(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2021-51172
(P2021-51172A)
(43)公開日　令和3年4月1日(2021.4.1)

| (51)Int.Cl. | | F I | | テーマコード（参考） |
|---|---|---|---|---|
| G10L 13/00 | (2006.01) | G10L | 13/00 | 100M |
| G10L 15/10 | (2006.01) | G10L | 15/10 | 500Z |
| G10L 15/22 | (2006.01) | G10L | 15/22 | 300Z |
| G10L 13/02 | (2013.01) | G10L | 13/02 | 130Z |
| G06F 3/16 | (2006.01) | G06F | 3/16 | 650 |

審査請求 未請求　請求項の数 14　ＯＬ　（全 63 頁）　最終頁に続く

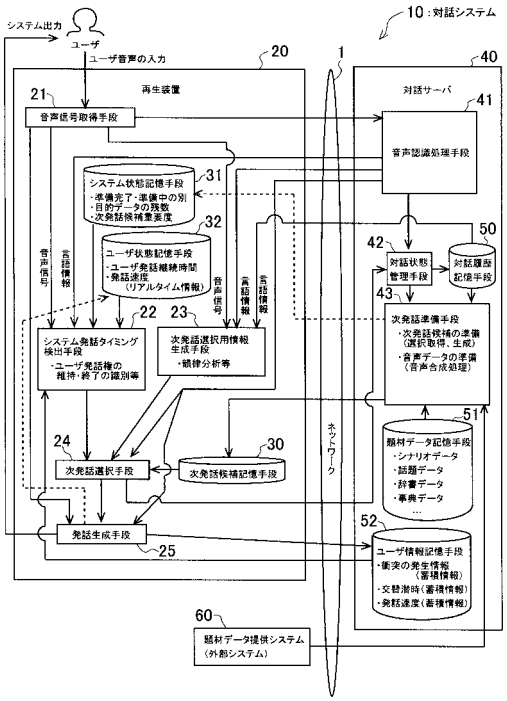
(21)出願番号　特願2019-173551 (P2019-173551)
(22)出願日　令和1年9月24日 (2019.9.24)

(71)出願人　899000068
学校法人早稲田大学
東京都新宿区戸塚町１丁目１０４番地
(74)代理人　100114638
弁理士　中野　寛也
(72)発明者　小林　哲則
東京都新宿区戸塚町１丁目１０４番地　学校法人早稲田大学内
(72)発明者　藤江　真也
東京都新宿区戸塚町１丁目１０４番地　学校法人早稲田大学内

最終頁に続く

(54)【発明の名称】対話システムおよびプログラム

(57)【要約】
【課題】システムの応答性を向上させ、衝突の発生を回避または抑制しつつ、不要に長いシステムの交替潜時の発生を回避または抑制することができる対話システムを提供する。
【解決手段】ユーザ発話の音声信号から抽出した音響特徴量を用いて、音声認識処理手段４１とは非同期で、ユーザ発話権の維持・終了を識別してシステム発話の開始タイミングを検出するシステム発話タイミング検出手段２２と、その検出前に、題材データ記憶手段５１等に記憶された題材データ、対話履歴情報や進行中のユーザ発話の途中までの音声認識処理の結果を用いて、システムの次発話を準備する次発話準備手段４３と、システム発話の開始タイミングの検出後に、次発話準備手段４３により準備された次発話を用いて、システム発話を再生する発話生成手段２５とを設け、対話システム１０を構成した。
【選択図】図１



【特許請求の範囲】
【請求項１】
ユーザとの音声対話のための処理を実行するコンピュータにより構成された対話システムであって、
ユーザ発話の音声信号を取得する音声信号取得手段と、
この音声信号取得手段により取得したユーザ発話の音声信号についての音声認識処理を実行する音声認識処理手段と、
前記音声信号取得手段により取得したユーザ発話の音声信号から音響特徴量を抽出し、抽出した音響特徴量を用いるか、または、この音響特徴量に加え、前記音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報から抽出した言語特徴量を用いて、前記音声認識処理手段による音声認識処理の実行タイミングに依拠しない周期で、ユーザが発話する地位または立場を有していることを示すユーザ発話権の維持または終了を識別するパターン認識処理を繰り返し実行し、このパターン認識処理の結果を用いて、システム発話の開始タイミングを検出する処理を実行するシステム発話タイミング検出手段と、
このシステム発話タイミング検出手段による前記パターン認識処理の前記周期に依拠しないタイミングで、かつ、このシステム発話タイミング検出手段によりシステム発話の開始タイミングが検出される前に、題材データ記憶手段に記憶された題材データまたはネットワークを介して接続された外部システムに記憶された題材データを用いるとともに、ユーザとシステムとの間の対話履歴情報の少なくとも一部および／または前記音声認識処理手段による進行中のユーザ発話についての途中までの音声認識処理の結果を用いて、システムの次発話の内容データを取得または生成する準備処理を実行する次発話準備手段と、
前記システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、前記次発話準備手段による準備処理で得られた次発話の内容データを用いて、システム発話の音声信号の再生を含むシステム発話生成処理を実行する発話生成手段と
を備えたことを特徴とする対話システム。
【請求項２】
前記次発話準備手段は、
次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、
前記システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、前記音声認識処理手段による音声認識処理の結果として得られた言語情報を用いて、前記次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、前記発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段を備えた
ことを特徴とする請求項１に記載の対話システム。
【請求項３】
前記次発話準備手段は、
次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、
前記音声信号取得手段により取得したユーザ発話の音声信号から得られる韻律情報を用いるか、若しくは、この韻律情報に加えて、前記音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いるか、またはこれらの韻律情報およびユーザ発話の言語情報に加えて、ユーザとシステムとの間の対話履歴情報のうちの直前のシステム発話の言語情報を用いて、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図を識別するパターン認識処理を繰り返し実行する次発話選択用情報生成手段と、
前記システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、前記次発話選択用情報生成手段による処理で得られた前記ユーザ発話意図の識別結果を用いて、前記次発話準備手段による準備処理で得られた複数の次発話候補の内容デ

ータの中から、前記発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段と
を備えたことを特徴とする請求項１に記載の対話システム。
【請求項４】
前記次発話準備手段は、
次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、
前記音声信号取得手段により取得したユーザ発話の音声信号から得られる韻律情報を用いるか、若しくは、この韻律情報に加えて、前記音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いるか、またはこれらの韻律情報およびユーザ発話の言語情報に加えて、ユーザとシステムとの間の対話履歴情報のうちの直前のシステム発話の言語情報を用いて、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図を識別するパターン認識処理を繰り返し実行する次発話選択用情報生成手段と、
前記システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、前記次発話選択用情報生成手段による処理で得られた前記ユーザ発話意図の識別結果と、前記音声認識処理手段による音声認識処理の結果として得られた言語情報とを組み合わせて用いて、前記次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、前記発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段と
を備えたことを特徴とする請求項１に記載の対話システム。
【請求項５】
前記次発話準備手段は、
次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、
前記システム発話タイミング検出手段は、
前記ユーザ発話権の維持または終了を識別するパターン認識処理を実行する際に、終了については、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図のうちのいずれのユーザ発話意図で終了するのかを識別するパターン認識処理を実行する構成とされ、
前記システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、前記システム発話タイミング検出手段による処理で得られたユーザ発話意図の識別結果を用いて、前記次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、前記発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段を備えた
ことを特徴とする請求項１に記載の対話システム。
【請求項６】
前記次発話準備手段による準備処理の状態を含むシステム状態を示す情報を記憶するシステム状態記憶手段を備え、
前記システム発話タイミング検出手段は、
前記ユーザ発話権の維持または終了を識別するパターン認識処理の結果および前記システム状態記憶手段に記憶されている前記システム状態を示す情報を用いて、システム発話の開始タイミングを検出する処理を実行する際に、
前記パターン認識処理の結果が前記ユーザ発話権の維持を示している場合には、システム発話の開始タイミングではないと判断し、
前記パターン認識処理の結果が前記ユーザ発話権の終了を示し、かつ、前記システム状態を示す情報が準備完了を示している場合には、システム発話の開始タイミングであると判断し、
前記パターン認識処理の結果が前記ユーザ発話権の終了を示し、かつ、前記システム状態を示す情報が準備中を示している場合には、前記次発話準備手段による準備中の処理内

容に応じ、直ぐに完了する処理内容として予め分類されている処理の準備中であるときには、準備完了になるまで待ってシステム発話の開始タイミングであると判断し、直ぐに完了しない処理内容として予め分類されている処理の準備中であるときには、システム発話の開始タイミングであると判断するとともに、フィラーの挿入タイミングである旨の情報を出力する処理を実行する構成とされている
ことを特徴とする請求項１～５のいずれかに記載の対話システム。
【請求項７】
ユーザ発話継続時間を含むユーザ状態を示す情報を記憶するユーザ状態記憶手段を備え、
前記システム発話タイミング検出手段は、
前記ユーザ発話権の維持または終了を識別するパターン認識処理の結果および前記ユーザ状態記憶手段に記憶されている前記ユーザ状態を示す情報を用いて、システム発話の開始タイミングを検出する処理を実行し、この際の処理として、
（１）前記ユーザ状態記憶手段に記憶されている前記ユーザ発話継続時間が、予め定められた短時間判定用閾値以下または未満の場合には、前記パターン認識処理の結果として得られる尤度に対して設定されているユーザ発話権終了判定用閾値を標準値よりも高く設定し、予め定められた長時間判定用閾値以上または超過の場合には、前記ユーザ発話権終了判定用閾値を標準値よりも低く設定する処理と、
（２）前記ユーザ状態記憶手段に記憶されている前記ユーザ発話継続時間を用いて、前記パターン認識処理の結果として得られる尤度に対するユーザ発話権終了判定用閾値を、前記ユーザ発話継続時間が短いときには当該ユーザ発話権終了判定用閾値が高くなり、前記ユーザ発話継続時間が長いときには当該ユーザ発話権終了判定用閾値が低くなるように予め定められた関数により設定する処理と、
（３）前記ユーザ状態記憶手段に記憶されている前記ユーザ発話継続時間が、予め定められた短時間判定用閾値以下または未満の場合には、前記パターン認識処理の結果が前記ユーザ発話権の終了を示していても、システム発話の開始タイミングではないと判断し、予め定められた長時間判定用閾値以上または超過の場合には、前記パターン認識処理の結果が前記ユーザ発話権の維持を示していても、システム発話の開始タイミングであると判断する処理とのうちのいずれかの処理を実行する構成とされている
ことを特徴とする請求項１～６のいずれかに記載の対話システム。
【請求項８】
システムによる発話開始に対する要求の強さの度合いを示すシステム発話意欲度の指標値として、対話目的を達成するためのシステムの最終の次発話候補の内容データとなり得る目的データの残数および／または前記次発話準備手段による準備処理で得られた次発話候補の内容データの重要度を含むシステム状態を示す情報を記憶するシステム状態記憶手段を備え、
前記システム発話タイミング検出手段は、
前記パターン認識処理の結果として得られる尤度に対するユーザ発話権終了判定用閾値を、前記システム状態記憶手段に記憶されている前記目的データの残数および／または前記重要度で定まる前記システム発話意欲度を用いて、前記システム発話意欲度が強いときには当該ユーザ発話権終了判定用閾値が低くなり、前記システム発話意欲度が弱いときには当該ユーザ発話権終了判定用閾値が高くなるように予め定められた関数により設定する処理を実行する構成とされている
ことを特徴とする請求項２～５のいずれかに記載の対話システム。
【請求項９】
前記次発話準備手段は、
前記音声認識処理手段によるユーザ発話の音声認識処理の結果が新たに出力された場合には、新たに出力された当該音声認識処理の結果を用いて、次発話の候補となる複数の次発話候補の内容データの少なくとも一部を入れ替えるか否かを判定し、入れ替えると判定した場合には、次発話の候補となる別の複数の次発話候補の内容データを取得または生成

する準備処理を実行する構成とされている
ことを特徴とする請求項２～５のいずれかに記載の対話システム。
【請求項１０】
前記次発話準備手段は、
新たに出力された前記音声認識処理の結果を用いて、この結果に含まれる単語のうち予め定められた重要度の高い単語を用いて、ユーザの関心のある話題を決定し、前記題材データ記憶手段に記憶された題材データまたは前記外部システムに記憶された題材データの中から、決定した話題に関連付けられて記憶されている題材データを選択し、次発話の候補となる別の複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされている
ことを特徴とする請求項９に記載の対話システム。
【請求項１１】
前記発話生成手段は、
前記音声信号取得手段により取得したユーザ発話の音声信号と、再生中のシステム発話の音声信号との衝突の発生を検出し、検出した衝突の発生情報を、ユーザ識別情報と関連付けてユーザ情報記憶手段に記憶させるとともに、ユーザ発話の終了からシステム発話の開始までの交替潜時を計測し、計測した交替潜時を、ユーザ識別情報と関連付けて前記ユーザ情報記憶手段に記憶させる処理も実行する構成とされ、
前記システム発話タイミング検出手段は、
前記ユーザ情報記憶手段に記憶されている音声対話相手のユーザとの衝突の発生情報を取得して当該ユーザとの衝突の発生頻度または累積発生回数を算出し、算出した衝突の発生頻度または累積発生回数が上方調整用閾値以上または超過の場合には、前記ユーザ発話権の維持または終了を識別するパターン認識処理の結果として得られる尤度に対して設定されているユーザ発話権終了判定用閾値を標準値または前回調整値よりも高く設定し、
前記ユーザ情報記憶手段に記憶されている音声対話相手のユーザについてのユーザ発話の終了からシステム発話の開始までの複数の交替潜時を取得して当該ユーザについての交替潜時の長短の傾向を示す平均値若しくはその他の指標値を算出し、算出した交替潜時の指標値が下方調整用閾値以上または超過の場合には、前記ユーザ発話権終了判定用閾値を標準値または前回調整値よりも低く設定する処理も実行する構成とされている
ことを特徴とする請求項１～１０のいずれかに記載の対話システム。
【請求項１２】
前記発話生成手段は、
前記音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いて発話速度を算出し、算出した発話速度を、ユーザ識別情報と関連付けて前記ユーザ情報記憶手段に記憶させる処理も実行する構成とされ、
前記システム発話タイミング検出手段は、
前記ユーザ情報記憶手段に記憶されている音声対話相手のユーザについてのユーザ発話の終了からシステム発話の開始までの複数の交替潜時を取得して当該ユーザについての交替潜時の長短の傾向を示す平均値若しくはその他の指標値を算出し、算出した交替潜時の指標値が下方調整用閾値以上または超過の場合に、前記ユーザ発話権終了判定用閾値を標準値または前回調整値よりも低く設定する処理を実行する際に、
前記ユーザ情報記憶手段に記憶されている音声対話相手の複数の発話速度を取得して当該ユーザの発話速度の傾向を示す平均値若しくはその他の指標値を算出し、前記下方調整用閾値を、算出した前記発話速度の指標値を用いて、前記発話速度の指標値が大きいときには当該下方調整用閾値が小さくなり、前記発話速度の指標値が小さいときには当該下方調整用閾値が大きくなるように予め定められた関数により設定する処理を実行する構成とされている
ことを特徴とする請求項１１に記載の対話システム。
【請求項１３】
ユーザのリアルタイムの発話速度を含むユーザ状態を示す情報を記憶するユーザ状態記

憶手段を備え、
前記発話生成手段は、
前記音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いてリアルタイムの発話速度を算出し、算出したリアルタイムの発話速度を前記ユーザ状態記憶手段に記憶させる処理も実行する構成とされ、
前記システム発話タイミング検出手段は、
前記音声信号取得手段により取得したユーザ発話の音声信号から音響特徴量を抽出し、抽出した音響特徴量および前記ユーザ状態記憶手段に記憶されているリアルタイムの発話速度を用いるか、または、これらの音響特徴量およびリアルタイムの発話速度に加え、前記音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報から抽出した言語特徴量を用いて、前記音声認識処理手段による音声認識処理の実行タイミングに依拠しない周期で、前記ユーザ発話権の維持または終了を識別するパターン認識処理を繰り返し実行し、このパターン認識処理の結果を用いて、システム発話の開始タイミングを検出する処理を実行する構成とされている
ことを特徴とする請求項１～１２のいずれかに記載の対話システム。
【請求項１４】
請求項１～１３のいずれかに記載の対話システムとして、コンピュータを機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ユーザとの音声対話のための処理を実行するコンピュータにより構成された対話システムおよびプログラムに係り、例えば、ニュースやコラムや歴史等の各種の話題を記載した記事データから生成したシナリオデータを用いてユーザに対して記事の内容を伝達するニュース対話システム、ユーザに対して機器の使用方法の説明や施設の案内等を行うガイダンス対話システム、選挙情勢や消費者志向等の各種のユーザの動向調査を行うアンケート対話システム、ユーザが店舗・商品・旅行先・聞きたい曲等の情報検索を行うための情報検索対話システム、ユーザが家電機器や車等の各種の機器や装置等を操作するための操作対話システム、子供や学生や新入社員等であるユーザに対して教育を行うための教育対話システム、システムがユーザ属性等の情報を特定するための情報特定対話システム等に利用できる。
【背景技術】
【０００２】
一般に、音声対話システムは、人であるユーザと、コンピュータシステムである自身との間で、互いに主に音声チャネルを通じた言語情報のやりとりを行うことにより、所望のタスクを実行し、その目的（例えば、ユーザへのニュース等の記事の内容の伝達、ユーザに対するガイダンス、ユーザへのアンケート、ユーザによる情報検索、ユーザによる機器等の操作、ユーザの教育、システムによる情報特定等）を達成するものである。
【０００３】
より詳細には、従来の音声対話システムでは、先ず、ユーザ発話の音声信号を取得し（音声信号取得）、連続的に得られる音声信号から、ユーザの発話が途切れたことを手がかりとして発話単位の音声信号を切り出す発話区間検出を行い（発話区間検出）、次に、得られた発話区間の音声信号を言語情報に変換する音声認識処理を行うことにより、検出したユーザ発話の意味を推定し（音声認識）、続いて、推定した意味に応じて次発話を決定し、すなわち得られたユーザの言語情報に適したシステム発話の内容を生成し（発話内容生成）、さらに、その発話内容を音声信号に変換する音声合成処理を行い（音声合成）、その後、システム発話の内容をユーザに伝達するため、生成したシステム発話の音声信号を再生する処理を行う（音声信号再生）。従来の音声対話システムは、これらの一連の処理を、原則的にはシーケンシャルに行うため、それぞれの処理における遅延が蓄積することで、ユーザが発話を完了してから、システムが応答するまでに長い遅延が生じることに

