 (全 15 頁)

(21) 出願番号	特願2003-403364 (P2003-403364)	(73) 特許権者	000001007
(22) 出願日	平成15年12月2日(2003.12.2)		キヤノン株式会社
(65) 公開番号	特開2005-164944 (P2005-164944A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成17年6月23日(2005.6.23)	(74) 代理人	100125254
審査請求日	平成18年12月4日(2006.12.4)		弁理士 別役 重尚
		(72) 発明者	中川 賢一郎
			東京都大田区下丸子3丁目30番2号 キ
			ヤノン株式会社内
		(72) 発明者	廣田 誠
			東京都大田区下丸子3丁目30番2号 キ
			ヤノン株式会社内
		(72) 発明者	山本 寛樹
			東京都大田区下丸子3丁目30番2号 キ
			ヤノン株式会社内

最終頁に続く

(54) 【発明の名称】 音声合成装置、音声合成方法、プログラム、及び記録媒体

(57) 【特許請求の範囲】

【請求項1】

音声出力する際の音声の種類を決定する動作パラメータとテキスト情報とを用いて、システムアナウンスとして当該テキスト情報の内容を示す合成音声出力する音声合成装置であって、

前記システムアナウンスがユーザの音声入力に対してどのような状況にあるかを示す対話状態と前記テキスト情報とを対応付けたセットを1セット以上有するコンテンツから、注目するテキスト情報に対応する対話状態を判断する判断手段と、

前記判断手段により、前記注目するテキスト情報に対応する前記対話状態がないと判断された場合、予め設定された第1の動作パラメータを検索し、前記判断手段により、前記注目するテキスト情報に対応する対話状態が音声を入力する直前のシステムアナウンスを出力している状態にあることを示す第2の対話状態と判断された場合、対話状態と動作パラメータとを対応付けて保持するテーブルから、前記第2の対話状態に対応する動作パラメータとして前記第1の動作パラメータと異なる第2の動作パラメータを検索し、前記判断手段により、前記注目するテキスト情報に対応する対話状態が前記第2の対話状態と判断され、かつ前記注目するテキスト情報に対応する対話状態がバージン可能なシステムアナウンスを出力している状態にあることを示す第3の対話状態と判断された場合、前記テーブルから、前記第2の対話状態に対応する前記第2の動作パラメータと前記第3の対話状態に対応する動作パラメータとして前記第1の動作パラメータ及び前記第2の動作パラメータと異なる第3の動作パラメータとを検索する検索手段と、

10

20

前記検索手段により、前記第 1 の動作パラメータが検索された場合、前記注目するテキスト情報と前記第 1 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声として第 1 の合成音声を生成し、前記検索手段により、前記第 2 の動作パラメータのみが検索された場合、前記注目するテキスト情報と前記第 2 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声であって前記第 1 の合成音声とは識別可能な第 2 の合成音声を生成し、前記検索手段により、前記第 2 の動作パラメータと前記第 3 の動作パラメータとが検索された場合、前記注目するテキスト情報と前記第 2 及び第 3 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声であって前記第 1 及び第 2 の合成音声と識別可能な第 3 の合成音声を生成する音声合成手段とを有する音声合成装置。

10

【請求項 2】

判断手段、検索手段、及び音声合成手段を有し、音声出力する際の音声の種類を決定する動作パラメータとテキスト情報とを用いて、システムアナウンスとして当該テキスト情報の内容を示す合成音声を出力する音声合成装置の音声合成方法であって、

前記システムアナウンスがユーザの音声入力に対してどのような状況にあるかを示す対話状態と前記テキスト情報とを対応付けたセットを 1 セット以上有するコンテンツから、前記判断手段が、注目するテキスト情報に対応する対話状態を判断する判断工程と、

前記検索手段が、前記判断工程により、前記注目するテキスト情報に対応する前記対話状態がないと判断された場合、予め設定された第 1 の動作パラメータを検索し、前記判断工程により、前記注目するテキスト情報に対応する対話状態が音声を入力する直前のシステムアナウンスを出力している状態にあることを示す第 2 の対話状態と判断された場合、対話状態と動作パラメータとを対応付けて保持するテーブルから、前記第 2 の対話状態に対応する動作パラメータとして前記第 1 の動作パラメータと異なる第 2 の動作パラメータを検索し、前記判断工程により、前記注目するテキスト情報に対応する対話状態が前記第 2 の対話状態と判断され、かつ前記注目するテキスト情報に対応する対話状態がバージン可能なシステムアナウンスを出力している状態にあることを示す第 3 の対話状態と判断された場合、前記テーブルから、前記第 2 の対話状態に対応する前記第 2 の動作パラメータと前記第 3 の対話状態に対応する動作パラメータとして前記第 1 の動作パラメータ及び前記第 2 の動作パラメータと異なる第 3 の動作パラメータとを検索する検索工程と、

20

前記音声合成手段が、前記検索工程により前記第 1 の動作パラメータが検索された場合、前記注目するテキスト情報と前記第 1 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声として第 1 の合成音声を生成し、前記検索工程により前記第 2 の動作パラメータのみが検索された場合、前記注目するテキスト情報と前記第 2 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声であって前記第 1 の合成音声とは識別可能な第 2 の合成音声を生成し、前記検索工程により前記第 2 の動作パラメータと前記第 3 の動作パラメータとが検索された場合、前記注目するテキスト情報と前記第 2 及び第 3 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声であって前記第 1 及び第 2 の合成音声と識別可能な第 3 の合成音声を生成する音声合成工程とを有する音声合成装置の音声合成方法。