なる。
【０００４】
音声対話における二者間の発話の間（ま）の長さを交替潜時と呼ぶが、人同士の円滑な対話における交替潜時は、平均的には０．６秒程度であり、長くとも１秒程度である。また、相手の発話が終了する前に、発話を開始することも多く、これを衝突と呼ぶ。一方、近年普及しているスマートスピーカ等の対話システムと人との対話においては、ユーザの発話終了からシステムの発話開始までの間（ま）（以下、特にユーザからシステムという方向性を持たせた交替潜時を指すときは、システムの交替潜時と呼ぶ。）が、１秒から数秒となることが多い。従来の研究によれば、一方の交替潜時が他方の交替潜時に影響を与えるとされているので、システムの交替潜時が不要に長くなると、これに影響されてユーザの間（ま）（システムの発話終了からユーザの応答開始までに要する時間）も長くなる。これにより、対話全体に要する時間が不要に長くなるため、タスク達成の効率や、ユーザ体験の観点から好ましくない。
【０００５】
従って、システムの応答性を向上させることにより、上述した従来生じていたユーザ発話とシステム発話との間に生じる不要に長い無音の時間を短くするか、あるいは発生そのものを避けることが望ましく、それを実現するためには、システム発話の開始タイミングを適切に検出することが必要となる。なぜなら、システムの交替潜時を短くするためにシステム発話の開始タイミングを不当に早めるような方法で検出処理を行えば、衝突が発生する可能性が高くなるので、単純にシステム発話の開始タイミングが早まる方法を採用すればよいというものではないからである。
【０００６】
より詳細には、従来の音声対話システムでは、ユーザ発話の終了時をシステム発話の開始タイミングとみなしていた。１対１の対話においては、これは極めて自然な考え方であるが、そもそもユーザ発話が終了する現象の定義が明確ではなかった。例えば、特定の長さ（例えば、１００ミリ秒以上）のポーズで区切られた音声区間をＩｎｔｅｒ－Ｐａｕｓａｌ　Ｕｎｉｔ（ＩＰＵ）と呼び、音声分析や会話分析では音声区間の単位として広く用いられているが、１００ミリ秒程度の無音区間は、１人の話者の発話区間内にも頻繁に生じるため、必ずしもその前後で話者交替が起こるわけではない。そのため、ユーザ発話の音声信号における短い無音区間をシステム発話の開始タイミングの検出に用いると、生成して再生を開始したシステム発話と、継続されたユーザ発話とがオーバーラップする衝突を起こし、対話を崩してしまう可能性がある。一方、より長い無音区間で区切ることにより、オーバーラップ（衝突）を防ぐことはできるが、システム発話の開始タイミングは、無音区間の長さだけ遅れ、ユーザ発話とシステム発話との間の無音区間を短くすることができなくなる。
【０００７】
また、従来の音声認識では、音声認識対象とする音声区間を決定するために音声区間検出（Ｖｏｉｃｅ　Ａｃｔｉｖｉｔｙ　Ｄｅｔｅｃｔｉｏｎ；ＶＡＤ）と呼ばれる処理を行う。音声信号の振幅やゼロ交差数を閾値処理する単純なものから、音声信号から得られる特徴量に基づき確率的に音声が含まれるか否かを確定するモデルなど、様々な手法が研究されてきた。しかし、システム発話の開始タイミングを早期に決定するということを意図した手法は提案されていなかった。
【０００８】
さらに、システム発話の開始タイミングを決定するために、ユーザ発話の継続または終了、あるいはシステムが次にどのような行動をとるべきか（発話だけに限らず、相槌なども含む）を検出する技術も、本願発明者らにより研究されているが、ユーザ発話途中でのシステムの相槌・復唱の生成技術を除けば、これらは全て音声認識と同様にＶＡＤを前提としており、ＶＡＤ処理による遅延の影響を排除することができない。
【０００９】
これらの従来技術に対し、本願発明者らは、音声信号を逐次処理し、短い周期（例えば

、１０ミリ秒～１００ミリ秒）で音声信号から音響特徴量を抽出し、抽出した音響特徴量を用いて、システムが発話をすべきか否かの識別を行う技術、換言すれば、ユーザが発話する地位または立場を有していることを示すユーザ発話権の維持または終了（終了には、譲渡、放棄が含まれる。）を識別する技術を開発した（非特許文献１，２参照）。このようにすることで、音声区間検出処理（ＶＡＤ処理）による遅延なしにシステム発話の開始タイミングを決定することができる。

【００１０】

なお、本発明では、複数の次発話候補が準備された場合に、その中から次発話を選択する処理が行われるが、この選択処理を行うために必要となる情報を生成する技術としては、本願発明者らにより開発された、韻律分析によりユーザ発話意図を推定する技術が知られている（非特許文献３参照）。

【００１１】

また、本発明は、例えば、ニュース対話システム、ガイダンス対話システム、アンケート対話システム、情報検索対話システム、操作対話システム、教育対話システム等の各種の対話システムに適用することができるが、ユーザへの効率的な情報伝達を実現することができる対話システムとしては、本願発明者らにより開発された、主計画および副計画からなるシナリオデータを用いてユーザに対してニュース等の記事の内容を伝達するニュース対話システムが知られている（非特許文献４参照）。

【先行技術文献】

【非特許文献】

【００１２】

【非特許文献１】藤江真也、横山勝矢、小林哲則、“音声対話システムのためのユーザ発話終了タイミングの逐次予測”、日本音響学会講演論文集、２０１８

【非特許文献２】藤江真也、横山勝矢、小林哲則、“音声対話システムのためのユーザの発話権維持状態の逐次推定”、人工知能学会全国大会、２Ｎ１－０３、Ｊｕｎｅ２０１８

【非特許文献３】高津弘明、横山勝矢、本田裕、藤江真也、小林哲則、“システム発話の文脈を考慮した発話意図理解”、言語処理学会　第２５回年次大会　発表論文集、ｐｐ．３２０－３２３、２０１９

【非特許文献４】高津弘明、福岡維新、藤江真也、林良彦、小林哲則、“意図性の異なる多様な情報行動を可能とする音声対話システム”、人工知能学会論文誌、ｖｏｌ．２２、ｎｏ．１、ｐ．ＤＳＨ－Ｃ＿１－２４、２０１８

【発明の概要】

【発明が解決しようとする課題】

【００１３】

従来の音声対話システムでは、前述したように、音声信号取得、発話区間検出、音声認識、発話内容生成、音声合成、音声信号再生という一連の処理を、シーケンシャルに行うため、それぞれの処理における遅延が蓄積するという問題があった。

【００１４】

また、前述した非特許文献１，２に記載された技術を用いれば、短い周期（例えば、１０ミリ秒～１００ミリ秒）で音声信号から抽出した音響特徴量を用いてユーザ発話権の維持または終了を識別するパターン認識処理を行うので、音声区間検出処理（ＶＡＤ処理）による遅延なしにシステム発話の開始タイミングを決定することができる。

【００１５】

しかし、非特許文献１，２に記載された技術を用いれば、システム発話の開始タイミングを、ＶＡＤ処理による遅延なしに早期に、かつ、衝突の発生を回避または抑制しながら適切に、決定することができるものの、その後の処理、すなわち、前述した一連の処理のうちの音声認識、発話内容生成、音声合成、音声信号再生の各処理を、従来通りにシーケンシャルに行うと、そこでの遅延が生じるという問題がある。

【００１６】

従って、非特許文献１，２に記載された技術を利用してシステム発話の開始タイミング

を早期かつ適切に決定しつつ、ユーザ発話とシステム発話との間に生じる不要に長い無音の時間を短くするか、あるいは発生を回避することができる技術の開発が望まれる。

【００１７】

本発明の目的は、システムの応答性を向上させることができ、衝突の発生を回避または抑制しつつ、不要に長いシステムの交替潜時の発生を回避または抑制することができる対話システムおよびプログラムを提供するところにある。

【課題を解決するための手段】

【００１８】

＜本発明の基本構成＞

【００１９】

本発明は、ユーザとの音声対話のための処理を実行するコンピュータにより構成された対話システムであって、

ユーザ発話の音声信号を取得する音声信号取得手段と、

この音声信号取得手段により取得したユーザ発話の音声信号についての音声認識処理を実行する音声認識処理手段と、

音声信号取得手段により取得したユーザ発話の音声信号から音響特徴量を抽出し、抽出した音響特徴量を用いるか、または、この音響特徴量に加え、音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報から抽出した言語特徴量を用いて、音声認識処理手段による音声認識処理の実行タイミングに依拠しない周期で、ユーザが発話する地位または立場を有していることを示すユーザ発話権の維持または終了を識別するパターン認識処理を繰り返し実行し、このパターン認識処理の結果を用いて、システム発話の開始タイミングを検出する処理を実行するシステム発話タイミング検出手段と、

このシステム発話タイミング検出手段によるパターン認識処理の周期に依拠しないタイミングで、かつ、このシステム発話タイミング検出手段によりシステム発話の開始タイミングが検出される前に、題材データ記憶手段に記憶された題材データまたはネットワークを介して接続された外部システムに記憶された題材データを用いるとともに、ユーザとシステムとの間の対話履歴情報の少なくとも一部および／または音声認識処理手段による進行中のユーザ発話についての途中までの音声認識処理の結果を用いて、システムの次発話の内容データを取得または生成する準備処理を実行する次発話準備手段と、

システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、次発話準備手段による準備処理で得られた次発話の内容データを用いて、システム発話の音声信号の再生を含むシステム発話生成処理を実行する発話生成手段と

を備えたことを特徴とするものである。

【００２０】

ここで、ユーザ発話権の「終了」には、放棄および譲渡の双方が含まれる。放棄は、自分の発話を終了させるだけの場合であり、譲渡は、相手への質問等のように、自分の発話を終了させるとともに、相手の発話開始を促す場合である。

【００２１】

また、「次発話準備手段」における「題材データ記憶手段に記憶された題材データまたはネットワークを介して接続された外部システムに記憶された題材データ」には、例えば、ニュース等の各種の話題をシナリオ化したシナリオデータ、シナリオ化されていない各種の話題データ、辞書データ、事典データ、機器の使用方法や施設等のガイダンス用データ、アンケート調査用データ、機器や装置等の操作補助用データ、教育用データ等が含まれる。

【００２２】

さらに、「次発話準備手段」による「システムの次発話の内容データを取得または生成」の「取得」には、題材データ記憶手段やネットワークを介して接続された外部システムに記憶されている複数の題材データの中からの必要な題材データ（使用するか、または使用する可能性のある題材データ）の選択的な取得と、題材データ記憶手段や外部システムに記憶されている任意の１つの題材データの構成要素の中からの必要な構成要素（使用す

るか、または使用する可能性のある構成要素）の選択的な取得とが含まれる。
【００２３】
また、上記の「次発話準備手段」における「生成」には、取得した言語情報（題材データまたはその構成要素であるテキストデータ）の加工（語尾等の部分的な変換調整、結合等）が含まれる。但し、題材データは、題材データ記憶手段や外部システムに記憶されている段階で、予め加工されていることが好ましい。そして、「生成」には、テキストデータから音声データへの変換（音声合成）も含まれる。なお、題材データ記憶手段や外部システムに、題材データまたはその構成要素として、音声データ（例えばｗａｖファイル等）が既に用意されている場合には、「次発話準備手段」による音声合成処理は行わなくてもよい。
【００２４】
さらに、「次発話準備手段」により準備される「システムの次発話の内容データ」は、テキストデータおよびこれに対応する音声データの場合と、テキストデータだけの場合とがある。但し、「発話生成手段」の処理負荷の軽減および遅延防止の観点からは、「次発話準備手段」により音声データも併せて準備することが好ましい。そして、対話中に、付帯的な情報として、映像（動画）や静止画を再生する場合には、「システムの次発話の内容データ」には、映像データや画像データが付随していてもよく、対話中に音楽を再生する場合には、「システムの次発話の内容データ」には、楽曲データが含まれていてもよい。
【００２５】
また、「次発話準備手段」における「ユーザとシステムとの間の対話履歴情報の少なくとも一部および／または音声認識処理手段による進行中のユーザ発話についての途中までの音声認識処理の結果を用いて」の「対話履歴情報の少なくとも一部」を用いることには、対話履歴情報（システム発話、ユーザ発話）の全体を用いること、直前のシステム発話のみを用いること、直前のシステム発話を用いずにそれよりも前のシステム発話やユーザ発話を用いること（例えば、ユーザの「さっき言ってたＸＸＸのこと、もう少し詳しく聞きたいんだけど・・・」等の要求に応答する場合等）、直前のユーザ発話のみを用いること等が含まれる。そして、「音声認識処理手段による進行中のユーザ発話についての途中までの音声認識処理の結果」を用いることは、進行中のユーザ発話の部分的な内容（ユーザ発話の発話区間全体の内容ではなく、その途中までの部分的な内容）を用いることである。なお、部分的な音声認識処理の結果が得られた場合に、それがユーザ発話の発話区間の最後の部分であるか否かは、その時点では判らないことがあるが、結果的にそれがユーザ発話の発話区間の最後の部分であった場合には、直前のユーザ発話ということになり、「対話履歴情報の少なくとも一部」に該当する。
【００２６】
さらに、「発話生成手段」における「システム発話の音声信号の再生を含むシステム発話生成処理」には、次発話のテキストデータについての音声合成が未だ済んでいない場合には、音声合成処理が含まれる。なお、前述した通り、「発話生成手段」の処理負荷の軽減および遅延防止の観点からは、音声データ（例えばｗａｖファイル等）は、次発話準備手段による準備処理で用意することが好ましい。さらに、「次発話準備手段」により準備された「システムの次発話の内容データ」に映像データや画像データが付随している場合には、「発話生成手段」における「システム発話生成処理」には映像や画像の再生処理も含まれ、「システムの次発話の内容データ」に楽曲データが含まれている場合には、「システム発話生成処理」には音楽の再生処理も含まれる。
【００２７】
＜本発明の基本構成の作用・効果＞
【００２８】
このような本発明の対話システムにおいては、システム発話タイミング検出手段により、ユーザが自己の発話権を維持しているか、または、譲渡若しくは放棄により終了させたかをパターン認識処理により逐次推定するとともに、次発話準備手段により、システム発

話タイミング検出手段によるパターン認識処理とは非同期で、かつ、システム発話タイミング検出手段によりシステム発話の開始タイミングが検出される前に、ユーザ発話に対するシステムの次発話の内容データを準備する。すなわち、ユーザ発話に対するシステムの次発話の内容データを、当該ユーザ発話の進行中に、または、それよりも前の段階である当該ユーザ発話の開始前に準備しておく。

【００２９】

このため、対話相手であるユーザが自己の発話権を譲渡若しくは放棄することによりユーザ発話権が終了し、システム発話タイミング検出手段により、このユーザ発話権の終了が捉えられ、システム発話の開始タイミングが検出された場合には、その検出直後に、発話生成手段により、タイミングよくシステム発話を開始させることが可能となるので、システムの応答性を向上させることが可能となる。

【００３０】

また、システム発話タイミング検出手段は、音声認識処理手段による音声認識処理とは非同期で、ユーザ発話権の維持または終了を識別するパターン認識処理を繰り返し実行する構成とされているので、音声区間検出処理（ＶＡＤ処理）を前提としない処理を実現することができるため、ＶＡＤ処理による遅延なしに早期に、システム発話の開始タイミングを決定することができるとともに、ユーザ発話とシステム発話との衝突も回避または抑制することができる。

【００３１】

以上より、本発明では、次発話準備手段により、システムが発話すべき内容を早期に確定したうえで、システム発話タイミング検出手段により、ユーザ発話権の終了が推定され、システム発話の開始タイミングが検出されるのを待って、発話生成手段により、システム応答を行うので、ユーザ発話の終了後、システム発話の開始までに、長い間（ま）が空くことを避けることができるうえ、両者の発話の衝突の発生も回避または抑制することができ、これらにより前記目的が達成される。

【００３２】

また、本発明では、次発話準備手段により、次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行し、この準備処理で得られた複数の次発話候補の内容データの中から、発話生成手段で用いる次発話の内容データを選択する処理を実行するようにしてもよい。これにより、様々な種別の対話に対応可能となる。具体的には、以下のような構成を採用することができる。

【００３３】

＜音声認識処理の結果として得られた言語情報を用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成＞

【００３４】

すなわち、前述した対話システムにおいて、

次発話準備手段は、

次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、

システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、音声認識処理手段による音声認識処理の結果として得られた言語情報を用いて、次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段を備えた構成を採用することができる。

【００３５】

このように音声認識処理の結果として得られた言語情報を用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成とした場合には、次発話選択手段により、ユーザ発話の内容に応じて、システムの次発話の内容データを選択することができる。

【００３６】

このため、例えば、次発話準備手段により、直前のシステム発話の内容に基づき、またはそれまでの対話履歴（システム発話、ユーザ発話）の内容に基づき、システムの次発話での使用が想定される複数の次発話候補の内容データを準備しておき、ユーザ発話の内容に応じて、準備した複数の次発話候補の内容データの中から、次発話の内容データを選択することができる。

【００３７】

また、例えば、次発話準備手段により、進行中のユーザ発話の途中までの部分的な内容（ユーザ発話の開始時点から途中の時点までの内容、あるいは、ユーザ発話の途中の時点から別の途中の時点までの内容）に応じ、システムの次発話での使用が想定される複数の次発話候補の内容データを準備しておき、その後の発話内容（途中の時点以降、あるいは別の途中の時点以降の発話内容）を含めたユーザ発話の発話区間全体の内容に応じて、準備した複数の次発話候補の内容データの中から、次発話の内容データを選択することもできる。

【００３８】

＜韻律分析で推定したユーザ発話意図を用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成＞

【００３９】

また、前述した対話システムにおいて、

次発話準備手段は、

次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、

音声信号取得手段により取得したユーザ発話の音声信号から得られる韻律情報を用いるか、若しくは、この韻律情報に加えて、音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いるか、またはこれらの韻律情報およびユーザ発話の言語情報に加えて、ユーザとシステムとの間の対話履歴情報のうちの直前のシステム発話の言語情報を用いて、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図を識別するパターン認識処理を繰り返し実行する次発話選択用情報生成手段と、

システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、次発話選択用情報生成手段による処理で得られたユーザ発話意図の識別結果を用いて、次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段と

を備えた構成を採用することができる。

【００４０】

ここで、「質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図」における質問、応答、相槌等は、ユーザ発話意図の例示列挙であり、ここに列挙されていない「その他」のユーザ発話意図を用意してもよい。また、質問、応答、相槌等は、例示列挙であるので、これらの各々は必須ではなく、別の定義のユーザ発話意図を用意してもよい。他の発明においても同様である。

【００４１】

このように韻律分析で推定したユーザ発話意図を用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成とした場合には、前述した＜音声認識処理の結果として得られた言語情報を用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成＞の場合と同様な作用・効果が得られることに加え、ユーザ発話意図を用いるので、音声認識処理の結果を得ることなく、次発話の内容データを選択することが可能となるため、システムの応答性を向上させることが可能となる。

【００４２】

なお、次発話選択用情報生成手段のユーザ発話意図の識別器と、システム発話タイミング検出手段のユーザ発話権の維持・終了の識別器とは、マルチタスクの識別器とすることにより一体化させてもよい。