30

【請求項 3】

請求項 2 に記載の音声合成方法をコンピュータに実行させるプログラム。

40

【請求項 4】

請求項 3 に記載のプログラムが記録されたことを特徴とするコンピュータが読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、カーナビゲーションシステムや自動チケット予約システム等に適用される音声合成装置及び音声合成方法、並びに前記音声合成方法を実現するためのプログラム、前記プログラムを記録した記録媒体に関する。

50

【背景技術】

【0002】

音声を用いた対話は、人間にとって使いなれたものである。このため、音声対話を用いた機器の操作は、子供からお年寄りまで機器に不慣れなユーザにも自然な操作を提供することが可能となる。また、音声対話が必要とするデバイスはマイクとスピーカだけであり、これはマウスやキーボードと比較して機器の小型化に貢献するメリットもある。

【0003】

近年、音声を用いた対話システムは、カーナビゲーションシステムや自動コールセンター等で実現され始めている。また、VoiceXML(<http://www.w3.org/TR/voicexml20/>)やMultimodal Interaction(<http://www.w3.org/TR/mmi-framework/>)のような音声を用いた対話の標準仕様も策定中であり、この標準化が進むにつれ、多くの音声対話製品が世の中に出回ると考えられる。

【0004】

この種の対話システムでは、システムが完全に対話の主導権を握る“システム主導型”で実装されることが多い。これは、システムが質問し、ユーザが答えることで、タスクを遂行する形式である。例えば、次のような経路案内システムの例が挙げられる。(Sはシステムの出力、Uはユーザの入力とする)。

S:「こちらは経路案内システムです。」

S:「出発駅を発声してください。」

U:「東京」

S:「到着駅を発声してください。」

U:「大阪」

S:「東京から大阪でよろしいですか？」

U:「はい」

:

このようなシステム主導型の対話システムにおいては、ユーザが適切な場所で適切な入力を行うように、対話を組み上げていくことが大切である。しかし、どこで、どのような入力を行うかをユーザに適切に通知することは意外に難しく、結果として次のような入力ミスが生じることが多い。

(1) システムアナウンスがまだ続くと勘違いすることにより、入力を躊躇してしまう。

(2) 入力可能位置と勘違いすることにより、システムアナウンス中に無駄な入力をしてしまう。

(3) 入力を急に促されることにより、入力内容を頭の中で整理しきれない、あるいは喉の調子を整えられていないため、「えー」「あー」等の不要語の入力や、咳払いをしてしまう。

【0005】

これらの入力ミスを回避するため、ユーザが入力可能となる前にガイド音(ピー音)を付けることが行われている(第1の従来手法)。その一例を次に示す。

S:「こちらは経路案内システムです。」

S:「発信音の後に出発駅を発声してください。」(ピー)

U:「東京」

:

また、従来技術には、ユーザの対話モードに合わせて音声合成パラメータを変更するようにしたものがある(例えば特許文献1を参照:第2の従来手法)。これにより、ユーザはシステムアナウンスを聞くだけで、どのような対話状況であるかが判るようになる。

【特許文献1】特開2002-123385号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

10

20

30

40

50

しかしながら、上記第 1 及び第 2 の従来手法では、上記の問題 (1) と (2) を回避することができるものの、(3) の問題は回避することができなかった。さらに、現在どのような種類の入力 (音声、プッシュボタン等) が行えるかをユーザに伝えることはできなかった。

【 0 0 0 7 】

本発明は上記従来の問題点に鑑み、ユーザとの対話状態に応じた動作パラメータを利用して、音声入力するタイミングや種類をユーザに的確に知らせることができる音声合成装置、音声合成方法、プログラム、及び記録媒体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】

【 0 0 0 9 】

上記目的を達成するために、本発明の音声合成装置は、音声出力する際の音声の種類を決定する動作パラメータとテキスト情報とを用いて、システムアナウンスとして当該テキスト情報の内容を示す合成音声を出力する音声合成装置であって、前記システムアナウンスがユーザの音声入力に対してどのような状況にあるかを示す対話状態と前記テキスト情報とを対応付けたセットを 1 セット以上有するコンテンツから、注目するテキスト情報に対応する対話状態を判断する判断手段と、前記判断手段により、前記注目するテキスト情報に対応する前記対話状態がないと判断された場合、予め設定された第 1 の動作パラメータを検索し、前記判断手段により、前記注目するテキスト情報に対応する対話状態が音声を入力する直前のシステムアナウンスを出力している状態にあることを示す第 2 の対話状態と判断された場合、対話状態と動作パラメータとを対応付けて保持するテーブルから、前記第 2 の対話状態に対応する動作パラメータとして前記第 1 の動作パラメータと異なる第 2 の動作パラメータを検索し、前記判断手段により、前記注目するテキスト情報に対応する対話状態が前記第 2 の対話状態と判断され、かつ前記注目するテキスト情報に対応する対話状態がバージン可能なシステムアナウンスを出力している状態にあることを示す第 3 の対話状態と判断された場合、前記テーブルから、前記第 2 の対話状態に対応する前記第 2 の動作パラメータと前記第 3 の対話状態に対応する動作パラメータとして前記第 1 の動作パラメータ及び前記第 2 の動作パラメータと異なる第 3 の動作パラメータとを検索する検索手段と、前記検索手段により、前記第 1 の動作パラメータが検索された場合、前記注目するテキスト情報と前記第 1 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声として第 1 の合成音声を生成し、前記検索手段により、前記第 2 の動作パラメータのみが検索された場合、前記注目するテキスト情報と前記第 2 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声であって前記第 1 の合成音声とは識別可能な第 2 の合成音声を生成し、前記検索手段により、前記第 2 の動作パラメータと前記第 3 の動作パラメータとが検索された場合、前記注目するテキスト情報と前記第 2 及び第 3 の動作パラメータとを用いて、前記注目するテキスト情報の内容を示す合成音声であって前記第 1 及び第 2 の合成音声と識別可能な第 3 の合成音声を生成する音声合成手段とを有する。