【００４３】
また、次発話選択用情報生成手段で用いる韻律情報を得るための分析処理は、システム発話タイミング検出手段で用いる音響特徴量を抽出するための分析処理と共通の処理としてもよい。
【００４４】
＜韻律分析で推定したユーザ発話意図と、音声認識処理の結果として得られた言語情報とを組み合わせて用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成＞
【００４５】
さらに、前述した対話システムにおいて、
次発話準備手段は、
次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、
音声信号取得手段により取得したユーザ発話の音声信号から得られる韻律情報を用いるか、若しくは、この韻律情報に加えて、音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いるか、またはこれらの韻律情報およびユーザ発話の言語情報に加えて、ユーザとシステムとの間の対話履歴情報のうちの直前のシステム発話の言語情報を用いて、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図を識別するパターン認識処理を繰り返し実行する次発話選択用情報生成手段と、
システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、次発話選択用情報生成手段による処理で得られたユーザ発話意図の識別結果と、音声認識処理手段による音声認識処理の結果として得られた言語情報とを組み合わせて用いて、次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段と
を備えた構成を採用することができる。
【００４６】
このように韻律分析で推定したユーザ発話意図と、音声認識処理の結果として得られた言語情報とを組み合わせて用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成とした場合には、ユーザ発話意図を用いるだけでは、対応できないときでも、あるいは、音声認識処理の結果を用いるだけでは、対応できないときでも、次発話の内容データの選択処理を行うことができるようになるので、あらゆるタイプの音声対話に対応可能となる。
【００４７】
＜システム発話タイミング検出手段によりユーザ発話意図の識別も行う構成＞
【００４８】
また、前述した対話システムにおいて、
次発話準備手段は、
次発話の候補となる複数の次発話候補の内容データを取得または生成する準備処理を実行する構成とされ、
システム発話タイミング検出手段は、
ユーザ発話権の維持または終了を識別するパターン認識処理を実行する際に、終了については、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図のうちのいずれのユーザ発話意図で終了するのかを識別するパターン認識処理を実行する構成とされ、
システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された後に、システム発話タイミング検出手段による処理で得られたユーザ発話意図の識別結果を用いて、次発話準備手段による準備処理で得られた複数の次発話候補の内容データの中から、発話生成手段で用いる次発話の内容データを選択する処理を実行する次発話選択手段を備えた構成を採用することができる。

【００４９】
このようにシステム発話タイミング検出手段によりユーザ発話意図の識別も行う構成とした場合には、システム発話タイミング検出手段によりシステム発話の開始タイミングが検出された時点で、同時にユーザ発話意図の識別結果も得られているので、システムの応答性を向上させることが可能となる。
【００５０】
＜システム発話タイミング検出手段により、システム状態を示す情報である準備完了・準備中の別を参照する構成＞
【００５１】
また、以上に述べた対話システムにおいて、
次発話準備手段による準備処理の状態を含むシステム状態を示す情報を記憶するシステム状態記憶手段を備え、
システム発話タイミング検出手段は、
ユーザ発話権の維持または終了を識別するパターン認識処理の結果およびシステム状態記憶手段に記憶されているシステム状態を示す情報を用いて、システム発話の開始タイミングを検出する処理を実行する際に、
パターン認識処理の結果がユーザ発話権の維持を示している場合には、システム発話の開始タイミングではないと判断し、
パターン認識処理の結果がユーザ発話権の終了を示し、かつ、システム状態を示す情報が準備完了を示している場合には、システム発話の開始タイミングであると判断し、
パターン認識処理の結果がユーザ発話権の終了を示し、かつ、システム状態を示す情報が準備中を示している場合には、次発話準備手段による準備中の処理内容に応じ、直ぐに完了する処理内容として予め分類されている処理の準備中であるときには、準備完了になるまで待ってシステム発話の開始タイミングであると判断し、直ぐに完了しない処理内容として予め分類されている処理の準備中であるときには、システム発話の開始タイミングであると判断するとともに、フィラーの挿入タイミングである旨の情報を出力する処理を実行する構成としてもよい。
【００５２】
このようにシステム発話タイミング検出手段により、システム状態を示す情報である準備完了・準備中の別を参照する構成とした場合には、システム状態を考慮し、より適切なシステム発話の開始タイミングを検出することが可能となる。
【００５３】
＜システム発話タイミング検出手段により、ユーザ状態を示す情報であるユーザ発話継続時間を用いて、ユーザ発話権終了判定用閾値の調整を行うか、またはシステム発話の開始タイミングであるか否かの判断を行う構成＞
【００５４】
さらに、以上に述べた対話システムにおいて、
ユーザ発話継続時間を含むユーザ状態を示す情報を記憶するユーザ状態記憶手段を備え、
システム発話タイミング検出手段は、
ユーザ発話権の維持または終了を識別するパターン認識処理の結果およびユーザ状態記憶手段に記憶されているユーザ状態を示す情報を用いて、システム発話の開始タイミングを検出する処理を実行し、この際の処理として、
（１）ユーザ状態記憶手段に記憶されているユーザ発話継続時間が、予め定められた短時間判定用閾値以下または未満の場合には、パターン認識処理の結果として得られる尤度に対して設定されているユーザ発話権終了判定用閾値を標準値よりも高く設定し、予め定められた長時間判定用閾値以上または超過の場合には、ユーザ発話権終了判定用閾値を標準値よりも低く設定する処理と、
（２）ユーザ状態記憶手段に記憶されているユーザ発話継続時間を用いて、パターン認識処理の結果として得られる尤度に対するユーザ発話権終了判定用閾値を、ユーザ発話継

続時間が短いときには当該ユーザ発話権終了判定用閾値が高くなり、ユーザ発話継続時間が長いときには当該ユーザ発話権終了判定用閾値が低くなるように予め定められた関数により設定する処理と、

（３）ユーザ状態記憶手段に記憶されているユーザ発話継続時間が、予め定められた短時間判定用閾値以下または未満の場合には、パターン認識処理の結果がユーザ発話権の終了を示していても、システム発話の開始タイミングではないと判断し、予め定められた長時間判定用閾値以上または超過の場合には、パターン認識処理の結果がユーザ発話権の維持を示していても、システム発話の開始タイミングであると判断する処理とのうちのいずれかの処理を実行する構成としてもよい。

【００５５】

ここで、「ユーザ発話権終了判定用閾値」の「標準値」は、別の情報に基づく別の趣旨での閾値調整が別途に行われている場合には、その別途の閾値調整後の値を指す。

【００５６】

このようにシステム発話タイミング検出手段により、ユーザ状態を示す情報であるユーザ発話継続時間を用いて、ユーザ発話権終了判定用閾値の調整を行う（上記（１）、（２））か、またはシステム発話の開始タイミングであるか否かの判断を行う（上記（３））構成とした場合には、ユーザ発話継続時間の長短に応じ、システム発話の開始タイミングを調整することが可能となる。

【００５７】

上記（１）、（２）では、ユーザ発話継続時間が短いときにはユーザ発話権終了判定用閾値が高くなり、ユーザ発話継続時間が長いときにはユーザ発話権終了判定用閾値が低くなるように設定することができるので、ユーザ発話の開始直後の時期には、ユーザ発話権が終了したという識別結果が出にくい設定状態とし、ユーザ発話の開始時点から比較的長時間が経過している時期には、ユーザ発話権が終了したという識別結果が出やすい設定状態とすることができる。

【００５８】

上記（３）では、ユーザ発話継続時間を、ユーザ発話権終了判定用閾値に反映させるのではなく、ユーザ発話権終了判定用閾値を用いて維持・終了の識別結果を出した後におけるシステム発話の開始タイミングの判断処理に反映させることにより、上記（１）、（２）と同様な作用・効果を得る。

【００５９】

＜システム発話タイミング検出手段により、システム状態を示す情報であるシステム発話意欲度を用いてユーザ発話権終了判定用閾値を動的に調整する構成＞

【００６０】

また、前述した次発話準備手段により準備した複数の次発話候補の内容データの中から、次発話選択手段により次発話の内容データを選択する構成とした場合において、

システムによる発話開始に対する要求の強さの度合いを示すシステム発話意欲度の指標値として、対話目的を達成するためのシステムの最終の次発話候補の内容データとなり得る目的データの残数および／または次発話準備手段による準備処理で得られた次発話候補の内容データの重要度を含むシステム状態を示す情報を記憶するシステム状態記憶手段を備え、

システム発話タイミング検出手段は、

パターン認識処理の結果として得られる尤度に対するユーザ発話権終了判定用閾値を、システム状態記憶手段に記憶されている目的データの残数および／または重要度で定まるシステム発話意欲度を用いて、システム発話意欲度が強いときには当該ユーザ発話権終了判定用閾値が低くなり、システム発話意欲度が弱いときには当該ユーザ発話権終了判定用閾値が高くなるように予め定められた関数により設定する処理を実行する構成としてもよい。

【００６１】

このようにシステム発話タイミング検出手段により、システム状態を示す情報であるシ

ステム発話意欲度を用いてユーザ発話権終了判定用閾値を動的に調整する構成とした場合には、システム発話意欲度が強いときには、ユーザ発話権が終了したという識別結果が出やすくなる設定状態とし、システム発話意欲度が弱いときには、ユーザ発話権が終了したという識別結果が出にくい設定状態とすることが可能となる。

【００６２】

＜音声認識処理の結果が新たに出力されたときに、その音声認識処理の結果を用いて、次発話候補の入替が可能な構成＞

【００６３】

さらに、前述した次発話準備手段により準備した複数の次発話候補の内容データの中から、次発話選択手段により次発話の内容データを選択する構成とした場合において、

次発話準備手段は、

音声認識処理手段によるユーザ発話の音声認識処理の結果が新たに出力された場合には、新たに出力された当該音声認識処理の結果を用いて、次発話の候補となる複数の次発話候補の内容データの少なくとも一部を入れ替えるか否かを判定し、入れ替えると判定した場合には、次発話の候補となる別の複数の次発話候補の内容データを取得または生成する準備処理を実行する構成としてもよい。

【００６４】

このように音声認識処理の結果が新たに出力されたときに、その音声認識処理の結果を用いて、次発話候補の入替が可能な構成とした場合には、進行中のユーザ発話の内容に応じて、既に準備されている複数の次発話候補の内容データの入替を行うことが可能となるので、ユーザ発話の内容に応じた適切な次発話候補の内容データを準備することが可能となる。

【００６５】

＜音声認識処理の結果が新たに出力されたときに、この結果に含まれる重要度の高い単語を用いてユーザの関心のある話題を決定し、決定した話題に従って次発話候補の入替を行う構成＞

【００６６】

また、上述した音声認識処理の結果が新たに出力されたときに、その音声認識処理の結果を用いて、次発話候補の入替が可能な構成とした場合において、

次発話準備手段は、

新たに出力された音声認識処理の結果を用いて、この結果に含まれる単語のうち予め定められた重要度の高い単語を用いて、ユーザの関心のある話題を決定し、題材データ記憶手段に記憶された題材データまたは外部システムに記憶された題材データの中から、決定した話題に関連付けられて記憶されている題材データを選択し、次発話の候補となる別の複数の次発話候補の内容データを取得または生成する準備処理を実行する構成としてもよい。

【００６７】

このように音声認識処理の結果が新たに出力されたときに、この結果に含まれる重要度の高い単語を用いてユーザの関心のある話題を決定し、決定した話題に従って次発話候補の入替を行う構成とした場合には、進行中のユーザ発話の内容に応じて、既に準備されている複数の次発話候補の内容データの入替を行い、次発話により提示する話題を変更することが可能となる。

【００６８】

＜システム発話タイミング検出手段により、衝突の発生情報やシステムの交替潜時を用いて、ユーザ発話権終了判定用閾値を調整する構成＞

【００６９】

また、以上に述べた対話システムにおいて、

発話生成手段は、

音声信号取得手段により取得したユーザ発話の音声信号と、再生中のシステム発話の音声信号との衝突の発生を検出し、検出した衝突の発生情報を、ユーザ識別情報と関連付け

てユーザ情報記憶手段に記憶させるとともに、ユーザ発話の終了からシステム発話の開始までの交替潜時を計測し、計測した交替潜時を、ユーザ識別情報と関連付けてユーザ情報記憶手段に記憶させる処理も実行する構成とされ、
システム発話タイミング検出手段は、
ユーザ情報記憶手段に記憶されている音声対話相手のユーザとの衝突の発生情報を取得して当該ユーザとの衝突の発生頻度または累積発生回数を算出し、算出した衝突の発生頻度または累積発生回数が上方調整用閾値以上または超過の場合には、ユーザ発話権の維持または終了を識別するパターン認識処理の結果として得られる尤度に対して設定されているユーザ発話権終了判定用閾値を標準値または前回調整値よりも高く設定し、
ユーザ情報記憶手段に記憶されている音声対話相手のユーザについてのユーザ発話の終了からシステム発話の開始までの複数の交替潜時を取得して当該ユーザについての交替潜時の長短の傾向を示す平均値若しくはその他の指標値を算出し、算出した交替潜時の指標値が下方調整用閾値以上または超過の場合には、ユーザ発話権終了判定用閾値を標準値または前回調整値よりも低く設定する処理も実行する構成としてもよい。

【００７０】
ここで、「システム発話タイミング検出手段」における「標準値または前回調整値」は、別の情報に基づく別の趣旨での閾値調整が別途に行われている場合には、その別途の閾値調整後の値を指す。

【００７１】
また、ここでの「衝突」は、ユーザ発話権が終了したという識別結果が出て、システム発話を開始したところ、実際にはユーザ発話権が維持されていて、両者の発話が重なった場合の衝突である。従って、ユーザ発話権が終了したものの、システム発話の開始が遅れたために、再び、ユーザ発話が開始されてしまい、ほぼ同時に両者の発話が開始されて重なった場合の衝突ではない。

【００７２】
さらに、ここでの「交替潜時」は、ユーザ発話の終了からシステム発話の開始までの間（ま）であり、システムの交替潜時である。従って、「当該ユーザについての交替潜時」とされているが、これは、当該ユーザとの音声対話を行うときのシステムの交替潜時のことであり、システム発話の終了からユーザ発話の開始までの間（ま）のことではない。

【００７３】
このようにシステム発話タイミング検出手段により、衝突の発生情報やシステムの交替潜時を用いて、ユーザ発話権終了判定用閾値を調整する構成とした場合には、各ユーザについて、衝突の発生が起きる傾向にあるときには、ユーザ発話権が終了したという識別結果が出にくい設定状態とし、システムの交替潜時が長い傾向にあるときには、ユーザ発話権が終了したという識別結果が出やすくなる設定状態とすることが可能となる。このため、ユーザ属性に応じたユーザ発話権終了判定用閾値の調整を実現することができる。

【００７４】
＜ユーザ発話権終了判定用閾値を下方調整することを決めるための下方調整用閾値を、ユーザの発話速度の関数とする構成＞

【００７５】
さらに、上述したシステム発話タイミング検出手段により、衝突の発生情報やシステムの交替潜時を用いて、ユーザ発話権終了判定用閾値を調整する構成とした場合において、
発話生成手段は、
音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いて発話速度を算出し、算出した発話速度を、ユーザ識別情報と関連付けてユーザ情報記憶手段に記憶させる処理も実行する構成とされ、
システム発話タイミング検出手段は、
ユーザ情報記憶手段に記憶されている音声対話相手のユーザについてのユーザ発話の終了からシステム発話の開始までの複数の交替潜時を取得して当該ユーザについての交替潜時の長短の傾向を示す平均値若しくはその他の指標値を算出し、算出した交替潜時の指標

値が下方調整用閾値以上または超過の場合に、ユーザ発話権終了判定用閾値を標準値または前回調整値よりも低く設定する処理を実行する際に、
ユーザ情報記憶手段に記憶されている音声対話相手の複数の発話速度を取得して当該ユーザの発話速度の傾向を示す平均値若しくはその他の指標値を算出し、下方調整用閾値を、算出した発話速度の指標値を用いて、発話速度の指標値が大きいときには当該下方調整用閾値が小さくなり、発話速度の指標値が小さいときには当該下方調整用閾値が大きくなるように予め定められた関数により設定する処理を実行する構成としてもよい。
【００７６】
このようにユーザ発話権終了判定用閾値を下方調整することを決めるための下方調整用閾値を、ユーザの発話速度の関数とする構成とした場合には、各ユーザの発話速度の傾向に応じ、下方調整用閾値の設定を変更することが可能となる。このため、ユーザ属性に応じたユーザ発話権終了判定用閾値の下方調整を実現することができる。すなわち、システムの交替潜時が長い傾向にあるときには、ユーザ発話権終了判定用閾値を下方調整することにより、ユーザ発話権が終了したという識別結果が出やすくなる設定状態とし、システムの交替潜時が短くなるようにすることができるが、この際、システムの交替潜時が長い傾向にあるか否かは、ユーザ毎に異なり、各ユーザの発話速度の傾向と関係するので、下方調整用閾値をユーザの発話速度の関数とすることで、ユーザ属性に応じてユーザ発話権終了判定用閾値の下方調整を行うか否かを決めることができる。
【００７７】
＜ユーザ発話の音声信号から抽出した音響特徴量、およびリアルタイムのユーザの発話速度を用いて、ユーザ発話権の維持または終了を識別するパターン認識処理を行う構成＞
【００７８】
また、以上に述べた対話システムにおいて、
ユーザのリアルタイムの発話速度を含むユーザ状態を示す情報を記憶するユーザ状態記憶手段を備え、
発話生成手段は、
音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報を用いてリアルタイムの発話速度を算出し、算出したリアルタイムの発話速度をユーザ状態記憶手段に記憶させる処理も実行する構成とされ、
システム発話タイミング検出手段は、
音声信号取得手段により取得したユーザ発話の音声信号から音響特徴量を抽出し、抽出した音響特徴量およびユーザ状態記憶手段に記憶されているリアルタイムの発話速度を用いるか、または、これらの音響特徴量およびリアルタイムの発話速度に加え、音声認識処理手段による音声認識処理の結果として得られたユーザ発話の言語情報から抽出した言語特徴量を用いて、音声認識処理手段による音声認識処理の実行タイミングに依拠しない周期で、ユーザ発話権の維持または終了を識別するパターン認識処理を繰り返し実行し、このパターン認識処理の結果を用いて、システム発話の開始タイミングを検出する処理を実行する構成としてもよい。
【００７９】
ここで、「リアルタイムの発話速度」における「リアルタイムの」という意味は、事後的に計算するのではなく、その場で計算されるという意味であり、「逐次得られる最新の」という意味である。発話速度の計算には、音声認識処理の結果が用いられるが、音声認識処理の結果は、若干の時間遅れで得られるため、厳密に言えば、ここでいう「リアルタイム」には、「略リアルタイム」が含まれる。
【００８０】
このようにユーザ発話の音声信号から抽出した音響特徴量、およびリアルタイムのユーザの発話速度を用いて、ユーザ発話権の維持または終了を識別するパターン認識処理を行う構成とした場合には、システム発話タイミング検出手段の識別器は、ユーザ発話の各時点における発話速度（蓄積した発話速度から得られるユーザ属性としての発話速度の傾向ではなく、瞬間的な発話速度という意味）を用いた学習を行うことにより構築されるので

、その時々のユーザの発話速度を加味した識別結果を得ることが可能となる。このため、ユーザ毎に異なる発話速度の傾向（蓄積した発話速度から得られるユーザ属性）に応じてユーザ発話権終了判定用閾値を調整する必要がなくなる。なお、閾値調整と併用してもよく、その場合には、閾値調整量が少なくなる。

【００８１】

＜プログラムの発明＞

【００８２】

本発明のプログラムは、以上に述べた対話システムとして、コンピュータを機能させるためのものである。

【００８３】

なお、上記のプログラムまたはその一部は、例えば、光磁気ディスク（ＭＯ）、コンパクトディスク（ＣＤ）、デジタル・バーサタイル・ディスク（ＤＶＤ）、フレキシブルディスク（ＦＤ）、磁気テープ、読出し専用メモリ（ＲＯＭ）、電気的消去および書換可能な読出し専用メモリ（ＥＥＰＲＯＭ）、フラッシュ・メモリ、ランダム・アクセス・メモリ（ＲＡＭ）、ハードディスクドライブ（ＨＤＤ）、ソリッドステートドライブ（ＳＳＤ）、フラッシュディスク等の記録媒体に記録して保存や流通等させることが可能であるとともに、例えば、ローカル・エリア・ネットワーク（ＬＡＮ）、メトロポリタン・エリア・ネットワーク（ＭＡＮ）、ワイド・エリア・ネットワーク（ＷＡＮ）、インターネット、イントラネット、エクストラネット等の有線ネットワーク、あるいは無線通信ネットワーク、さらにはこれらの組合せ等の伝送媒体を用いて伝送することが可能であり、また、搬送波に載せて搬送することも可能である。さらに、上記のプログラムは、他のプログラムの一部分であってもよく、あるいは別個のプログラムと共に記録媒体に記録されていてもよい。