【発明の効果】

【 0 0 1 7 】

本発明によれば、外部との対話状態に応じた動作パラメータを出力することが可能になる。

【 0 0 1 8 】

また、外部との対話状態に応じて決定された動作パラメータを利用して、入力するタイミングや種類をユーザに的確に知らせることが可能になる。

【発明を実施するための最良の形態】

【 0 0 1 9 】

まず、例えばカーナビゲーションシステムや自動チケット予約システムに適用される本実施形態の動作パラメータ決定装置について説明する。

【 0 0 2 0 】

[実施の形態]

10

20

30

40

50

図１は、本発明の実施の形態に係る動作パラメータ決定装置の機能構成図である。

【００２１】

この動作パラメータ決定装置１０１は、問い合わせシグナルが入力された瞬間に検知した対話状態によって、動作パラメータを生成し出力する機能を有し、装置外部に、ユーザとの対話を管理する対話管理装置１００と、装置１０１から出力された動作パラメータを受信する動作パラメータ受信部１０３と、動作パラメータを問い合わせる場合に問い合わせシグナルを装置１０１内に入力する問い合わせシグナル入力部１０４とが接続されている。対話管理装置１００は、現在の対話状態を検出する対話状態検出部１０２を有している。

【００２２】

動作パラメータ決定装置１０１の内部には、問い合わせシグナル受付部１１０が設けられている。問い合わせシグナル受付部１１０は、装置外部から入力される問い合わせシグナルを監視し、問い合わせシグナル入力部１０４から問い合わせシグナルが入力された場合に、その後の処理を開始する。問い合わせシグナルは、押しボタン等から送られるボタンイベントでもよいし、前もって決まったメモリ領域に、特定のメモリイメージがセットされることでもよい。

【００２３】

問い合わせシグナルが入力されると、問い合わせシグナル受付部１１０は、対話状態取り込み部１０７と動作パラメータ統合部１０９にその旨を通知する。対話状態取り込み部１０７は、装置外部の対話状態検出部１０２を動作させ、これによって現在の対話状態を検知するようになっている。

【００２４】

取り込まれた対話状態は、動作パラメータ検索部１０６へ送られる。動作パラメータ検索部１０６は、対話状態と動作パラメータとが対となって格納された対話状態／動作パラメータ対応テーブル１０５（例えば後述の図５参照）にアクセスし、対話状態に対応する動作パラメータを検索する。

【００２５】

検索された全動作パラメータは、動作パラメータ統合部１０９へ送られる。この動作パラメータ統合部１０９では、選択された動作パラメータが複数の場合に、これらが互いに矛盾したパラメータとならないように統合処理を施す。そして、動作パラメータ群は、動作パラメータ出力部１０８へ送られ、ここから装置外部の動作パラメータ受信部１０３に動作パラメータが出力される。

【００２６】

図２は、図１に示した動作パラメータ決定装置の詳細な動作を示すフローチャートである。本装置１０１が立ち上がると、このフローに入るものとする。

【００２７】

まず、終了シグナルを受信したかどうかの判定を行う（ステップＳ２０１）。終了シグナルは、例えば装置１０１に設置された終了ボタン（図示省略）が押された場合などに発行されるものである。終了シグナルが検出されない場合は、そのまま処理を続け、検出された場合は処理を終了する。

【００２８】

次に、問い合わせシグナル受付部１１０で問い合わせシグナルが受信されたかどうかの判定を行う（ステップＳ２０２）。この問い合わせシグナルは、本装置１０１の外部の問い合わせシグナル入力部１０４から送信されるものである。このシグナルを受信するまで、本処理は待機することになる。

【００２９】

問い合わせシグナルを受信すると、問い合わせシグナル受付部１１０は、対話状態取り込み部１０７と動作パラメータ統合部１０９にその旨を通知し、対話状態取り込み部１０７は、装置外部の対話状態検出部１０２を動作させ、現在の対話状態を検知し、現在の対話状態の取り込みを行う（ステップＳ２０３）。ここで、対話状態とは、“ユーザ入力待