【発明の効果】

【００８４】

以上に述べたように本発明によれば、システム発話タイミング検出手段により、ユーザが自己の発話権を維持しているか、または、譲渡若しくは放棄により終了させたかをパターン認識処理により逐次推定するとともに、次発話準備手段により、システム発話タイミング検出手段によるパターン認識処理とは非同期で、かつ、システム発話タイミング検出手段によりシステム発話の開始タイミングが検出される前に、ユーザ発話に対するシステムの次発話の内容データを準備するので、システムの応答性を向上させることができ、衝突の発生を回避または抑制しつつ、不要に長いシステムの交替潜時の発生を回避または抑制することができるという効果がある。

【図面の簡単な説明】

【００８５】

【図１】本発明の一実施形態の対話システムの全体構成図。

【図２】前記実施形態のシステム発話タイミング検出手段の詳細構成図。

【図３】前記実施形態の次発話選択用情報生成手段の詳細構成図。

【図４】前記実施形態の次発話準備手段の詳細構成図。

【図５】前記実施形態のユーザからシステムへの話者交替時の処理の流れを示すフローチャートの図。

【図６】前記実施形態のシステム発話、ユーザ発話、各処理の時間的な前後関係を示す説明図。

【図７】前記実施形態の各処理のタイミングとその結果との関係を示す説明図。

【図８】前記実施形態のシステム発話タイミング検出手段によるシステム発話の開始タイミングの判断処理のロジックを示すブロック図。

【図９】前記実施形態のシステム発話タイミング検出手段によるユーザ発話権終了判定用閾値のリアルタイム調整（その１）の説明図。

【図１０】前記実施形態のシステム発話タイミング検出手段によるユーザ発話権終了判定用閾値のリアルタイム調整（その２）の説明図。

【図１１】前記実施形態のシステム発話タイミング検出手段によるユーザ発話権終了判定用閾値の事前調整（その１）の説明図。
【図１２】前記実施形態のシステム発話タイミング検出手段によるユーザ発話権終了判定用閾値の事前調整（その２）の説明図。
【図１３】前記実施形態のシナリオのデータ構成の具体例を示す図。
【図１４】前記実施形態のシナリオ再生（システム発話）とユーザの反応（ユーザ発話）との関係を示す説明図。
【図１５】前記実施形態の次発話候補の準備処理の具体例（１）を示す図。
【図１６】前記実施形態の次発話候補の準備処理の具体例（２）を示す図。
【図１７】前記実施形態の次発話候補の準備処理の具体例（３）を示す図。
【発明を実施するための形態】
【００８６】
以下に本発明の一実施形態について図面を参照して説明する。図１には、本実施形態の対話システム１０の全体構成が示されている。図２には、システム発話タイミング検出手段２２の詳細構成が示され、図３には、次発話選択用情報生成手段２３の詳細構成が示され、図４には、次発話準備手段４３の詳細構成が示されている。また、図５には、ユーザからシステムへの話者交替時の処理の流れがフローチャートで示され、図６には、システム発話、ユーザ発話、各処理の時間的な前後関係が示され、図７には、各処理のタイミングとその結果との関係が示され、図８には、システム発話タイミング検出手段２２によるシステム発話の開始タイミングの判断処理のロジックが示されている。さらに、図９～図１２は、システム発話タイミング検出手段２２によるユーザ発話権終了判定用閾値の調整の説明図である。また、図１３には、シナリオのデータ構成の具体例が示され、図１４には、シナリオ再生（システム発話）とユーザの反応（ユーザ発話）との関係の具体例が示され、図１５～図１７には、次発話候補の準備処理の具体例が示されている。
【００８７】
＜対話システム１０の全体構成＞
【００８８】
図１において、対話システム１０は、ユーザとの音声対話を行うシステムであり、１台または複数台のコンピュータにより構成され、本実施形態では、一例として、再生装置２０と、対話サーバ４０とをネットワーク１で接続した構成とされている。また、ネットワーク１には、外部システムである題材データ提供システム６０も接続されている。
【００８９】
ここで、ネットワーク１は、主としてインターネットのような外部ネットワークであるが、これとイントラネットやＬＡＮ等の内部ネットワークとの組合せ等でもよく、有線であるか、無線であるか、有線・無線の混在型であるかは問わない。また、ネットワーク１は、例えば、社内、工場内、事業所内、グループ企業内、学内、病院内、マンション内、建物内、公園・遊園地・動物園・博物館・美術館・博覧会場等の施設内、所定の地域内等に限定されたイントラネットやＬＡＮ等の内部ネットワークであってもよい。
【００９０】
再生装置２０は、１台または複数台のコンピュータにより構成され、音声信号取得手段２１と、システム発話タイミング検出手段２２と、次発話選択用情報生成手段２３と、次発話選択手段２４と、発話生成手段２５と、次発話候補記憶手段３０と、システム状態記憶手段３１と、ユーザ状態記憶手段３２とを備えている。この再生装置２０は、例えば、スマートフォン、タブレット、モバイルＰＣ（パーソナル・コンピュータ）等の携帯機器であってもよい。また、汎用機器ではなく、音声対話の専用機器としてもよい。
【００９１】
このうち、音声信号取得手段２１（但し、マイクロフォンの部分を除く。）、システム発話タイミング検出手段２２（但し、ユーザ発話権終了判定モデル記憶手段２２Ｅ（図２参照）の部分を除く。）、次発話選択用情報生成手段２３（但し、第１、第２の発話意図識別モデル記憶手段２３Ｄ，２３Ｇ（図３参照）の部分を除く。）、次発話選択手段２４

、および発話生成手段２５（但し、スピーカやディスプレイの部分を除く。）は、再生装置２０を構成するコンピュータ本体の内部に設けられた中央演算処理装置（ＣＰＵ）、およびこのＣＰＵの動作手順を規定する１つまたは複数のプログラムにより実現される。また、次発話候補記憶手段３０、システム状態記憶手段３１、ユーザ状態記憶手段３２、システム発話タイミング検出手段２２を構成するユーザ発話権終了判定モデル記憶手段２２Ｅ（図２参照）、および次発話選択用情報生成手段２３を構成する第１、第２の発話意図識別モデル記憶手段２３Ｄ，２３Ｇ（図３参照）としては、例えば、ハードディスクドライブ（ＨＤＤ）、ソリッドステートドライブ（ＳＳＤ）等を採用することができる。

【００９２】

対話サーバ４０は、１台または複数台のコンピュータにより構成され、音声認識処理手段４１と、対話状態管理手段４２と、次発話準備手段４３と、対話履歴記憶手段５０と、題材データ記憶手段５１と、ユーザ情報記憶手段５２とを備えている。

【００９３】

このうち、音声認識処理手段４１、対話状態管理手段４２、および次発話準備手段４３（但し、先行次発話候補情報記憶手段４３Ｄ（図４参照）の部分は除く。）は、対話サーバ４０を構成するコンピュータ本体の内部に設けられた中央演算処理装置（ＣＰＵ）、およびこのＣＰＵの動作手順を規定する１つまたは複数のプログラムにより実現される。また、対話履歴記憶手段５０、題材データ記憶手段５１、ユーザ情報記憶手段５２、および次発話準備手段４３を構成する先行次発話候補情報記憶手段４３Ｄとしては、例えば、ハードディスクドライブ（ＨＤＤ）、ソリッドステートドライブ（ＳＳＤ）等を採用することができる。なお、先行次発話候補情報記憶手段４３Ｄ（図４参照）は、主メモリ等の揮発性メモリとしてもよい。

【００９４】

題材データ提供システム６０は、外部システムであり、１台または複数台のコンピュータにより構成され、対話サーバ４０を構成する題材データ記憶手段５１に相当する外部題材データ記憶手段（不図示）を備えている。

【００９５】

なお、本実施形態では、図１に示すように、対話システム１０は、再生装置２０と、対話サーバ４０とをネットワーク１で接続した構成とされているが、本発明の対話システムは、スタンドアローンのシステムとしてもよい。また、図１に示したネットワーク構成は、一例に過ぎないので、ネットワーク構成とする場合でも、各機能の分散形態として、図１の状態とは異なる様々な形態を採用することができる。

【００９６】

例えば、再生装置２０は、音声対話相手であるユーザと音声によるやりとりを行うので、ユーザの近く（音声の届く範囲）に配置する必要があることから、これを本体と端末とに分割して無線または有線で通信を行うようにし、端末をユーザの近くに配置する一方、本体をユーザから比較的離れた位置（音声が届かない位置でもよい）に配置する構成とすることができる。この場合、例えば、端末は、再生装置２０を構成する音声信号取得手段２１またはその一部であるマイクロフォンの部分と、再生装置２０を構成する発話生成手段２５またはその一部であるスピーカの部分（映像や静止画像の再生を伴う場合には、ディスプレイの部分を含む。）とにより構成することができる。そして、例えば、本体を固定設置された機器とし、端末を移動機器とすること等ができるが、本体、端末のいずれについても、固定機器でも移動機器でもよい。また、本体と端末との個数の関係は、１対１でも、１対多でもよい。さらに、例えば、対話種別（ニュース対話、ガイダンス対話、アンケート対話、情報検索対話、操作対話、教育対話等の別）に対応させて異なる本体を設置する場合、新旧異なるタイプの本体を併用する場合、機能の異なる本体を使い分ける場合等には、本体と端末との個数の関係は、多対１、多対多でもよく、この場合には、任意の１つの端末と、複数の本体から目的に応じて選択された１つの本体とが接続されることになる。また、ユーザとの関係では、１つの端末は、同時使用でなければ、複数のユーザが交代して使用することができる。本体は、複数の端末と同時接続可能な構成とすれば、

複数のユーザの同時使用に対応可能な構成とすることができるが、複数のユーザの同時使用を許容しない構成としてもよい。

【００９７】

また、再生装置２０を構成するシステム発話タイミング検出手段２２と、次発話選択用情報生成手段２３と、次発話選択手段２４と、次発話候補記憶手段３０と、システム状態記憶手段３１と、ユーザ状態記憶手段３２とは、それぞれ別々のコンピュータに設けてもよく、適宜組み合わせて同じコンピュータに設けてもよい。

【００９８】

さらに、対話サーバ４０も同様であり、対話サーバ４０を構成する各機能の部分は、それぞれ別々のコンピュータに設けてもよく、適宜組み合わせて同じコンピュータに設けてもよい。また、再生装置２０を構成する１つまたは複数の機能の部分と、対話サーバ４０を構成する１つまたは複数の機能の部分とを適宜組み合わせて同じコンピュータに設けてもよい。

【００９９】

＜再生装置２０／音声信号取得手段２１の構成＞

【０１００】

音声信号取得手段２１は、ユーザ発話の音声信号を取得するものであり、音（ここでは、音声）をアナログの電気信号に変換する機器であるマイクロフォン、Ａ／Ｄ変換手段、Ａ／Ｄ変換で得られたデジタルの音声信号を保持する音声信号記憶手段、音声信号を各所に送信する送信手段等を含んで構成されている。

【０１０１】

＜再生装置２０／システム発話タイミング検出手段２２の構成＞

【０１０２】

図２において、システム発話タイミング検出手段２２は、音響特徴量抽出手段２２Ａと、言語特徴量抽出手段２２Ｂと、ユーザ発話権終了判定用パターン認識器２２Ｃと、システム発話開始タイミング判断手段２２Ｆと、ユーザ発話権終了判定用閾値調整手段２２Ｇとを含んで構成されている。このうち、音響特徴量抽出手段２２Ａと、言語特徴量抽出手段２２Ｂと、ユーザ発話権終了判定用パターン認識器２２Ｃとについては、例えば、前述した非特許文献１，２に記載された技術を採用することができる。

【０１０３】

このシステム発話タイミング検出手段２２による処理は、音声認識処理手段４１による音声認識処理の実行タイミングに依拠しない周期で、すなわち音声認識処理とは非同期で実行される。具体的には、例えば、１０ｍｓ（ミリ秒）～１００ｍｓという短い周期で実行される。図６の最下部に示した処理の周期Ｑ（時間間隔）である。なお、音声区間（ＩＰＵ）を形成する際のポーズは、通常は１００ｍｓ以上であるから、周期Ｑは、そのＩＰＵ形成用の閾値よりも短いか、同等の周期ということになる。

【０１０４】

＜再生装置２０／システム発話タイミング検出手段２２／音響特徴量抽出手段２２Ａの構成＞

【０１０５】

音響特徴量抽出手段２２Ａは、音声信号取得手段２１により取得したユーザ発話の音声信号から音響特徴量を抽出する処理を実行するものである。前述した非特許文献１，２に記載された技術を採用する場合には、狭帯域スペクトログラムを符号化、復号化する自己符号化器（オートエンコーダ）をニューラルネットワーク（ＣＮＮ）により構築し、その中間層の出力を音響特徴量（ボトルネック特徴量）とする。具体的には、周波数分析により例えば１０ｍｓ毎に得られる２５６点のパワースペクトルを１０本分並べたものを入力とし、中間層の出力２５６次元を特徴量とする。すなわち、ＣＮＮオートエンコーダの入力は、例えば、フレームサイズ＝８００サンプル（５０ｍｓ）、フレームシフト＝１６０サンプル（１０ｍｓ）で切り出した音声信号から生成したスペクトログラムを時系列に１０本分（図６の下部に示したＲ本分）並べたものとし、そのサイズを１０×２５６次元と

する。そして、この入力サイズを２５６次元に圧縮し、音響特徴量とする。
【０１０６】
このようにして音響特徴量を抽出する場合、２５６点のパワースペクトルを１０本分並べた入力データを作成する際に、２５６点のパワースペクトルを１本ずつずらしていけば、図６の最下部に示した処理の周期Ｑ（時間間隔）は、周波数分析のフレームシフト＝１０ｍｓと同じになり、２本ずつずらしていけば、２倍の２０ｍｓとなり、同様に１０本ずつずらしていけば、１０倍の１００ｍｓとなる。従って、２５６点のパワースペクトルを用いる際に、ずらす本数を変えることにより、処理の周期Ｑを変更することができる。なお、ずらす本数を多くすることにより、使用する音声信号の区間に重なりがないようにしてもよいが、使用されない音声信号の区間が生じることは避ける必要がある。なお、図７では、周波数分析のタイミングは、各フレームの終点を指し、フレーム毎に図示されているので、時間間隔はフレームシフトを示しているのに対し、システム発話タイミング検出手段２２による処理の周期Ｑは、それよりも広い時間間隔で図示されているので、パワースペクトルを数本ずつ（例えば３本ずつ）ずらして用いていることを示している。また、使用するパワースペクトルの本数は、Ｒ本＝１０本である必要はなく、任意である。さらに、フレームサイズ＝５０ｍｓ、フレームシフト＝１０ｍｓという数値も一例に過ぎず、これらの数値に限定されるものではない。
【０１０７】
また、音声信号からの音響特徴量の抽出処理は、上述した非特許文献１，２に記載された技術による抽出処理に限定されるものではなく、ユーザ発話権終了判定用パターン認識器２２Ｃの入力に用いる音響特徴量は、ユーザ発話の音声信号から得られる音響特徴量であれば、いずれの特徴量でもよい。
【０１０８】
例えば、音響特徴量は、基本周波数（Ｆ０）や、メル周波数ケプストラム係数（ＭＦＣＣ）等でもよい。但し、特徴量を計算すること自体に遅延が生じないことが好ましい。なお、ＭＦＣＣ等の音響特徴量を用いると、処理遅れは無くなるが、韻律情報が失われるという欠点がある。
【０１０９】
＜再生装置２０／システム発話タイミング検出手段２２／言語特徴量抽出手段２２Ｂの構成＞
【０１１０】
言語特徴量抽出手段２２Ｂは、音声認識処理手段４１による音声認識処理の結果として得られたユーザ発話の言語情報から言語特徴量を抽出する処理を実行するものである。この言語特徴量抽出手段２２Ｂの設置は省略してもよい。
【０１１１】
具体的には、例えば、ＬＳＴＭ言語モデルの中間出力（５１２次元）を言語特徴量とすることができる（非特許文献２参照）。なお、ＬＳＴＭ（Long short-term memory）は、リカレントニューラルネットワーク（ＲＮＮ）の一種である。
【０１１２】
また、音声認識処理手段４１による音声認識処理は、上述した音響特徴量抽出手段２２Ａの処理と非同期で実行されるため、音響特徴量抽出手段２２Ａにより音響特徴量が抽出されたときに、この言語特徴量抽出手段２２Ｂによる言語特徴量の抽出が行われていない場合があるので、その場合には、言語特徴量はゼロベクトルとする。
【０１１３】
＜再生装置２０／システム発話タイミング検出手段２２／ユーザ発話権終了判定用パターン認識器２２Ｃの構成＞
【０１１４】
ユーザ発話権終了判定用パターン認識器２２Ｃは、音響特徴量抽出手段２２Ａにより抽出した音響特徴量を入力とするか（非特許文献１参照）、あるいは、この音響特徴量および言語特徴量抽出手段２２Ｂにより抽出した言語特徴量を入力とし（非特許文献２参照）

、ユーザが発話する地位または立場を有していることを示すユーザ発話権の維持または終了（終了には、譲渡、放棄が含まれる。）を識別するパターン認識処理を繰り返し実行するものである。

【０１１５】

このユーザ発話権終了判定用パターン認識器２２Ｃは、識別アルゴリズムによるパターン認識処理を実行するユーザ発話権終了判定用パターン認識処理手段２２Ｄと、このパターン認識処理で用いるモデル（パラメータ）を記憶するユーザ発話権終了判定モデル記憶手段２２Ｅとにより構成されている。

【０１１６】

具体的には、例えば、音響特徴量（２５６次元）および言語特徴量（５１２次元）を入力とし、ユーザ発話権の維持または終了を逐次推定するモデルをニューラルネットワークにより構築し、ユーザ発話権終了判定用パターン認識器２２Ｃとすることができる（非特許文献２参照）。この際、ニューラルネットワークには、時系列情報を考慮するため、ＬＳＴＭ（ＲＮＮの一種）を用いることができる（非特許文献１参照）。

【０１１７】

このユーザ発話権終了判定用パターン認識器２２Ｃは、ユーザ発話権が終了したことの確からしさを示す尤度を出力するので、その尤度が、予め定められたユーザ発話権終了判定用閾値（但し、この閾値は、ユーザ発話権終了判定用閾値調整手段２２Ｇにより、事前に、またはリアルタイムで動的に調整されることがある。）以上であるか、またはこの閾値を超えているかを判定する閾値処理を行い（図７の最下部を参照）、ユーザ発話権終了判定用閾値以上または超過と判定した場合には、ユーザ発話権が終了したという識別結果を出力し、ユーザ発話権終了判定用閾値未満または以下と判定した場合には、ユーザ発話権が維持されているという識別結果を出力する。

【０１１８】

この際、尤度は、ユーザ発話権が終了したことの確からしさを示す尤度としているので、尤度の値が大きい程（１に近い程）、ユーザ発話権の終了の状態に近く、尤度の値が小さい程（０に近い程）、ユーザ発話権の維持の状態に近い（図７の最下部を参照）。従って、尤度がユーザ発話権終了判定用閾値以上になるか、超えれば、ユーザ発話権が終了したという識別結果が出力されることになるので、ユーザ発話権終了判定用閾値の上方調整というのは、ユーザ発話権が終了したという識別結果が出にくくなる方向への調整であり、下方調整というのは、ユーザ発話権が終了したという識別結果が出やすくなる方向への調整である。本願の請求項は、この場合の記載とされている。

【０１１９】

一方、尤度は、ユーザ発話権が維持されていることの確からしさを示す尤度としてもよく、この場合には、尤度の値が大きい程（１に近い程）、ユーザ発話権の維持の状態に近く、尤度の値が小さい程（０に近い程）、ユーザ発話権の終了の状態に近い。従って、尤度がユーザ発話権終了判定用閾値以下になるか、未満になれば、ユーザ発話権が終了したという識別結果が出力されることになるので、ユーザ発話権終了判定用閾値の上方調整というのは、ユーザ発話権が終了したという識別結果が出やすくなる方向への調整であり、下方調整というのは、ユーザ発話権が終了したという識別結果が出にくくなる方向への調整である。このため、本願の請求項は、この場合とは逆の記載とされているが（上方、下方が逆の表現となっているが）、両者は等価なことであり、また、１から尤度の値を減じれば、逆の意味の尤度になるので、本願の請求項は、いずれの場合も含むものである。

【０１２０】

また、閾値処理を行う際には、フィルタをかけた後の尤度（出力される直近の幾つかの尤度を用いて平準化した後の尤度）を用いてもよい。

【０１２１】

さらに、ユーザ発話権終了判定用パターン認識器２２Ｃは、ユーザ発話権の維持または終了を識別するパターン認識処理を実行する際に、終了については、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図のうちの

いずれのユーザ発話意図で終了するのかを識別するパターン認識処理を実行してもよい。この場合、識別器の学習段階で、終了については、質問、応答、相槌等のユーザ発話意図を含むラベル（タグ）を付すことになる。すなわち、「維持」、「質問で終了」、「応答で終了」、「相槌で終了」等のタグ付けを行った学習用データをそれぞれ多数用意して学習を行い、３クラス識別以上のユーザ発話権終了判定モデルを構築する。そして、運用段階では、ユーザ発話権が「維持」されていることの確からしさを示す尤度、「質問で終了」したことの確からしさを示す尤度、「応答で終了」したことの確からしさを示す尤度、「相槌で終了」したことの確からしさを示す尤度等のように、いずれのユーザ発話意図で終了したのかを示す情報が出力される。例えば、質問で終了の尤度＝０．９０、応答で終了の尤度＝０．０４、相槌で終了の尤度＝０．０３等のように出力される。従って、ユーザ発話権終了判定用閾値は、質問、応答、相槌等のユーザ発話意図毎に設定し、ユーザ発話意図毎に閾値処理を行う。但し、ユーザ発話意図毎に設定したユーザ発話権終了判定用閾値が、同じ値となってもよい。このようにした場合、次発話選択手段２４は、ユーザ発話権終了判定用パターン認識器２２Ｃによるユーザ発話意図の識別結果（いずれのユーザ発話意図で終了したのかという情報）を用いて、次発話準備手段４３による準備処理で得られた複数の次発話候補の内容データの中から、発話生成手段２５で用いる次発話の内容データを選択することができる。そして、以上のように質問、応答、相槌等のユーザ発話意図を含むタグ付けをした学習で得られたユーザ発話権終了判定用パターン認識器２２Ｃは、次発話選択用情報生成手段２３のユーザ発話意図の識別器（図３に示す第１、第２の発話意図識別器２３Ｂ，２３Ｅ）と、いずれのユーザ発話意図で終了したのかというタグ付けをしないで単純に維持・終了を識別するための学習で得られた識別器とを、マルチタスクでまとめて一体化させたユーザ発話権終了判定用パターン認識器２２Ｃとは、異なるものである。

【０１２２】

なお、ユーザ発話の音声信号を逐次処理して短い周期で識別を行う技術として、スマートスピーカのウェイクワードのスポッティングが挙げられるが、ユーザ発話権終了判定用パターン認識器２２Ｃは、特定の語のスポッティングではなく、ユーザ発話権の維持または終了を、その発話内容に依らずに検出することを目的とする点で異なる。

【０１２３】

＜再生装置２０／システム発話タイミング検出手段２２／システム発話開始タイミング判断手段２２Ｆの構成＞

【０１２４】

システム発話開始タイミング判断手段２２Ｆは、ユーザ発話権終了判定用パターン認識器２２Ｃによるパターン認識処理の結果（維持または終了の識別結果）を用いるか、またはこのパターン認識処理の結果に加え、システム状態記憶手段３１に記憶されているシステム状態を示す情報（準備完了・準備中の別）や、ユーザ状態記憶手段３２に記憶されているユーザ状態を示す情報（ユーザ発話継続時間）を用いて、システム発話の開始タイミングを検出する処理を実行するものである。

【０１２５】

具体的には、図８に示すように、システム発話開始タイミング判断手段２２Ｆは、先ず、ユーザ発話権終了判定用パターン認識器２２Ｃによるパターン認識処理の結果が、ユーザ発話権の維持を示している場合（Ｐ１）と、終了を示している場合（Ｐ２）とに判断分岐する。

【０１２６】

次に、維持を示している場合（Ｐ１）には、システム発話の開始タイミングではないと判断する（Ｐ７）。

【０１２７】

但し、図８中および図２中の二点鎖線で示すように、ユーザ状態記憶手段３２に記憶されているユーザ発話継続時間が、予め定められた長時間判定用閾値以上または超過の場合には、パターン認識処理の結果がユーザ発話権の維持を示していても（Ｐ１）、システム

発話の開始タイミングであると判断する処理（Ｐ８）を行ってもよい。
【０１２８】
一方、終了を示している場合（Ｐ２）において、システム状態記憶手段３１に記憶されているシステム状態を示す情報が準備完了（ステータス＝「準備完了」）を示している場合（Ｐ３）には、システム発話の開始タイミングであると判断する（Ｐ８）。
【０１２９】
但し、図８中および図２中の二点鎖線で示すように、ユーザ状態記憶手段３２に記憶されているユーザ発話継続時間が、予め定められた短時間判定用閾値以下または未満の場合には、パターン認識処理の結果がユーザ発話権の終了を示していても（Ｐ２）、システム発話の開始タイミングではないと判断する処理（Ｐ７）を行ってもよい。
【０１３０】
また、終了を示している場合（Ｐ２）において、システム状態記憶手段３１に記憶されているシステム状態を示す情報が準備中を示している場合（Ｐ４）には、その準備中を示すステータスに応じ（次発話準備手段４３による準備中の処理内容）に応じ、判断を分岐させる。
【０１３１】
そして、準備中を示している場合（Ｐ４）において、その準備中の処理内容が、直ぐに完了する処理内容として予め分類されている処理の準備中である場合（Ｐ５）には、準備完了になるまで待ってシステム発話の開始タイミングであると判断するために（但し、結果的に、直ぐに準備完了にならない場合もある。）、その時点では、システム発話の開始タイミングではないと判断する（Ｐ９）。直ぐに完了する処理内容として予め分類されている処理の準備中とは、例えば、ステータス＝「自サーバ検索中」等である。
【０１３２】
一方、準備中を示している場合（Ｐ４）において、その準備中の処理内容が、直ぐに完了しない処理内容として予め分類されている処理の準備中である場合（Ｐ６）には、システム発話の開始タイミングであると判断するとともに、フィラーの挿入タイミングである旨の情報を出力する（Ｐ１０）。直ぐに完了しない処理内容として予め分類されている処理の準備中とは、例えば、ステータス＝「外部システムアクセス中」、「音声合成処理中」等である。フィラーの挿入タイミングである旨の情報には、どのような種別のフィラーを挿入するかの情報を含めてもよく、この場合、準備中のステータスの種別と、フィラーの種別との対応関係を、予め定めておけばよい。例えば、直ぐに完了しない処理内容にも、その程度があるので、かなり長時間の準備を要する場合には、「ちょっと待ってね、今、調べてるから。」、「少々お待ちください、処理中です。」等のフィラーを挿入することができ、そこまで長時間を要しない場合には、「えー。」、「あのね。」等のフィラーを挿入することができる。
【０１３３】
なお、準備中を示すステータスのうち、どのようなステータスが、直ぐに完了する処理内容なのか、直ぐに完了しない処理内容なのかは、システムの構築、運用、管理を行う者が適宜設計すればよく、対話の種別（ニュース対話、アンケート対話、情報検索対話、操作対話、教育対話等の別）に応じて定めてもよい。
【０１３４】
また、図８において、Ｐ９の下流部分で点線で示されているように、Ｐ９の判断を行って待った結果、システム状態が変化することもあるので、次回以降の判断時の状態に従って、判断分岐が行われることになる。
【０１３５】
すなわち、待った結果、直ぐに準備処理が完了した場合には、Ｐ２→Ｐ３→Ｐ８という流れとなり、一方、直ぐに完了しない別の準備処理に移行した場合には、Ｐ２→Ｐ４→Ｐ６→Ｐ１０という流れとなる。直ぐに完了しない別の準備処理に移行した場合とは、例えば、ステータス＝「自サーバ検索中」であったが、自サーバ内で目的の情報が得られなかったため、外部システムにアクセスし、ステータス＝「外部システムアクセス中」となっ

た場合等である。
【０１３６】
さらに、図８において、Ｐ１０の判断に基づきフィラーの再生を開始した後、フィラーの再生を行っている間に準備が完了すれば、Ｐ２→Ｐ３→Ｐ８という流れとなり、フィラーの再生を中断するか、またはフィラーの再生終了後に、準備が完了した複数の次発話候補の中からの次発話の選択が行われ、選択された次発話の再生が行われることになる。一方、フィラーの挿入（Ｐ１０）を行っても、未だ準備が続いていた場合には、直ぐに完了しない準備処理が続いていることになるので、Ｐ２→Ｐ４→Ｐ６→Ｐ１０という流れとなり、再び、フィラーの挿入（Ｐ１０）が行われる。なお、フィラーの再生中に、Ｐ１０の判断が再びなされた場合には、再生中のフィラーを優先させて再生を続ける。新たなＰ１０の判断を優先させると、例えば「ちょっと待っ」「ちょっと待っ」「ちょっと待っ」のような繰り返しをする不自然な発話になってしまうからである。
【０１３７】
＜再生装置２０／システム発話タイミング検出手段２２／ユーザ発話権終了判定用閾値調整手段２２Ｇの構成＞
【０１３８】
ユーザ発話権終了判定用閾値調整手段２２Ｇは、ユーザ発話権終了判定用パターン認識器２２Ｃによる維持・終了の識別処理で用いるユーザ発話権終了判定用閾値の事前調整、ユーザ発話権終了判定用閾値の下方調整を行うことを決めるための下方調整用閾値の事前調整、およびユーザ発話権終了判定用閾値のリアルタイム調整の各処理を実行するものである。
【０１３９】
ここで、事前調整は、ユーザとの対話（その日またはその時における対話、あるいは、その週、その月、その季節、その年等の所定の区切りの期間における対話）を開始する前に行う調整であり、ユーザ情報記憶手段５２に記憶されているユーザの属性情報（当該ユーザとの対話中における一時的な情報ではなく、当該ユーザとの複数回の対話を通じて得られた蓄積情報）を用いて行われる。一方、リアルタイム調整は、ユーザとの対話の開始後（特に、ユーザ発話の進行中）に行う調整であり、ユーザ状態記憶手段３２に記憶されているユーザ状態を示す情報（対話中における一時的な情報）や、システム状態記憶手段３１に記憶されているシステム状態を示す情報（対話中における一時的な情報）を用いて行われる。
【０１４０】
具体的には、図１１に示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、対話相手のユーザについてのユーザ識別情報を用いてユーザ情報記憶手段５２に記憶されている当該ユーザの衝突の発生情報（蓄積情報）を取得し、当該ユーザとの衝突の発生頻度または累積発生回数を算出する。この際、衝突の発生頻度は、例えば、１日、１週間、１か月等の所定の長さの期間における衝突の発生回数としてもよく、対話の総数に対する衝突の発生回数としてもよく、ユーザ発話からシステム発話への交替の総数に対する衝突の発生回数としてもよく、発生頻度の単位は、任意である。そして、算出した衝突の発生頻度または累積発生回数が、予め定められた上方調整用閾値以上または超過の場合には、ユーザ発話権終了判定用閾値を標準値または前回調整値よりも高く設定する上方調整を実行する。これにより、図１１に示すように、システム発話の開始タイミングが遅れる方向、すなわち衝突回避方向に調整される。
【０１４１】
また、図１２に示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、対話相手のユーザについてのユーザ識別情報を用いてユーザ情報記憶手段５２に記憶されている当該ユーザについてのユーザ発話の終了からシステム発話の開始までの複数の交替潜時（システムの交替潜時の蓄積情報）を取得し、当該ユーザを対話相手とするときのシステムの交替潜時の長短の傾向を示す平均値若しくはその他の指標値を算出する。この際、交替潜時の長短の傾向を示す指標値は、複数の交替潜時をまとめた指標値であれば、いずれでも

よく、例えば、平均値、中央値、最頻値等とすることができ、中央値や最頻値とする場合は、交替潜時を幾つかに区分していずれかの区分に帰属させ、各区分の代表値の中のいずれかを中央値、最頻値とすること等ができる。そして、算出した交替潜時の指標値が、予め定められた下方調整用閾値以上または超過の場合には、ユーザ発話権終了判定用閾値を標準値または前回調整値よりも低く設定する下方調整を実行する。これにより、図１２に示すように、システム発話の開始タイミングが早まる方向、すなわち交替潜時が短くなる方向に調整される。

【０１４２】

さらに、図１２に示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、対話相手のユーザについてのユーザ識別情報を用いてユーザ情報記憶手段５２に記憶されている当該ユーザの複数の発話速度（蓄積情報）を取得し、当該ユーザの発話速度の傾向を示す平均値若しくはその他の指標値を算出する。なお、発話速度の単位は「モーラ／秒」等である。この際、ユーザの発話速度の傾向を示す指標値は、複数の発話速度をまとめた指標値であれば、いずれでもよく、例えば、平均値、中央値、最頻値等とすることができる。そして、下方調整用閾値を、算出した発話速度の指標値を用いて、発話速度の指標値が大きい（速い）ときには当該下方調整用閾値が小さくなり、発話速度の指標値が小さい（遅い）ときには当該下方調整用閾値が大きくなるように予め定められた関数により設定する。この関数は、上述した前提条件を満たす関数であれば、どのような関数でもよく、図１２の例では、１次関数とされているが、例えば、２次以上の関数でもよく、１段または多段のステップ関数等でもよい。これにより、早口のユーザについては、下方調整用閾値が小さくなり、比較的短い交替潜時でも、ユーザ発話権終了判定用閾値の下方調整を行うことができるようになり（下方調整の条件が成立するようになり）、交替潜時が短くなる方向へのシステム発話の開始タイミングの調整を行うことができるようになる。一方、ゆっくり発話するユーザについては、下方調整用閾値が大きくなり、比較的長い交替潜時でないと、ユーザ発話権終了判定用閾値の下方調整を行うことができないようになり（下方調整の条件が成立しなくなり）、交替潜時が短くなる方向へのシステム発話の開始タイミングの調整を行うことができないようになる。

【０１４３】

また、図９中の実線で示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、ユーザ状態記憶手段３２に記憶されている対話相手のユーザについてのユーザ発話継続時間（リアルタイム情報）を逐次取得し、取得したユーザ発話継続時間が、予め定められた短時間判定用閾値以下または未満の場合には、ユーザ発話権終了判定用閾値を標準値よりも高く設定し、予め定められた長時間判定用閾値以上または超過の場合には、ユーザ発話権終了判定用閾値を標準値よりも低く設定する処理を逐次実行する。これにより、ユーザ発話の開始直後には、ユーザ発話権が終了したという識別結果が出にくくなり、ユーザ発話の開始時からの経過時間が長くなると、ユーザ発話権が終了したという識別結果が出やすくなる。

【０１４４】

また、図９中の二点鎖線で示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、ユーザ状態記憶手段３２に記憶されている対話相手のユーザについてのユーザ発話継続時間（リアルタイム情報）を逐次取得し、ユーザ発話権終了判定用閾値を、取得したユーザ発話継続時間を用いて、ユーザ発話継続時間が短いときには当該ユーザ発話権終了判定用閾値が高くなり、ユーザ発話継続時間が長いときには当該ユーザ発話権終了判定用閾値が低くなるように予め定められた関数（図９中の実線で示された階段状の関数に限らず、それ以外の様々な関数）により逐次設定してもよい。この関数は、上述した前提条件を満たす関数であれば、どのような関数でもよく、例えば、１次関数でもよく、２次以上の関数でもよく、１段のステップ関数や、図９中の実線で示された２段のステップ関数以外の多段（３段以上）のステップ関数等でもよい。

【０１４５】

さらに、図１０に示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、システ

ム状態記憶手段３１に記憶されている対話相手のユーザについての目的データの残数（対話目的を達成するためのシステムの最終の次発話候補の内容データとなり得る題材データである目的データの残数）および／または次発話候補の重要度（次発話準備手段４３による準備処理で得られた複数の次発話候補の内容データの各々に付されている重要度）を取得する処理を逐次実行する。これらの目的データの残数および／または次発話候補の重要度は、システムによる発話開始に対する要求の強さの度合いを示すシステム発話意欲度の指標値である。

【０１４６】

ここで、目的データの残数については、例えば、目的データの残数が１であれば、システム発話意欲度が強く、目的データの残数が２以上であれば、システム発話意欲度が弱い設定とすること等ができる。例えば、情報検索対話において、ユーザ発話の進行に伴ってユーザによる条件提示が進み、その条件提示の内容に応じて目的データの残数が１になった時点で、システム発話意欲度を強く設定することができる。具体例を挙げると、飲食店を検索するときに、ユーザが、システムの「食べる場所はどこ？」に対して「東京駅周辺のお店を探したい。」と答え、システムの「何が食べたいの？」に対して「中華料理が食べたい。」と答え、システムの「どんなお店がいいの？」に対して「おいしいと評判のお店がよくて、それと・・・」と答える等の条件提示を積み重ねていった結果、目的データ（情報提供する飲食店のデータ）が１つに絞り込まれる場合があり、この場合、ユーザは、それ以上、条件提示を行う必要はなく（つまり、「それと・・・」以降の条件提示を行う必要はなく）、１つに絞り込まれた目的データ（飲食店のデータ）を早く再生した方がよいという状況になるので、システム発話意欲度が強くなる。また、ユーザが「おいしいと評判のお店がよい。」と言った後に「待って、やっぱり評判はどうでもいいから、安いお店がいいな。」と訂正の発話を行い、それに基づき、再び、目的データ（飲食店のデータ）の残数が２以上になったときには、システム発話意欲度が弱くなる。システム発話意欲度の数値化方法は任意であり、例えば、１～１０の１０段階（段階数は任意）、０～１の連続値、０～１００％の連続値等とすることができる。例えば、目的データの残数＝１の場合には、システム発話意欲度＝１０段階のうちの１０とし、目的データの残数＝２または３の場合には、システム発話意欲度＝１０段階のうちの７とし、目的データの残数＝４以上の場合には、システム発話意欲度＝１０段階のうちの２とすること等ができる。この対応関係は、予め定めておけばよい。なお、上記の例の対応関係では、１０段階のうち使用されないシステム発話意欲度が存在するが、これは、下記の次発話候補の重要度により定まるシステム発話意欲度とのレベル合わせをしているからである。