10

20

30

40

50

ち”、“システム出力中”等、何らかの対話の状態を表す情報である。場合によっては、複数の状態が検出されることもあり得る。

【 0 0 3 0 】

次に、検知された全対話状態に対応する動作パラメータを対話状態 / 動作パラメータ対応テーブル 1 0 5 から検索する (ステップ S 2 0 4)。検知された対話状態に対応する動作パラメータがテーブル 1 0 5 に存在した場合は (ステップ S 2 0 5)、その検索された動作パラメータを全て選択する (ステップ S 2 0 6)。もし、検知された対話状態に対応する動作パラメータがない場合は、デフォルトの動作パラメータを選択する (ステップ S 2 0 7)。

【 0 0 3 1 】

そして、選択された動作パラメータが複数の場合は、動作パラメータ統合部 1 0 9 において、互いに矛盾したパラメータとならないように統合処理がなされる (ステップ S 2 0 8)。この動作パラメータ統合部 1 0 9 の統合処理では、動作パラメータ検索部 1 0 6 により互いに相容れないパラメータが検索された場合に、例えば、“ A に 1 0 を加えよ ” という動作パラメータと “ A に 3 0 を加えよ ” という動作パラメータが得られた場合は、“ A に 4 0 を加えよ ” という 1 つの動作パラメータに変換される。また、“ A = 1 0 とせよ ” という動作パラメータと、“ A = 3 0 とせよ ” という動作パラメータが検索された場合は、その間を採り “ A = 2 0 とせよ ” という 1 つの動作パラメータに変換される。

【 0 0 3 2 】

こうして、互いに相容れないものが解消された動作パラメータ群は、動作パラメータ出力部 1 0 8 より装置外部に出力される (ステップ S 2 0 9)。出力後は処理の先頭に戻り、再び問い合わせシグナルを受信するまで待機する。

【 0 0 3 3 】

このように本実施形態では、ユーザとの対話状態に応じた動作パラメータを出力することが可能となる。

【 0 0 3 4 】

< 第 1 の実施例 >

次に、図 1 に示した動作パラメータ決定装置をカーナビゲーションシステムに利用した例について、図 3 ~ 図 6 を参照して説明する。

【 0 0 3 5 】

図 3 は、本発明の第 1 の実施例を示すブロック図であり、図 1 に示した動作パラメータ決定装置 1 0 1 を組み込んだカーナビゲーションシステム (以下、単にカーナビと記す) を示している。また、図 4 は、本実施例のカーナビの概観と G U I 画面の例を示す図である。

【 0 0 3 6 】

このカーナビ 4 0 1 は、前述した動作パラメータ決定装置 1 0 1 を搭載しており、動作パラメータ決定装置 1 0 1 から出力された動作パラメータは、動作パラメータ受信部 1 0 3 を介して表示制御部 3 0 2 に供給される。本例では、一定間隔で問い合わせシグナルを送信し、動作パラメータを得るものとする。

【 0 0 3 7 】

表示制御部 3 0 2 は、ナビゲーション本体部 3 0 1 からの地図データ等の画像データを入力して G U I 画面 4 0 5 に表示する機能を有すると共に、動作パラメータ決定装置 1 0 1 から得た動作パラメータに応じて、 G U I 画面 4 0 5 に表示するアイコン等の形状を変更する G U I 変更機能やマイクランプ 4 0 3 の点灯状態を制御する機能を有する。また、ナビゲーション本体部 3 0 1 には、スピーカ 4 0 4 やマイク 4 0 8 が接続されている。

【 0 0 3 8 】

カーナビは、一般的に混合主導のシステムとして構築されている。混合主導とは、システムが主導権を持つ場合とユーザが主導権を持つ場合とが混ざった対話である。例えば、次のような対話が想定される。

U 0 1 : (ボタンを押しながら) 「 近くのコンビニ 」

10

20

30

40

50

S 0 2 : 「 進行方向 5 分圏内に、 4 軒あります。 」

S 0 3 : 「 最も近いコンビニは A B C です。 」

S 0 4 : 「 こちらでよろしいですか？ 」

U 0 5 : 「 はい 」

:

(S はシステムの出力アナウンス、 U はユーザの入力)

システムからのアナウンス (問いかけ) に対し、ユーザがどのアナウンス後に入力すべきかは、そのシステムアナウンスの内容により判断することが可能である。しかし、運転により対話に気が回らない場合やシステム自体に不慣れな場合には、入力すべきタイミングを適切に判断することができないことがある。そこで、発声のガイドとして、本例では、図 4 に示すようなアニメーションアイコン 4 0 2 を表示する。

10

【 0 0 3 9 】

本実施例の動作パラメータ決定装置 1 0 5 が利用する対話状態 / 動作パラメータ対応テーブル 1 0 5 は、例えば図 5 に示すような内容の、対話状態と動作パラメータとが対となったデータが格納されている。

【 0 0 4 0 】

その結果、例えば、ユーザが音声入力可能となる直前のアナウンス出力時 (上の例では S 0 4 のシステムアナウンス出力時) は、“ アニメーション A を出力、マイクランプ点滅 ” という動作パラメータを得る。これにより、カーナビ 4 0 1 内の G U I 画面 4 0 5 には、図 6 (a) に示すようなアニメーションアイコン A (4 0 6) が表示され、マイクランプ 4 0 3 が点滅する。