【０１４７】

また、次発話候補の重要度については、重要度が高ければ、システム発話意欲度が強くなり、重要度が低ければ、システム発話意欲度が弱くなる関係にある。この重要度は、記事データ（ニュースやコラムや歴史等の各種の話題を記載した記事の原文データ）を要約してシナリオデータを生成する際の元の記事データの各構成文の重要度と同じとしてもよいが、本実施形態では、それだけではなく、防災関連情報の緊急性や日常生活への影響の大きさ等を加味した重要度としている。例えば、ニュース対話において、重要度が、「ＸＸＸで大きな地震が発生しましたので、ＹＹＹ沿岸地域の方は、すぐに高台に避難してください。」＝１０、「ＸＸＸ地方に大雨洪水警報が出ました。」＝８、「明日から消費税が１０％となります。」＝６、「早稲田花子選手が女子１００ｍの日本新記録を出しました。」＝４等のように、１～１０の１０段階の数値で設定されていれば、これらの数値をそのままシステム発話意欲度を示す数値とすること等ができる。また、重要度が８以上は、システム発話意欲度＝３段階のうちの３とし、重要度が７～５は、システム発話意欲度＝３段階のうちの２とし、重要度が４以下は、システム発話意欲度＝３段階のうちの１とすること等ができる。この対応関係は、予め定めておけばよい。そして、複数の次発話候補の内容データが次発話候補記憶手段３０に記憶され、システム状態記憶手段３１にそれらの複数の次発話候補の内容データの各々についての重要度が記憶されている場合には、複数の重要度の平均値、中央値、最頻値等を代表の重要度としてもよく、最も大きい重要

度や最も小さい重要度を代表の重要度としてもよい。

【０１４８】

なお、目的データの残数や、複数の次発話候補の内容データの各々の重要度をシステム状態記憶手段３１に記憶させるのではなく、次発話準備手段４３によりこれらを換算して求めたシステム発話意欲度を、システム状態記憶手段３１に記憶させてもよい。また、複数の次発話候補の内容データの各々の重要度をシステム状態記憶手段３１に記憶させるのではなく、次発話準備手段４３により求めた代表の重要度を、システム状態記憶手段３１に記憶させてもよい。さらに、目的データの残数により定まるシステム発話意欲度と、次発話候補の重要度により定まるシステム発話意欲度とを、対話の種別に応じて使い分けてもよいが、両者の平均値や加重平均値等を求めて統合して用いてもよい。

【０１４９】

そして、図１０に示すように、ユーザ発話権終了判定用閾値調整手段２２Ｇは、取得した目的データの残数および／または複数の次発話候補の内容データの各々の重要度からシステム発話意欲度を求め、ユーザ発話権終了判定用閾値を、求めたシステム発話意欲度を用いて、システム発話意欲度が強いときには当該ユーザ発話権終了判定用閾値が低くなり、システム発話意欲度が弱いときには当該ユーザ発話権終了判定用閾値が高くなるように予め定められた関数により設定する処理を逐次実行する。これにより、システム発話意欲度が強いときには、ユーザ発話権が終了したという識別結果が出やすくなり、システム発話意欲度が弱いときには、ユーザ発話権が終了したという識別結果が出にくくなる。

【０１５０】

＜再生装置２０／次発話選択用情報生成手段２３の構成＞

【０１５１】

図３において、次発話選択用情報生成手段２３は、韻律特徴量抽出手段２３Ａと、第１の発話意図識別器２３Ｂと、第２の発話意図識別器２３Ｅとを含んで構成されている。この次発話選択用情報生成手段２３には、例えば、前述した非特許文献３に記載された技術を採用することができる。

【０１５２】

韻律特徴量抽出手段２３Ａは、音声信号取得手段２１により取得したユーザ発話の音声信号から韻律特徴量（韻律情報）を抽出する処理を実行するものである。この韻律特徴量抽出手段２３Ａは、システム発話タイミング検出手段２２の音響特徴量抽出手段２２Ａと同様な構成を採用することができる。すなわち、音響特徴量抽出手段２２Ａで抽出された音響特徴量を、韻律特徴量（韻律情報）とすることができる。従って、この韻律特徴量抽出手段２３Ａと、システム発話タイミング検出手段２２の音響特徴量抽出手段２２Ａとは、共通化することができる。従って、例えば、韻律特徴量抽出手段２３Ａで得られる韻律特徴量（韻律情報）は、ＣＮＮオートエンコーダの中間層から取り出した２５６次元のボトルネック特徴量とすることができる。

【０１５３】

第１の発話意図識別器２３Ｂは、韻律特徴量抽出手段２３Ａで抽出した韻律特徴量（韻律情報）を用いて、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図を識別する処理を実行する第１の発話意図識別処理手段２３Ｃと、この識別処理で用いるモデル（パラメータ）を記憶する第１の発話意図識別モデル記憶手段２３Ｄとにより構成されている。この第１の発話意図識別器２３Ｂは、例えば、ＬＳＴＭ（ＲＮＮの一種）により構築することができる。

【０１５４】

具体的には、第１の発話意図識別器２３Ｂは、例えば、ＣＮＮオートエンコーダの中間層から取り出した２５６次元の韻律特徴量（韻律情報）を逐次入力し、ＬＳＴＭによるパターン認識処理を行って発話意図を識別し、その識別結果を出力する構成とすることができる（非特許文献３参照）。そして、この第１の発話意図識別器２３Ｂから出力された発話意図を、次発話選択手段２４に送ってもよい。

【０１５５】

第２の発話意図識別器２３Ｅは、第１の発話意図識別器２３Ｂで得られた韻律情報（例えば、ＬＳＴＭの隠れ層の値）と、音声認識処理手段４１による音声認識処理の結果として得られたユーザ発話の言語情報と、対話履歴記憶手段５０に記憶されているユーザとシステムとの間の対話履歴情報のうちの直前のシステム発話の言語情報とを用いて、質問、応答、相槌、補足要求、反復要求、理解、不理解、無関心、若しくはその他のユーザ発話意図を識別する処理を実行する第２の発話意図識別処理手段２３Ｆと、この識別処理で用いるモデル（パラメータ）を記憶する第２の発話意図識別モデル記憶手段２３Ｇとにより構成されている。この第２の発話意図識別器２３Ｅは、例えば、ＢＥＲＴにより構築することができる（非特許文献３参照）。ＢＥＲＴは、自然言語処理モデルであり、トランスフォーマのエンコーダ部分をユニットとする双方向トランスフォーマモデルである。この第２の発話意図識別器２３Ｅから出力された発話意図は、次発話選択手段２４に送られる。

【０１５６】

また、次発話選択用情報生成手段２３は、ユーザの顔画像やジェスチャー画像（身振り・手振りの画像）を取得し、顔の表情やジェスチャーの内容を解析し、その解析結果（表情の識別結果、身振り・手振りの意図の識別結果）を、次発話選択用情報として次発話選択手段２４に送ってもよい。

【０１５７】

＜再生装置２０／次発話選択手段２４の構成＞

【０１５８】

次発話選択手段２４は、システム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出された後に（システム発話タイミング検出手段２２からシステム発話の開始タイミングであるという判断結果を受け取ったときに）、次発話選択用情報生成手段２３による処理で得られたユーザ発話意図の識別結果と、音声認識処理手段４１による音声認識処理の結果として得られた言語情報（文字列）とを組み合わせて用いて、次発話準備手段４３による準備処理で得られて次発話候補記憶手段３０に記憶されている複数（但し、１つの場合もある。）の次発話候補の内容データの中から、発話生成手段２５で用いる次発話の内容データを選択し、選択した次発話の内容データを、発話生成手段２５に送るとともに、選択した次発話の内容データまたはその識別情報（例えば、シナリオＩＤ、発話節ＩＤ等）を、ネットワーク１を介して対話状態管理手段４２へ送信する処理を実行するものである。

【０１５９】

なお、ユーザ発話意図の識別結果（例えば、質問、相槌等の別）と、音声認識処理の結果として得られた言語情報（文字列）とのうちのいずれか一方だけで、次発話の内容データを選択することができる場合には、これらを組み合わせて用いなくてもよい。また、次発話選択用情報生成手段２３からではなく、システム発話タイミング検出手段２２から、システム発話の開始タイミングであるという判断結果とともにユーザ発話意図の識別結果を受け取った場合（システム発話タイミング検出手段２２において、どのようなユーザ発話意図で終了したのかを識別した場合）には、そのユーザ発話意図の識別結果を用いて、次発話の内容データを選択してもよい。さらに、次発話選択用情報生成手段２３から、ユーザの顔の表情やジェスチャーの内容についての解析結果を受け取った場合には、それらの解析結果を用いて、またはそれらの解析結果と他の情報とを組み合わせて、次発話の内容データを選択してもよい。

【０１６０】

また、次発話選択手段２４は、システム発話タイミング検出手段２２から、システム発話の開始タイミングであるという判断結果とともに、フィラーの挿入タイミングである旨の情報（どのような種別のフィラーを挿入するかの情報を含む）を受け取った場合には、指定された種別のフィラーの内容データ（音声データを含む）を、発話生成手段２５に送る処理を実行する。この際、挿入するフィラーの内容データまたは当該フィラーの種別の識別情報を、ネットワーク１を介して対話状態管理手段４２へ送信する処理を実行しても

よく、実行しなくてもよいが、実行した場合でも、対話状態管理手段４２は、フィラーの挿入を、対話履歴上は、システム発話として取り扱うのではなく、システム発話の準備用繋ぎ発話として取り扱う。この点については、後述する対話状態管理手段４２の説明で詳述するので、ここでは詳しい説明を省略する。なお、各種のフィラーの内容データ（音声データを含む）は、フィラーの種別の識別情報と対応付けて再生装置２０に設けられたフィラー記憶手段（不図示）に記憶させておけばよいが、フィラーは常に次発話候補になり得ると考え、次発話候補記憶手段３０に記憶させておいてもよい。後者とする場合、各種のフィラーの内容データ（音声データを含む）を、次発話準備手段４３から毎回ネットワーク１を介して受信する必要はなく、次発話候補記憶手段３０に固定的に準備されているデータとすればよい。なお、このようにフィラーを次発話候補であると考えて次発話候補記憶手段３０に記憶させたとしても、上述したように、対話状態管理手段４２は、フィラーの挿入が行われても、対話履歴上は、それをシステム発話として取り扱わないので、次発話候補記憶手段３０に固定的に記憶しておくフィラーと、それ以外の複数の次発話候補の内容データ（頻繁に更新されるデータ）とは別のものであり、単に同じ次発話候補記憶手段３０に記憶させるに過ぎない。

【０１６１】

具体的には、次発話選択手段２４は、音声認識処理手段４１による音声認識処理の結果（文字列）を用いて次発話の選択を行う場合には、音声認識処理の結果に含まれる各単語と、次発話候補記憶手段３０に記憶されている複数の次発話候補の内容データの各々に含まれる各単語とを用いて、キーワードマッチングを行い、マッチングした次発話候補の内容データを、次発話の内容データとして選択することができる。また、言語処理と機械学習とを合わせた複雑なマッチングを行ってもよい。さらに、音声認識処理の結果として得られた文字列と、複数の次発話候補の内容データ（文字列）の各々との類似度を、ｄｏｃ２ｖｅｃ等により求め、類似度の高い次発話候補の内容データを、次発話の内容データとして選択してもよい。

【０１６２】

また、電話の自動応答における音声対話等のように、システムがユーザに質問し、ユーザがそれに答えていく場合には、ユーザ発話の内容は、システムから与えられた選択肢等のように限られたものになるので、ユーザ発話の内容は予測することができる。この場合、次発話候補記憶手段３０に記憶されている複数の次発話候補の内容データの各々に、対応するユーザ発話の予測データが付随していれば、その付随しているユーザ発話の予測データのうちのいずれが、音声認識処理の結果として得られた文字列と一致するのかを判断することにより、一致したユーザ発話の予測データに対応する次発話候補の内容データを、次発話の内容データとして選択することができる。

【０１６３】

例えば、システム発話Ｓ（Ｎ）が「ＸＸＸ党と、ＹＹＹ党のどちらを支持しますか？」であり、システム発話Ｓ（Ｎ＋１）の複数（２つ）の候補として、「ＸＸＸ党のどの政治家が総理大臣になると思いますか？」という内容データおよびそれに付随する「ＸＸＸ党」というユーザ発話Ｕ（Ｎ）の予測データと、「ＹＹＹ党のどの政治家が党首に相応しいですか？」という内容データおよびそれに付随する「ＹＹＹ党」というユーザ発話Ｕ（Ｎ）の予測データとを、次発話準備手段４３により準備し、次発話候補記憶手段３０に記憶させたとする。このとき、ユーザ発話Ｕ（Ｎ）の音声認識処理の結果が「ＸＸＸ党」であれば、「ＸＸＸ党」というユーザ発話Ｕ（Ｎ）の予測データと一致するので、それに対応する「ＸＸＸ党のどの政治家が総理大臣になると思いますか？」がシステム発話Ｓ（Ｎ＋１）として選択され、発話生成手段２５により再生される。

【０１６４】

さらに、次発話選択手段２４は、ユーザ発話意図の識別結果（質問、相槌等）を用いて次発話の選択を行う場合には、ユーザ発話意図に対応するシステム発話種別が、次発話候補記憶手段３０に記憶されている複数の次発話候補の内容データの各々について定められているので、得られたユーザ発話意図の識別結果に対応するシステム発話種別である次発

話候補の内容データを、次発話の内容データとして選択することができる。
【０１６５】
例えば、ユーザ発話意図が「相槌」、「理解」等であれば、システム発話種別が主計画である次発話候補の内容データを選択し、ユーザ発話意図が「定義型質問」（用語の意義を問う質問）であれば、システム発話種別が副計画（定義）である次発話候補の内容データを選択し、ユーザ発話意図が「反復要求」、「不理解」であれば、システム発話種別が繰り返し用の主計画である次発話候補の内容データを選択し、ユーザ発話意図が「補足要求」等であれば、補足説明用の副計画（トリビア等）である次発話候補の内容データを選択する等のように、ユーザ発話意図と、システム発話種別との対応関係を予め定めておけばよい。この対応関係は、次発話選択手段２４を構成するプログラム内に記述されていてもよく、再生装置２０に設けられた発話意図・システム発話種別対応関係記憶手段（不図示）に記憶しておいてもよい。従って、シナリオデータの構成要素が、主計画要素であるか、副計画要素であるかも、システム発話種別に該当する。なお、主計画、副計画についての詳細は、図１３および図１４を用いて後述する。
【０１６６】
なお、ユーザ発話意図が「無関心」、「既知」であれば、長短２つ用意された主計画のうちの短い方の主計画である次発話候補の内容データを選択し、伝達情報量を減らすことができる。但し、これらのユーザ発話意図の場合は、次発話準備手段４３の入替準備手段４３Ｃにより、複数の次発話候補の内容データを入れ替える準備処理が進行しているか（ステータス＝準備中）、あるいは、既にその準備が完了し、次発話候補記憶手段３０に、別の話題のシナリオデータ内の先頭の構成要素（主計画要素）が記憶されているか、同じシナリオデータ内の別の構成要素（主計画要素）が記憶されていることもある。その場合は、その主計画要素を選択すればよい。
【０１６７】
また、次発話選択手段２４は、ユーザ発話意図の識別結果（質問、相槌等）と、音声認識処理の結果として得られた言語情報（文字列）とを組み合わせて用いて、次のように、次発話を選択することができる。
【０１６８】
例えば、システム発話Ｓ（Ｎ）が「早稲田太郎選手が４回転フリップを成功させたよ。」であり、システム発話Ｓ（Ｎ＋１）の複数の候補として、「グランプリシリーズのカナダ大会で跳んだそうだ。」（主計画要素）と、「早稲田太郎選手は、…」という早稲田太郎の人物の説明データ（副計画要素の定義）と、「４回転フリップっていうのは、…」という４回転フリップの技の説明データ（副計画要素の定義）と、繰り返し用の「早稲田太郎選手が４回転フリップを成功させたよ。」（主計画要素）とが、次発話準備手段４３により準備され、次発話候補記憶手段３０に記憶されているとする。このとき、Ｕ（Ｎ）のユーザ発話意図が「相槌」、「理解」であったとすると、「グランプリシリーズのカナダ大会で跳んだそうだ。」（主計画要素）を次発話Ｓ（Ｎ＋１）として選択すればよく、ユーザ発話意図が「反復要求」であったとすると、繰り返し用の「早稲田太郎選手が４回転フリップを成功させたよ。」（主計画要素）を選択すればよい。しかし、Ｕ（Ｎ）のユーザ発話意図が「質問」であったとすると、次発話候補記憶手段３０に記憶されている複数の次発話候補の内容データの中には、定義型質問に対するシステム応答（副計画要素の定義）が２つ用意されているので、早稲田太郎選手について質問しているのか、４回転フリップについて質問しているのかが判明しないと、システム応答を行うことができないが、いずれの質問であるかは、ユーザ発話意図だけでは判断することができない。そこで、音声認識処理の結果として得られた言語情報（文字列）を用いて、どちらの質問であるかを判断し、どちらのシステム応答（副計画要素の定義）を選択するのかを判断する。一方、次発話候補記憶手段３０に記憶されている複数の次発話候補の内容データの中に、定義型質問に対するシステム応答（副計画要素の定義）が１つしかない場合には、音声認識処理の結果を使用せずに（つまり、ユーザ発話意図だけで）、その１つのシステム応答（副計画要素の定義）を選択することができる。

【０１６９】
また、逆に音声認識処理の結果だけでは、次発話を選択できない場合もある。例えば、「えっ？」というユーザ発話は、驚きなのか、質問なのか、聞き返し（反復要求）なのかは判断できないので、ユーザ発話意図を用いて、次発話を選択することができる。
【０１７０】
なお、次発話候補の内容データが、次発話候補記憶手段３０に１つも記憶されていない期間があるが、これは、次発話選択手段２４による処理には影響しない。なぜなら、次発話候補の内容データが次発話候補記憶手段３０に１つも記憶されていない期間は、次発話準備手段４３による準備処理が完了していない期間（準備中の期間）であるが、図８のＰ５→Ｐ９の流れの場合（直ぐに準備が完了する場合）には、システム発話の開始タイミングではないと判断されるので、次発話選択手段２４による処理には進まず、一方、図８のＰ６→Ｐ１０の流れの場合（直ぐに準備が完了しない場合）には、次発話選択手段２４による処理に進むものの、フィラーの挿入になるので、次発話候補の内容データは必要ないからである。
【０１７１】
＜再生装置２０／発話生成手段２５の構成＞
【０１７２】
発話生成手段２５は、システム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出された後に、次発話選択手段２４で選択された次発話の内容データ（次発話準備手段４３による準備処理で得られた複数の次発話候補の内容データの中から選択された次発話の内容データ）を用いて、システム発話の音声信号の再生を含むシステム発話生成処理を実行するものである。この発話生成手段２５には、スピーカ、ディスプレイも含まれる。
【０１７３】
この際、発話生成手段２５は、次発話選択手段２４から受け取った次発話の内容データに音声データ（例えばｗａｖファイル等）が含まれていない場合には、次発話選択手段２４から受け取ったテキストデータから音声データを生成する音声合成処理も実行する。但し、音声合成処理は、システム応答の遅延防止の観点から、次発話準備手段４３で実行するか、または題材データとして予め用意されていることが好ましい。
【０１７４】
また、発話生成手段２５は、システム発話の音声信号の再生処理を実行するとともに、次発話選択手段２４から受け取った次発話の内容データに映像データや静止画データ、あるいは楽曲データが付随している場合には、ディスプレイでの動画や静止画の再生処理、あるいは音楽の再生処理も実行する。例えば、直前のシステム発話が「早稲田太郎選手が４回転フリップを跳びました。」であり、それに対するユーザの反応が「４回転フリップってどんな技？」という質問だった場合に、４回転フリップの技の説明用の映像を再生し、「早稲田太郎選手ってどんな選手なの？」という質問だった場合に、早稲田太郎選手の顔画像を再生すること等ができる。また、直前のシステム発話が「ＸＸＸホールで第九が演奏されました。」であり、それに対するユーザの反応が「第九ってどんな曲なの？」という質問だった場合に、第九の楽曲データを再生すること等ができる。なお、システム発話中に、システム発話の音声信号の再生と同期または略同期させて、ディスプレイでシステム発話の内容を示すテキスト表示を行ってもよい。
【０１７５】
さらに、発話生成手段２５は、音声信号取得手段２１により取得したユーザ発話の音声信号と、再生中のシステム発話の音声信号との衝突の発生を検出し、検出した衝突の発生情報を、ネットワーク１を介して対話サーバ４０へ送信し、ユーザ識別情報と関連付けてユーザ情報記憶手段５２に記憶させる処理を実行する。この際、衝突には２種類あり、ここで検出する衝突は、［１］ユーザ発話権が終了したという識別結果が出て、システム発話を開始したところ、実際にはユーザ発話権が維持されていて、両者の発話が重なった場合の衝突である。従って、［２］ユーザ発話権が終了したものの、システム発話の開始が

遅れたために、再び、ユーザ発話が開始されてしまい、ほぼ同時に両者の発話が開始されて重なった場合の衝突ではないので、この［２］の場合の衝突を排除する処理を実行する。例えば、衝突を起こしたときの直前の無音区間の長さ（衝突を起こしたユーザ発話の音声区間の開始時点とその直前のユーザ発話の音声区間の終了時点との間の時間間隔）が、予め定めた衝突種別判定用閾値以上または超過の場合に、［２］の場合の衝突であると判断し、排除すること等ができる。また、衝突の発生前後の関連するデータを全て保存しておき、事後的に［２］場合の衝突であるか否かを判断し、［２］の衝突を排除する処理を行ってもよい。関連するデータとは、例えば、衝突の直前のユーザ発話の音声区間の終了時刻、システム発話タイミング検出手段２２によるシステム発話の開始タイミングの検出時刻、発話生成手段２５によるシステム発話の音声信号の再生開始時刻、衝突を起こしたユーザ発話の音声区間の開始時刻、衝突の前後双方のユーザ発話の音声認識処理の結果としての言語情報、衝突を起こしたシステム発話の内容データ（テキストデータ）等である。さらに、これらの関連するデータを用いて学習を行い、［１］と［２］の衝突を識別する識別器を構築し、事後的に、またはリアルタイム若しくは略リアルタイムで、識別器による識別結果に従って、［２］の衝突を排除する処理を行ってもよい。なお、本発明は、予め次発話候補を準備する等、システム応答の遅延防止が図られているので、［２］の場合の衝突は、殆ど発生しないようになっている。

【０１７６】

また、発話生成手段２５は、音声信号取得手段２１により取得したユーザ発話の音声信号を用いてユーザ発話の終了時刻を検出するとともに、システム発話の音声信号の再生開始時刻を検出することにより、ユーザ発話の終了からシステム発話の開始までの交替潜時を計測し、計測したシステムの交替潜時を、ネットワーク１を介して対話サーバ４０へ送信し、ユーザ識別情報と関連付けてユーザ情報記憶手段５２に記憶させる処理を実行する。

【０１７７】

さらに、発話生成手段２５は、音声信号取得手段２１により取得したユーザ発話の音声信号を用いてユーザ発話の開始時刻を検出し、検出した開始時刻と現在時刻との差分によりシステム発話継続時間を逐次計測し、計測したシステム発話継続時間を、ユーザ状態記憶手段３２に逐次記憶させる処理を実行する。

【０１７８】

そして、発話生成手段２５は、音声認識処理手段４１による音声認識処理の結果として得られたユーザ発話の言語情報をネットワーク１を介して逐次取得し、取得した言語情報およびその取得時刻（または、言語情報とともに取得した時刻情報若しくは時間情報）を用いて発話速度をリアルタイムで算出し、算出したリアルタイムの発話速度を、ユーザ状態記憶手段３２に逐次記憶させる処理を実行する。また、発話生成手段２５は、対話全体におけるユーザ発話についての発話速度を算出し、算出した発話速度を、ネットワーク１を介して対話サーバ４０へ送信し、ユーザ識別情報と関連付けてユーザ情報記憶手段５２に記憶させる処理を実行する。

【０１７９】

＜再生装置２０／次発話候補記憶手段３０、システム状態記憶手段３１、ユーザ状態記憶手段３２の構成＞

【０１８０】

次発話候補記憶手段３０は、次発話準備手段４３からネットワーク１を介して送信されてきた複数の次発話候補の内容データを、それらのデータの識別情報（例えば、シナリオＩＤ、発話節ＩＤ等）と対応付けて記憶するものである。記憶する次発話候補の内容データには、テキストデータの他に、音声データ（例えばｗａｖファイル等）が含まれ、さらに映像データや静止画データ、あるいは楽曲データが付随している場合もある。なお、再生装置２０にフィラー記憶手段（不図示）を設けない場合には、各種のフィラーの内容データ（音声データを含む）を、既に述べたように固定的に準備されているデータとして、フィラーの種別の識別情報と対応付けて次発話候補記憶手段３０に記憶しておいてもよい

。
【０１８１】
また、複数の次発話候補の内容データの各々には、データの属性を示すシステム発話種別（例えば、シナリオデータにおける主計画・副計画の別等）が対応付けられて記憶されている。このシステム発話種別については、次発話選択手段２４の説明で既に詳述しているので、ここでは詳しい説明を省略する。
【０１８２】
システム状態記憶手段３１は、システム状態を示す情報として、次発話準備手段４３による準備処理の状態（準備完了・各種の準備中の別を示すステータス）、目的データの残数（対話目的を達成するためのシステムの最終の次発話候補の内容データとなり得る題材データである目的データの残数）、および次発話候補の重要度（次発話準備手段４３による準備処理で得られた複数の次発話候補の内容データの各々についての重要度）を記憶するものである。このシステム状態は、現在進行しているユーザとの対話中に得られる一時的な情報（逐次更新されるリアルタイム情報）である。このうち、目的データの残数および次発話候補の重要度は、システム発話意欲度の指標値であるが、システム発話意欲度については、システム発話タイミング検出手段２２のユーザ発話権終了判定用閾値調整手段２２Ｇの説明で既に詳述しているので（図１０参照）、ここでは詳しい説明を省略する。準備完了・各種の準備中の別を示すステータス、目的データの残数および次発話候補の重要度のいずれについても、次発話準備手段４３からネットワーク１を介して送信されてきてシステム状態記憶手段３１に記憶されて逐次更新される。
【０１８３】
ユーザ状態記憶手段３２は、ユーザ状態を示す情報として、進行中のユーザ発話についての発話開始からのユーザ発話継続時間、および進行中のユーザ発話についての発話速度を記憶するものである。このユーザ状態は、現在進行しているユーザとの対話中に得られる一時的な情報（逐次更新されるリアルタイム情報）であるため、ユーザ情報記憶手段５２に記憶されているユーザの属性情報（複数回の対話を通じて得られた蓄積情報）とは異なる。本実施形態では、ユーザ発話継続時間および発話速度のいずれについても、発話生成手段２５により計測され、ユーザ状態記憶手段３２に記憶されて逐次更新される。
【０１８４】
＜対話サーバ４０／音声認識処理手段４１の構成＞
【０１８５】
音声認識処理手段４１は、音声信号取得手段２１により取得したユーザ発話の音声信号をネットワーク１を介して逐次取得し、取得したユーザ発話の音声信号についての音声認識処理を実行し、音声認識処理の結果として得られた言語情報を逐次出力し、出力した言語情報を、対話状態管理手段４２を介して次発話準備手段４３に逐次送るとともに、ネットワーク１を介してシステム発話タイミング検出手段２２、次発話選択用情報生成手段２３、次発話選択手段２４、および発話生成手段２５にも逐次送信する処理を実行するものである。
【０１８６】
この音声認識処理手段４１による音声認識処理は、システム発話タイミング検出手段２２によるユーザ発話権の維持・終了を識別するパターン認識処理とは非同期で実行される。
【０１８７】
具体的には、図６および図７に示すように、音声認識処理手段４１は、ショートポーズセグメンテーションと呼ばれる機能により、音声信号取得手段２１により取得したユーザ発話の音声信号の中に短時間の無音区間が現れるたびに音声信号を細かく区切り、音声認識対象とする区間を順次確定させていく。これにより、長時間の音声入力を自動的に区切りながら逐次的に音声認識処理を実行することができる。このショートポーズセグメンテーションでの音声認識対象とする音声信号の区間（対応する音声認識処理の時間長は、図６中および図７中の点線で示されている。）は、通常の音声区間検出（ＶＡＤ）で決定さ

れる音声区間（対応する音声認識処理の時間長は、図６中および図７中の実線で示されている。）よりも短い。

【０１８８】

このショートポーズセグメンテーションの機能は、対話サーバ４０の音声認識処理手段４１の中に設けてもよく、あるいは、音声認識処理手段４１により、図示されない外部サーバにアクセスしてストリーミング音声認識を行うようにしてもよい。後者の場合は、例えば、グーグル・クラウド・スピーチＡＰＩ（ｈｔｔｐｓ：／／ｃｌｏｕｄ．ｇｏｏｇｌｅ．ｃｏｍ／ｓｐｅｅｃｈ／）のストリーミング音声認識等を用いることができる（非特許文献２参照）。

【０１８９】

＜対話サーバ４０／対話状態管理手段４２の構成＞

【０１９０】

対話状態管理手段４２は、進行中のユーザとシステムとの間の対話状態を管理する処理を実行するものである。ここで、システム発話とユーザ発話との時間的な前後関係の説明を容易にするため、図６の最上部に示すように、最初のシステム発話をＳ（１）、最初のユーザ発話をＵ（１）とし、以降、Ｓ（２）、Ｕ（２）、Ｓ（３）、Ｕ（３）、…、Ｓ（Ｎ－１）、Ｕ（Ｎ－１）と対話が進み、直前のシステム発話をＳ（Ｎ）、進行中のユーザ発話をＵ（Ｎ）とし、さらにＵ（Ｎ）を音声区間Ｕ（Ｎ，１）、Ｕ（Ｎ，２）、Ｕ（Ｎ，３）、…に分割し、Ｕ（Ｎ，Ｋ）まで進んでいるものとする。そして、未来の新たな音声区間をＵ（Ｎ，Ｋ＋１）とする。なお、実際にはショートポーズセグメンテーションにより処理が進行するので、Ｕ（Ｎ，Ｋ＋１）よりも細かい区間で、新たな出力が得られる。

【０１９１】

具体的には、図４に示すように、対話状態管理手段４２は、次発話選択手段２４からネットワーク１を介して送信されてくる選択結果を受信する処理を実行する。この選択結果は、次発話選択手段２４により選択された次発話Ｓ（Ｎ＋１）の内容データまたはその識別情報（例えば、シナリオＩＤ、発話節ＩＤ等）である。

【０１９２】

そして、対話状態管理手段４２は、次発話選択手段２４からの選択結果の受信により、システム発話の開始タイミングが検出されたこと、すなわちユーザ発話権が終了したことを把握することができるので、その時点まで対話状態管理手段４２のメモリ（主メモリでよい）で保持していた言語情報、すなわちユーザ発話権が終了したユーザ発話Ｕ（Ｎ）の発話区間全体の内容データを、対話履歴記憶手段５０に記憶させる処理を実行する。この時点の前までには、図６の最上部に示すように、対話履歴記憶手段５０には直前のシステム発話Ｓ（Ｎ）までが保存されているので、これにユーザ発話Ｕ（Ｎ）が追加されることになる。

【０１９３】

また、対話状態管理手段４２は、次発話選択手段２４からの選択結果の受信により、システム発話の開始タイミングが検出されたこと、すなわち次発話Ｓ（Ｎ＋１）の再生が開始されることを把握することができるので、選択結果として受信した次発話Ｓ（Ｎ＋１）の内容データを、対話履歴記憶手段５０に記憶させる処理を実行する。これにより、直前のシステム発話Ｓ（Ｎ）まで保存されていた対話履歴記憶手段５０には、ユーザ発話Ｕ（Ｎ）およびシステム発話Ｓ（Ｎ＋１）が追加されることになる。

【０１９４】

さらに、対話状態管理手段４２は、次発話選択手段２４からの選択結果の受信により、システム発話の開始タイミングが検出されたこと、すなわち次発話Ｓ（Ｎ＋１）の再生が開始されることを把握することができるので、さらに次の次発話候補の準備処理を開始させるための準備開始指示情報を、次発話準備手段４３に送る処理を実行する。これにより、次発話準備手段４３によるシステム発話Ｓ（Ｎ＋２）についての複数の候補の準備処理が開始されることになる。

【０１９５】

また、対話状態管理手段４２は、ユーザ発話Ｕ（Ｎ）の進行中には、音声認識処理手段４１から逐次出力される音声認識処理の結果としてのユーザ発話の言語情報を逐次受け取り、受け取った言語情報を、次発話準備手段４３に入替要否判断のために逐次送るとともに、対話状態管理手段４２のメモリ（主メモリでよい）に保持する。この際、音声認識処理手段４１から逐次受け取る音声認識処理の結果は、ショートポーズセグメンテーションによる短い区間についての音声認識処理の分であるから、Ｕ（Ｎ．１）、Ｕ（Ｎ，２）、Ｕ（Ｎ，３）、…よりも細かい区間についての音声認識処理の結果である（図６参照）。なお、音声認識処理手段４１から受け取った言語情報を、次発話準備手段４３に逐次送る際には、受け取った言語情報だけを送ってもよく、対話状態管理手段４２のメモリに保持しているその時点までのユーザ発話Ｕ（Ｎ）の全部の言語情報を送ってもよい。

【０１９６】

また、対話状態管理手段４２は、上述したように、ユーザ発話Ｕ（Ｎ）の進行中には、音声認識処理手段４１から逐次受け取った言語情報を、次発話準備手段４３に入替要否判断のために逐次送るが、結果的に、それがユーザ発話Ｕ（Ｎ）の発話区間全体における最後の部分（または発話区間全体）であったとしても、次発話準備手段４３に送る。そして、次発話準備手段４３において、ユーザ発話Ｕ（Ｎ）の最後の部分を含めて入替要否判断を行い、システム発話Ｓ（Ｎ＋１）の複数の候補を、新しい複数の候補に入れ替えると判定し、入替の準備を行った場合には、それらの新しい複数の次発話候補（システム発話Ｓ（Ｎ＋１）の複数の候補）の内容データが、ネットワーク１を介して再生装置２０へ送信され、複数の次発話候補として次発話候補記憶手段３０に既に記憶されているシステム発話Ｓ（Ｎ＋１）の複数の候補が、新しい複数の候補に更新される。その後、システム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出されたときには、次発話選択手段２４によりシステム発話Ｓ（Ｎ＋１）の新しい複数の候補のうちの１つが選択され、その選択結果が、ネットワーク１を介して対話状態管理手段４２に送信されてくるので、対話状態管理手段４２のメモリに保持されているユーザ発話Ｕ（Ｎ）の発話区間全体の内容データ、および選択結果として受信したシステム発話Ｓ（Ｎ＋１）の内容データを、対話履歴記憶手段５０に記憶させるとともに、システム発話Ｓ（Ｎ＋２）の複数の候補を準備するための準備開始指示情報を、次発話準備手段４３に送る。

【０１９７】

なお、上記において、仮に、システム発話Ｓ（Ｎ＋１）の複数の候補を、新しい複数の候補に入れ替える準備処理に多少時間がかかった場合でも、システム状態記憶手段３１に記憶されている準備状態を示すステータスが準備中になるので、システム発話の開始タイミングは検出されないことから（図８のＰ９参照）、次発話候補記憶手段３０に記憶されているシステム発話Ｓ（Ｎ＋１）の複数の候補は更新されないまま保たれ（但し、入替の準備を開始した時点でクリアしてもよい。）、新しい複数の候補への入替を待つことになる。一方、準備処理にかなりの時間がかかった場合には、フィラーが挿入されるが（図８のＰ１０参照）、この場合も次発話候補記憶手段３０に記憶されているシステム発話Ｓ（Ｎ＋１）の複数の候補は更新されないまま保たれ（但し、入替の準備を開始した時点でクリアしてもよい。）、新しい複数の候補への入替を待つことになる。この場合、フィラーの挿入情報は、次発話選択手段２４から対話状態管理手段４２へ送信してもよく、送信しなくてもよいが、送信した場合でも、フィラーの挿入情報を受信した対話状態管理手段４２は、フィラーの挿入を、選択されたシステム発話Ｓ（Ｎ＋１）として取り扱うわけではないので、システム発話Ｓ（Ｎ＋２）の準備のための準備開始指示情報を次発話準備手段４３に送る処理は行わない。新しいシステム発話Ｓ（Ｎ＋１）の準備処理に時間がかかっているので、フィラーを挿入したのに、そのフィラーの挿入をもって、さらに次のシステム発話Ｓ（Ｎ＋２）の準備を開始するための処理を行うのは不合理だからである。但し、挿入したフィラーの情報を、システム発話Ｓ（Ｎ＋１）として取り扱うのではなく、システム発話Ｓ（Ｎ＋１）の準備用繋ぎ発話Ｓ（Ｎ＋１：準備）として対話履歴記憶手段５０に記憶させてもよい。

【０１９８】

また、対話状態管理手段４２は、対話の開始時には、最初のシステム発話Ｓ（１）についての準備開始指示情報を次発話準備手段４３に送るが、Ｓ（１）については、複数の候補を準備する必要はないので、次発話準備手段４３は、Ｓ（１）の準備開始指示情報を受け取った場合には、１つの次発話（システム発話Ｓ（１））の内容データを、ネットワーク１を介して再生装置２０へ送信し、次発話候補記憶手段３０にＳ（１）の内容データを記憶させればよい。Ｓ（１）の内容データを記憶させた時点で、ユーザは発話していないので、すぐにシステム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出され、次発話選択手段２４によりＳ（１）が選択され、発話生成手段２５により（１）の再生が開始されることになる。また、次発話選択手段２４によりＳ（１）が選択されると、その選択結果がネットワーク１を介して対話状態管理手段４２へ送信されるので、対話状態管理手段４２は、選択結果として受信したシステム発話Ｓ（１）を対話履歴記憶手段５０に記憶させるとともに、システム発話Ｓ（２）の複数の候補の準備のための準備開始指示情報を次発話準備手段４３へ送る。なお、この時点で、ユーザは未だ発話していないので、対話履歴記憶手段５０へのユーザ発話の保存はない。