20

【 0 0 4 1 】

さらに、システムアナウンス S 0 4 が終わり、ユーザの音声入力可能時になると、図 5 のテーブル 1 0 5 の内容から、“ アニメーション B を出力、マイクランプ点灯 ” という動作パラメータが得られるようになる。これにより、G U I 画面 4 0 5 には、図 6 (b) に示すようなアニメーションアイコン B (4 0 7) が表示され、マイクランプ 4 0 3 が点灯状態となる。

【 0 0 4 2 】

このような見た目の変化により、ユーザは、このシステムアナウンスの終了後に音声入力が可能になること、あるいは現在音声入力が可能であることが判断できるようになる。これにより、運転中でシステムアナウンスに気が回らなかった場合や、周囲の騒音等で一時的にシステムアナウンスが聞こえなかった場合でも、入力のタイミングが的確に判るようになる。

30

【 0 0 4 3 】

< 第 2 の実施例 >

本実施例では、図 1 に示した動作パラメータ決定装置を利用した音声合成装置の例について、図 7 ~ 図 1 2 を参照して説明する。

【 0 0 4 4 】

図 7 は、本発明の第 2 の実施例を示すブロック図であり、図 1 に示した動作パラメータ決定装置を組み込んだ音声合成装置の機能構成を示している。

40

【 0 0 4 5 】

この音声合成装置 5 0 1 は、図 1 に示した動作パラメータ決定装置 1 0 1 を備えるほか、図 1 中の動作パラメータ受信部 1 0 3 及び問い合わせシグナル入力部 1 0 4 にそれぞれ相当する音声合成パラメータ受信部 5 0 2 及び問い合わせシグナル送信部 5 0 4 と、装置外部からテキスト情報を取り込むテキスト情報取り込み部 5 0 7 と、音声合成用データを格納する音声合成用データ格納部 5 0 3 と、音声合成処理を行う音声合成部 5 0 6 と、音声合成部 5 0 6 で生成された合成音声出力する合成音出力部 5 0 5 とを備えている。

【 0 0 4 6 】

そして、装置外部には、テキスト情報取り込み部 5 0 7 に対してテキスト情報を供給するテキスト入力部 5 0 9 と、合成音出力部 5 0 5 からの合成音出力するスピーカ等の音

50

声出力装置 508 とが接続されている。テキスト入力部 509 は、対話管理装置 100 内に設けられている。

【0047】

図 8 は、本実施例に係る音声合成装置の動作を示すフローチャートである。

【0048】

音声合成装置 501 は、外部のテキスト入力部 509 からテキスト情報取り込み部 507 を介してテキスト情報を取り込む（ステップ S601）。テキスト情報が取り込まれると、その旨が問い合わせシグナル送信部 504 へ送られる。

【0049】

問い合わせシグナル送信部 504 は、動作パラメータ決定装置 101 内の問い合わせシグナル受付部 110 に対し、動作パラメータの問い合わせシグナルを発行する（ステップ S602）。これにより、現在の対話状態に応じた動作パラメータが決定され、音声合成パラメータ受信部（502）に動作パラメータ（この場合は音声合成パラメータ）が戻される（ステップ S603）。

【0050】

一方、テキスト情報取り込み部 507 に取り込まれたテキスト情報は、音声合成部 506 にも送られる。音声合成部 506 では、動作パラメータ決定装置 101 によって得られた音声合成パラメータと、テキスト情報と、音声合成用データとを用いて音声合成処理を行う（ステップ S604）。音声合成処理に関しては公知の技術を使う。

【0051】

音声合成部 506 で作成された合成音声は、合成音出力部 505 を介して装置外部の音声出力装置 508 へ送られて外部へ出力される（ステップ S605）。

【0052】

図 9 は、本実施例で使用する対話状態 / 動作パラメータ対応テーブル 105 の一例を示す図であり、検知した対話状態と、それに対応する音声合成用の動作パラメータとが対になって格納されている。

【0053】

このように本実施例では、音声合成装置において、検知された対話状態により、合成音声のパラメータを変えることが可能になる。

【0054】

以下に、上記音声合成装置 501 を自動コールセンター（自動チケット予約システム）に応用した例を説明する。

【0055】

この例では、ユーザは電話を通じてシステムと対話を行う。従って、ユーザが用いることができる入力デバイスは音声とプッシュボタンであり、システムからの出力は全て音声情報である。本実施例の対話内容の一例を図 10 に示す。

【0056】

図 11 は、本実施例の対話コンテンツを VoiceXML で記述した一部分を示す図である。

【0057】

対話管理装置 100 は、図 11 に示したような対話コンテンツ 901 を読み込み、ユーザとシステムの対話を管理する。対話管理装置 100 は、各システムアナウンスを出力する際に、テキスト入力部 509 を用いて上記の音声合成装置 501 にテキスト情報を入力する。例えば、図 11 の記述部分 903 のシステムアナウンス（図 10 の S02 に相当）を出力する際の処理は、次のようになる。

【0058】

対話管理装置 100 が、このシステムアナウンス（S02）を出力するために、テキスト入力部 509 を用いて上記の音声合成装置 501 に、当該アナウンス（S02）に相当するテキスト情報を入力する。音声合成装置 501 のテキスト情報取り込み部 507 は、このテキスト情報を取り込み、問い合わせシグナル送信部 504 は、動作パラメータ決定

10

20

30

40

50

装置 101 へ問い合わせシグナルを発行する。

【0059】

動作パラメータ決定装置 101 は、問い合わせシグナル受付部 110 によって問い合わせシグナルを受信すると、対話状態取り込み部 107 を用いて装置外部の対話管理装置 100 にアクセスし、対話状態検出部 102 から現在の対話状態の取り込みを行う。

【0060】

ここで、対話状態とは、音声入力直前のシステムアナウンス出力状態、プッシュボタン入力直前のシステムアナウンス出力状態、及びバージン可能なシステムアナウンス出力状態、のいずれかの状態を示す。場合によっては、複数の状態が検出されることもあり得る。また、バージン可能とは、システムアナウンスをユーザ入力により中断できること
10
であり、VoiceXML では <prompt> タグの bargein 属性で指定することができる。また、VoiceXML では、<prompt> の兄弟要素の <grammar> や <dtmf> を調べることにより、そのアナウンスが音声入力直前のアナウンスであるか、プッシュボタン入力直前のアナウンスであるかを判断可能である。

【0061】

動作パラメータ決定装置 101 は、システムの内部状態や対話コンテンツ 901 を解釈することにより、903 の部分で記述されたシステムアナウンス (S02) の出力時は、“バージン可能なシステムアナウンス出力時”、及び“ユーザが音声入力可能となる直前アナウンスの出力時”であると判断する。従って、この対話状態に対応する動作パラメータは、図 9 に示したテーブル 105 から、“ピッチ周波数 + 40”、“合成話者 = A”
20
となる。

【0062】

動作パラメータ決定装置 101 は、得られた 2 つの動作パラメータを出力し、音声合成装置 501 は、これらの動作パラメータとテキスト情報「ご希望の日を仰って下さい。」とを用いて、合成波形を生成する。ここでは、合成音の話者を A に設定し、基本となるピッチ周波数を 40 Hz 分高く設定し、合成音を生成する。

【0063】

生成された合成音声は、電話回線を介してユーザ側に出力される。ユーザは、このシステムアナウンスの合成音を聞くだけで、そのアナウンスが終了した後に例えば音声入力が可能になることが予測でき、さらにそのシステムアナウンス中にバージン可能であること
30
が判るようになる。

【0064】

なお、タスク (チケット予約等) が終了するまでに必要な対話数により、動作パラメータを変更することも可能である。例えば図 9 のテーブルには、タスク完了までの対話数に合わせて、合成音に特定の音声データを重畳させる指示が記述されている。これにより、ユーザは合成音に重畳された音声データにより、タスクの完了までにどれくらいの対話を行わなければならないかを知ることができる。

【0065】

< 第 3 の実施例 >

本実施例は、図 1 に示した動作パラメータ決定装置を、GUI と音声によるフォーム入力に利用したものである。
40

【0066】

図 12 は、本発明の第 3 の実施例を示す図であり、第 2 の実施例で説明した自動チケット予約システムのタスクを一般的なフォーム入力画面で表した例を示している。

【0067】

同図に示すようなフォーム入力画面 1001 において、ユーザは、マウスとキーボードで各フォーム内容を埋めていてもよいし、マイクからの音声入力によってフォームを埋めてもよい。

【0068】

このような音声入力を許すフォーム入力画面 1001 においては、ユーザが音声入力で
50

きないデータを延々と発声し続ける恐れがあるため、どのデータが音声入力可能であるのかをユーザに示すことが効果的である。図12においては、現時点で音声入力可能なフォーム近傍に、アニメーションアイコン1002を付けている。

【0069】

このアニメーションアイコン1002は、ユーザとの対話状態に応じて形態や動作を変更する。例えば、システムアナウンス中であるか、そうでないかに応じてアイコンの形態や動作を変えてもよい。また、同じシステムアナウンス中であっても、そのアナウンス終了後に音声入力が可能になるか、そうでないかに応じてアイコンの形態や動作を変更してもよい。

【0070】

本発明は、上述した実施形態の装置に限定されず、複数の機器から構成されるシステムに適用しても、1つの機器から成る装置に適用してもよい。前述した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を記憶した記憶媒体をシステムあるいは装置に供給し、そのシステムあるい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し実行することによっても、完成されることは言うまでもない。

【0071】

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を記憶した記憶媒体は本発明を構成することになる。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、フロッピー（登録商標）ディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMを用いることができる。また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、その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形態の機能が実現される場合も含まれることは言うまでもない。

【0072】

さらに、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、次のプログラムコードの指示に基づき、その拡張機能を拡張ボードや拡張ユニットに備わるCPUなどが処理を行って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形態の機能が実現される場合も含まれることは言うまでもない。

【図面の簡単な説明】

【0073】

【図1】本発明の実施形態に係る動作パラメータ決定装置の機能構成図である。

【図2】図1に示した動作パラメータ決定装置の詳細な動作を示すフローチャートである。

【図3】本発明の第1の実施例を示す構成ブロック図である。

【図4】カーナビの概観とGUI画面の例を示す図である。

【図5】第1の実施例に係る対話状態／動作パラメータ対応テーブルの一例を示す図である。

【図6】GUI画面に表示されるアニメーションアイコンの一例を示す図である。

【図7】本発明の第2の実施例を示す構成ブロック図である。

【図8】第2の実施例に係る音声合成装置の動作を示すフローチャートである。

【図9】第2の実施例で使用する対話状態／動作パラメータ対応テーブル105の一例を示す図である。

【図10】第2の実施例の対話内容の一例を示す図である。

【図11】第2の実施例の対話コンテンツをVoiceXMLで記述した一部分を示す図である。

【図12】本発明の第3の実施例を示す図である。

10

20

30

40

50

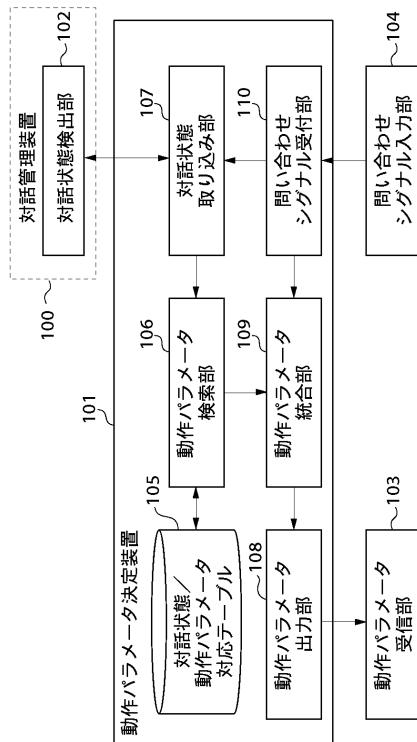
【符号の説明】

【 0 0 7 4 】

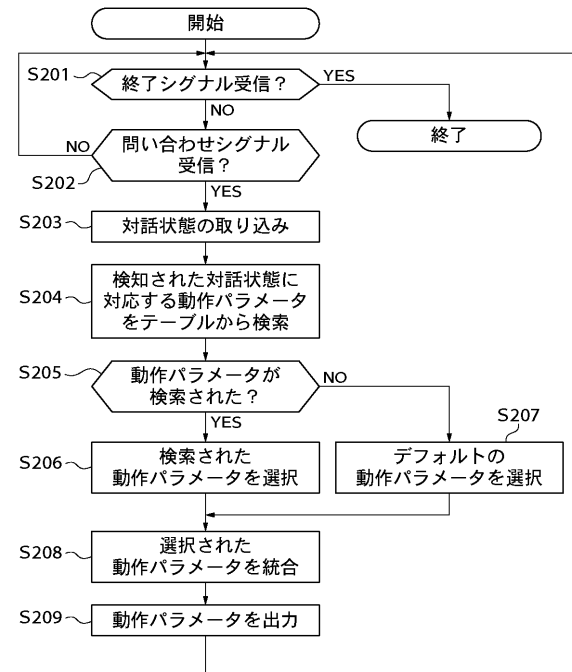
- 1 0 0 対話管理装置
 1 0 1 動作パラメータ決定装置
 1 0 2 対話状態検出部
 1 0 3 動作パラメータ受信部
 1 0 4 問い合わせシグナル入力部
 1 0 5 対話状態 / 動作パラメータ対応テーブル
 1 0 6 動作パラメータ検索部
 1 0 7 対話状態取り込み部
 1 0 8 動作パラメータ出力部
 1 0 9 動作パラメータ統合部
 1 1 0 問い合わせシグナル受付部

10

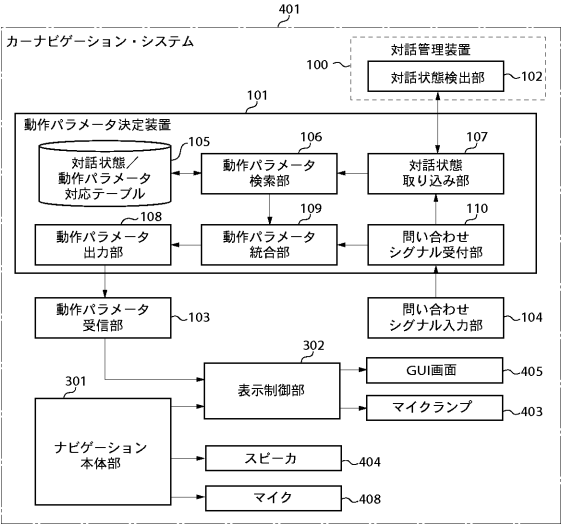
【図 1】



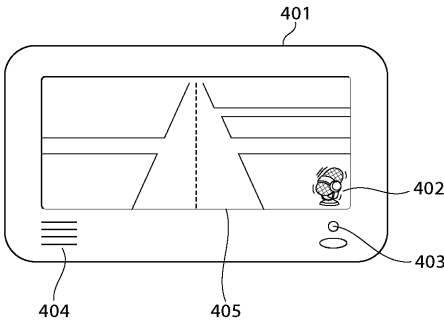
【図 2】



【図 3】



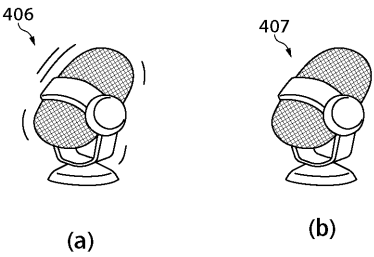
【図 4】



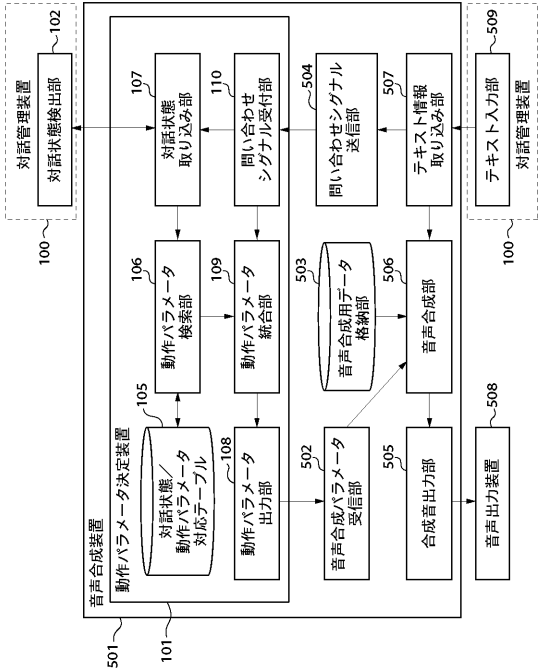
【図 5】

対話状態	動作パラメータ
ユーザが音声入力可能となる 直前アナウンスの出力時	アニメーションAを出力、 マイクランプ点滅
ユーザが音声入力可能時	アニメーションBを出力、 マイクランプ点灯
⋮	⋮

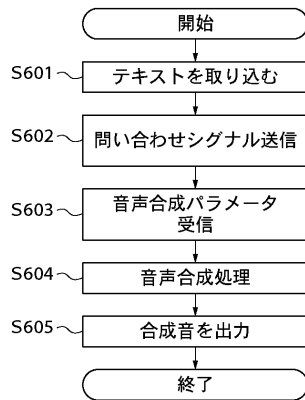
【図 6】



【図 7】



【図 8】



【図 9】

対話状態	動作パラメータ
ユーザが音声入力可能となる直前アナウンスの出力時	合成話者 = A
ユーザがプッシュボタン入力可能となる直前アナウンスの出力時	合成話者 = B
バージン可能なアナウンスの出力時	ピッチ周波数 + 40
タスクが完了するまでに必要な対話数が 1 ~ 2 回	音声データ X を重畳
タスクが完了するまでに必要な対話数が 3 ~ 4 回	音声データ Y を重畳
タスクが完了するまでに必要な対話数が 5 ~ 6 回	音声データ Z を重畳
⋮	⋮

【図 10】

S01: 「こちらはチケット予約システムです。」
 S02: 「ご希望の日を仰って下さい。」
 U03: 「1 2 月 2 日。」
 S04: 「何時の公演を希望しますか。」
 S05: 「1 4 時、1 8 時、2 0 時から選択して下さい。」
 U06: 「1 8 時。」
 S07: 「1 2 月 2 日、1 8 時の公演でよろしいですか。」
 U08: 「はい。」
 S09: 「それでは、予約を致します。」
 S10: 「お客様のユーザIDを打ち込んでください。」
 U11: (“0125” と打ち込む。)
 S12: 「次に、パスワードを打ち込んでください。」
 U13: (“0042” と打ち込む。)
 S14: 「確認しました。」
 S15: 「お客様が登録している、**カードから決済いたします。」
 S16: 「よろしいですね。」
 U17: 「はい。」
 S18: 「決済しました。」
 S19: 「ご利用ありがとうございました。」

(Sはシステムの出力とし、Uはユーザの入力とする)

【図 11】

901

```

<form id="ticket">
  <block>
    <prompt bargein="false">
      こちらはチケット予約システムです。
    </prompt>
  </block>

  <field name="date">
    <prompt bargein="true">
      ご希望の日を仰ってください。
    </prompt grammar src="date.grammar"/>
  </field>
  ⋮

  <field name="UserID">
    <prompt bargein="true">
      お客様のユーザIDを打ち込んでください。
    </prompt>
    <dtmf src="ID.dmf"/>
  </field>
  ⋮
</form>
  
```

902

903

903

【図 12】

Figure 12 is a diagram of a login form window. The window has a title bar with two small square buttons on the right. Inside the window, there are four input fields and one button. The first field is labeled '公演日' (Performance Date) and contains a calendar icon, a box with the number '1', a dropdown arrow, and the character '月' (Month). The second field is labeled '開始時間' (Start Time) and contains a clock icon, a box with the number '14', a dropdown arrow, and the character '時' (Hour). The third field is labeled 'ユーザID' (User ID) and is an empty rectangular box. The fourth field is labeled 'パスワード' (Password) and is an empty rectangular box. At the bottom right of the window is a button labeled '確認' (Confirm). Two callout lines point to the calendar icon and the clock icon, labeled '1001' and '1002' respectively.

1001

1002

公演日 1 ▼ 月

開始時間 14 ▼ 時

ユーザID

パスワード

確認

フロントページの続き

審査官 井上 健一

- (56)参考文献 特開2003-241797(JP,A)
特開2001-142483(JP,A)
特開平10-301675(JP,A)
特開2001-162573(JP,A)
国際公開第2003/065349(WO,A1)
市川 薫,対話理解に対する抑揚情報の役割,情報処理学会研究報告,日本,社団法人情報処理学会,1994年 7月 8日,Vol.94 No.57,p.51-58

- (58)調査した分野(Int.Cl.,DB名)
G10L 15/22
G10L 15/22