【０１９９】

＜対話サーバ４０／次発話準備手段４３の構成＞

【０２００】

次発話準備手段４３は、システム発話タイミング検出手段２２によるパターン認識処理の周期に依拠しないタイミングで、かつ、システム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出される前に、題材データ記憶手段５１に記憶された題材データまたはネットワーク１を介して接続された外部システムである題材データ提供システム６０に記憶された題材データを用いるとともに、ユーザとシステムとの間の対話履歴情報の少なくとも一部および／または音声認識処理手段４１による進行中のユーザ発話についての途中までの音声認識処理の結果を用いて、システムの次発話の内容データ（本実施形態では、複数の次発話候補の内容データ）を取得または生成する準備処理を実行するものである。

【０２０１】

より詳細には、図４に示すように、次発話準備手段４３は、次発話候補初期準備手段４３Ａと、入替要否判断手段４３Ｂと、入替準備手段４３Ｃと、先行次発話候補情報記憶手段４３Ｄとを含んで構成されている。

【０２０２】

次発話候補初期準備手段４３Ａは、対話状態管理手段４２からの準備開始指示情報を受け取ったときに、システムの複数の次発話候補の内容データを取得または生成する準備処理を実行するものである。

【０２０３】

入替要否判断手段４３Ｂは、次発話候補初期準備手段４３Ａにより準備した複数の次発話候補の内容データ、または入替要否判断手段４３Ｂ自身により前回準備した複数の次発話候補の内容データを、別の複数の次発話候補の内容データに入れ替えるか否かを判断する処理を実行するものである。

【０２０４】

入替準備手段４３Ｃは、入替要否判断手段４３Ｂにより入替が必要であると判断した場合に、現在、次発話候補記憶手段３０に記憶されている最新の複数の次発話候補の内容データとは別の複数の次発話候補の内容データを取得または生成する準備処理を実行するものである。

【０２０５】

先行次発話候補情報記憶手段４３Ｄは、次発話候補初期準備手段４３Ａや入替準備手段４３Ｃによる準備処理を行って得られた複数の次発話候補の内容データ、すなわちネットワーク１を介して再生装置２０へ送信し、現在、次発話候補記憶手段３０に記憶されている最新の複数の次発話候補の内容データについての情報（先行情報）を記憶するものである。この先行情報は、入替要否判断手段４３Ｂによる判断処理を行う際に、先行する複数

の次発話候補の内容を把握するために利用される。

【０２０６】

具体的には、次発話候補初期準備手段４３Ａは、システム発話Ｓ（Ｎ＋１）の準備開始指示情報を受け取ったときに、対話履歴記憶手段５０に記憶されているそれまでの対話履歴情報（システム発話Ｓ（Ｎ）までの対話履歴情報）の少なくとも一部、すなわちＳ（１）、Ｕ（１）、Ｓ（２）、Ｕ（２）、…、Ｓ（Ｎ）（図６の最上部を参照）の少なくとも一部を用いて、題材データ記憶手段５１または題材データ提供システム６０に記憶されている題材データやその構成要素の中から、システムの複数の次発話候補Ｓ（Ｎ＋１）の内容データを選択取得する処理を実行する。但し、最初のシステム発話Ｓ（１）の準備開始指示情報を受け取ったときには、選択取得するＳ（１）は、１つだけでよい。また、常に複数の次発話候補を選択取得しなければならないわけではなく、選択取得した次発話候補が、結果的に１つになる場合があってもよい。

【０２０７】

一方、入替準備手段４３Ｃは、次発話候補記憶手段３０に複数の次発話候補Ｓ（Ｎ＋１）の内容データが既に記憶されている状態において、入替要否判断手段４３Ｂにより入替が必要であると判断した場合に、対話状態管理手段４２から逐次送られてくる音声認識処理手段４１による現在進行中のユーザ発話Ｕ（Ｎ）の音声認識処理の結果である言語情報を用いて、題材データ記憶手段５１または題材データ提供システム６０に記憶されている題材データやその構成要素の中から、システムの別の複数の次発話候補Ｓ（Ｎ＋１）の内容データを選択取得する処理を実行する。但し、常に複数の次発話候補を選択取得しなければならないわけではなく、選択取得した次発話候補が、結果的に１つになる場合があってもよいのは、上述した次発話候補初期準備手段４３Ａの場合と同様である。また、上記の説明では、現在進行中のユーザ発話Ｕ（Ｎ）の音声認識処理の結果を用いるとしているが、準備処理を行う時点では、対話状態管理手段４２から逐次送られてくる音声認識処理の結果がユーザ発話Ｕ（Ｎ）の発話区間全体における最後の部分（または発話区間全体）であるか否かは判らない場合があるので（あるいは、最後の部分であるか否かの区別をする必要はないので）、対話状態管理手段４２の説明で既に詳述したように、対話状態管理手段４２から受け取った音声認識処理の結果が、結果的に、ユーザ発話Ｕ（Ｎ）の発話区間全体における最後の部分（または発話区間全体）であった場合でも、この入替準備手段４３Ｃによる準備処理は実行される。この場合、結果的には、進行中のユーザ発話Ｕ（Ｎ）ではなく、対話履歴情報として対話履歴記憶手段５０に記憶されることになる発話終了後のユーザ発話Ｕ（Ｎ）の情報を用いていることになる。なお、入替準備手段４３Ｃは、ユーザ発話Ｕ（Ｎ）だけではなく、次発話候補初期準備手段４３Ａの場合と同様に、対話履歴情報Ｓ（１）、Ｕ（１）、…、Ｓ（Ｎ）を用いてもよい。

【０２０８】

また、用意されている題材データには、様々な状態のものがあり、例えば、テキストデータだけの場合、テキストデータおよびそれに対応する音声データがある場合、それらのテキストデータや音声データに、映像データや静止画データ、あるいは楽曲データが付随している場合、付随させる映像データや静止画データ、あるいは楽曲データだけの場合等があり、更にはテキストデータにも様々な語調のものがある。このため、次発話候補初期準備手段４３Ａおよび入替準備手段４３Ｃは、必要な場合には、テキストデータの加工調整（例えば、語尾の調整、一部削除、結合・分割・組替・その他の編集等）、テキストデータから音声データ（例えばｗａｖファイル等）を生成する音声合成、動画や静止画の画質・サイズ調整といった各種の生成処理も行う。但し、システムの応答性を向上させる観点からは、次発話候補の準備処理自体に時間がかかることを避ける必要がある。準備処理は、ユーザ発話中に行い、原則として、ユーザ発話が終了する前に準備が完了していることが好ましいからである。従って、テキストデータの語調等の加工調整、音声合成処理、動画や静止画の画質・サイズ調整等は、予め実行しておき、それらの処理を実行済の題材データを、題材データ記憶手段５１または題材データ提供システム６０に用意しておくことが好ましい。

【０２０９】
そして、次発話候補初期準備手段４３Ａによるシステム発話Ｓ（Ｎ＋１）の複数の候補の選択取得では、通常は、直前のシステム発話Ｓ（Ｎ）の内容が最も重要な選択用判断材料となるが、それよりも前のＳ（１）、Ｕ（１）、…、Ｓ（Ｎ－１）、Ｕ（Ｎ－１）も使用されることがある。例えば、シナリオデータ内における各構成要素（本実施形態では、主計画要素、副計画要素がある。）を予め定めた順序で再生していくときに、ユーザ発話の内容に応じて各構成要素の再生順序を変更する場合がある。この場合、例えば、１回再生した構成要素については、２度目の再生は行わないというルールがあれば、それまでにいずれの構成要素が再生されたのかを把握する必要があるので、それまでの対話履歴情報の全部を使用する必要がある。

【０２１０】
また、入替準備手段４３Ｃによるシステム発話Ｓ（Ｎ＋１）の複数の候補の選択取得では、ユーザ発話Ｕ（Ｎ）が、例えば「さっき言っていたＸＸＸ選手について、別の情報が知りたいな。」、「その話は知っているから、さっきのＹＹＹ事件の話を詳しく聞きたいな。」等であれば、選択用判断材料として、そのユーザ発話Ｕ（Ｎ）の情報を使用することは勿論であるが、直前のシステム発話Ｓ（Ｎ）を使用せず、それよりも前の情報Ｓ（１）、Ｕ（１）、…、Ｓ（Ｎ－１）、Ｕ（Ｎ－１）を使用する場合もある。つまり、少し前（例えば数分前等）の対話履歴情報に基づき、ＸＸＸ選手やＹＹＹ事件について、どこまで話していたのかを把握し、それとは別の情報を、題材データ記憶手段５１または題材データ提供システム６０から選択取得する場合等がある。

【０２１１】
また、題材データ記憶手段５１または題材データ提供システム６０に記憶されている題材データには、様々な種類のデータがあり、情報量の多少も異なっている。例えば、題材データがシナリオデータであれば、複数の構成要素により構成され、一方、題材データの中には、シナリオデータ内の１つの構成要素に相当するような比較的短い題材データも存在する。従って、次発話候補初期準備手段４３Ａや入替準備手段４３Ｃにより複数の次発話候補を「選択」することには、複数（多数）のシナリオデータの中から１つのシナリオデータを選択し、かつ、選択した１つのシナリオデータの中から１つ（Ｓ（１）の場合）または複数の構成要素を選択すること、既に選択されている１つのシナリオデータの中から複数の構成要素を選択すること、複数（多数）の比較的短い題材データの中から１つ（Ｓ（１）の場合）または複数の題材データを選択すること等が含まれる。

【０２１２】
また、次発話候補初期準備手段４３Ａおよび入替準備手段４３Ｃは、準備処理で得られた複数（結果的に１つの場合もあり、また、Ｓ（１）の場合は１つである。）の次発話候補の内容データまたはそれらに加えてそれらの識別情報（例えば、シナリオＩＤ、発話節ＩＤ等）を、ネットワーク１を介して再生装置２０へ送信し、次発話候補記憶手段３０に記憶させる処理も実行する。２回目以降は、更新処理である。この更新により、次発話候補記憶手段３０に記憶されている次発話候補の内容データの数は、ユーザ発話の進行に伴って、例えば、図７中の中央部に示すように、Ｎ１，Ｎ２．Ｎ３のように変化する。また、入替要否判断手段４３Ｂにより入替が必要であるという判断結果が出た場合に、入替準備手段４３Ｃによる準備処理が開始されるが、この準備期間中は、図７中の中央部に示すように、次発話候補記憶手段３０に記憶されている複数の次発話候補の内容データを削除し、次発話候補の内容データの数をゼロにクリアしてもよく、あるいは、削除せずに維持し、ゼロにクリアしない処理を行ってもよい。

【０２１３】
さらに、次発話候補初期準備手段４３Ａおよび入替準備手段４３Ｃは、準備状態を示すステータス、目的データの残数および次発話候補の重要度（準備処理で得られた複数の次発話候補の内容データの各々の重要度）を、ネットワーク１を介して再生装置２０へ送信し、システム情報記憶手段３１に記憶させる処理も実行する。２回目以降は、更新処理である。

【０２１４】
なお、目的データの残数は、対話目的を達成するためのシステムの最終の次発話候補の内容データとなり得る目的データの残数であるが、次発話候補記憶手段３０に記憶させる次発話候補の内容データの数とは異なる。例えば、情報検索対話で、ユーザが自分の利用する飲食店を探すときには、飲食店のデータが目的データとなる。しかし、条件提示による絞り込みが進んでいない段階では、目的データ（例えば、飲食店のデータ等）は多数存在し、それらの目的データの全部を、次発話候補の内容データとして次発話候補記憶手段３０に記憶させるわけではなく、情報検索対話の初期の段階や途中の段階では、次発話候補記憶手段３０には、「何を食べたいですか？」、「費用はどれぐらいですか？」等が記憶されるだけである。そして、絞り込みが進んだ段階や絞り込みが完了した最終段階で、目的データ（例えば、飲食店のデータ等）は、次発話候補の内容データとして次発話候補記憶手段３０に記憶されることになる。従って、目的データの残数は、潜在的な次発話候補の内容データの数である。

【０２１５】
先行次発話候補情報記憶手段４３Ｄには、例えば、次発話候補初期準備手段４３Ａおよび入替準備手段４３Ｃによる準備処理で得られた複数の次発話候補の内容データ（現在、次発話候補記憶手段３０に記憶されている最新の複数の次発話候補の内容データ）、それらの内容データについての各分野（例えば、ＩＴ・科学、テニス、野球等）、分野以外の属性（例えば、男性向け、１０代～３０代向け等）、それらの内容データに含まれる１つまたは複数の重要度の高い単語等が記憶されている。

【０２１６】
入替要否判断手段４３Ｂは、対話状態管理手段４２から逐次送られてくる音声認識処理手段４１によるユーザ発話Ｕ（Ｎ）の音声認識処理の結果である言語情報を受け取り、受け取った言語情報と、先行次発話候補情報記憶手段４３Ｄに記憶されている再生装置２０へ送信済の複数の次発話候補の内容データ（現在、次発話候補記憶手段３０に記憶されている最新の複数の次発話候補の内容データ）についての情報（先行情報）とを用いて、次発話の候補となる複数の次発話候補の内容データの少なくとも一部を入れ替えるか否かを逐次判定し、入れ替えると判定した場合には、その結果を入替準備手段４３Ｃに送る処理を実行する。

【０２１７】
具体的には、入替要否判断手段４３Ｂは、現在までに、図６に示すＵ（Ｎ，Ｋ）までの音声認識処理の結果を用いた入替要否判断処理およびそれに伴う入替準備処理が行われていたとすると、例えば、新たに出力されたＵ（Ｎ，Ｋ＋１）（但し、ショートセグメンテーションであるから、正確には、その一部）の音声認識処理の結果である言語情報の中に、重要度の高い単語が含まれているか否かを判断する。ここで、単語の重要度としては、例えば、ＴＦ（Term Frequency：文書における単語の出現頻度）およびＩＤＦ（Inverse Document Frequency：逆文書頻度）によるＴＦ－ＩＤＦ、Ｏｋａｐｉ－ＢＭ２５等を採用することができ、予め算出して単語重要度記憶手段（不図示）に記憶しておけばよい。

【０２１８】
そして、例えば、Ｕ（Ｎ，Ｋ＋１）（正確には、その一部）の中に、単語α，β，γが含まれていたとすると、これらの単語α，β，γの全ての重要度が、予め定めた重要度判定用閾値以下または未満であった場合（単語α，β，γがいずれも重要度の高い単語ではなかった場合）には、入替は不要であると判断すること等ができる。一方、単語α，β，γの中に重要度の高い単語が含まれていた場合には、その重要度の高い単語（例えば単語α）が、先行次発話候補情報記憶手段４３Ｄに記憶されている重要度の高い単語の中に含まれているか否かを判断し、含まれていれば、入替は不要であると判断すること等ができる。あるいは、単語α，β，γの中に重要度の高い単語が含まれていた場合には、その重要度の高い単語（例えば単語α）と、先行次発話候補情報記憶手段４３Ｄに記憶されている重要度の高い単語の各々との類似度を、例えばｗｏｒｄ２ｖｅｃやＧｌｏＶｅ等により求め、求めた各類似度のうちのいずれかが類似度判定用閾値以上または超過であった場合

（類似する重度語の高い単語があった場合）には、入替は不要であると判断すること等ができる。また、上記のように単語α，β，γの中の重要度の高い単語（例えば単語α）と、先行次発話候補情報記憶手段４３Ｄに記憶されている１つまたは複数の重要度の高い単語とを用いて判断するのではなく、単語α，β，γの中の重要度の高い単語（例えば単語α）と、先行次発話候補情報記憶手段４３Ｄに記憶されている複数の次発話候補の内容データの全体（それらに含まれる全ての単語）とを用いて判断してもよい。

【０２１９】

また、上記の例の単語α，β，γの中の重要度の高い単語（例えば単語α）が、いずれの分野に属する単語であるかを判断し、判断した分野が（複数の分野でもよく、その場合には、いずれかの分野が）、先行次発話候補情報記憶手段４３Ｄに記憶されている１つまたは複数の分野（通常は１つの分野であることが多い。）の中に含まれていれば、入替は不要であると判断すること等ができる。なお、各単語（重要度の高い単語）と各分野（例えば、ＩＴ・科学、テニス、ゴルフ、エンタメ、政治経済、国際等）との対応関係は、予め定めて単語帰属分野記憶手段（不図示）に記憶しておけばよく、１つの単語が複数の分野に帰属していてもよい。この対応関係は、例えば、各分野の文書における各単語の出現頻度や、累積出現回数等により定めることができる。

【０２２０】

なお、題材データまたはその構成要素には、分野の識別情報が関連付けられている。分野の粒度は、システム設計者が適宜定めればよく、例えば、テニス、ゴルフ、野球等を別々の分野とするか、スポーツで１つの分野にまとめるか、あるいは、政治、経済を別々の分野とするか、１つにまとめるか等は任意である。１つの題材データまたはその構成要素は、複数の分野に帰属していてもよい。また、題材データまたはその構成要素が、女子プロゴルフの話題のみである場合に、例えば、女子プロゴルフ＜ゴルフ＜スポーツのように、包含関係にある分野の識別情報を全て関連付けるようにしてもよい。

【０２２１】

また、音声対話には、各種の目的の対話（例えば、ニュース対話、アンケート対話、ガイダンス対話、情報検索対話、操作対話、教育対話、情報特定対話等）があり、対話の進行も各種のタイプのものがある。対話の進行のタイプとの関係では、次のようになる。

【０２２２】

次発話候補初期準備手段４３Ａは、シナリオデータ（主計画および副計画を有する複雑な分岐を行うシナリオに限らず、より単純なシナリオも含む。）があり、シナリオとして予め定められた順序に従って対話を進めていく場合には、そのシナリオの順序に従って、複数の次発話候補を選択していく。この場合、入替準備手段４３Ｃにより、予め定められた順序が変更された場合には、その変更を反映させ、例えば、１回再生したシナリオ構成要素については、２度目の再生は行わないというルールがあれば、そのルールに従いつつ、当初の順序をなるべく維持した順序で複数の次発話候補を選択していく。

【０２２３】

また、シナリオデータがなく、対話の進行や分岐のパターンが予め定まっているわけではないが、システム発話の内容については、予定外の情報を外部システムから取得しなければならない場合を除き、予め用意されていて、毎回のユーザ発話の内容に従って、その都度、次のシステム発話の内容を定める場合がある。このような場合には、直前のシステム発話Ｓ（Ｎ）で、次のシステム発話Ｓ（Ｎ＋１）の複数の候補が定まることは少ない。なぜなら、Ｓ（Ｎ）でＳ（Ｎ＋１）の候補が定まるということは、結局、広い意味で、または部分的にシナリオが形成されていると考えることができるので、シナリオがない場合に該当しないからである。

【０２２４】

例えば、自動車のカーナビ操作のための操作対話において、システム発話Ｓ（Ｎ）＝「住所で目的地を設定しますか？」であった場合、次発話候補初期準備手段４３Ａにより、システム発話Ｓ（Ｎ＋１）＝「最初に都道府県を教えてください。」、「市町村を教えてください。」、「何丁目何番地ですか？」等を準備して次発話候補記憶手段３０に記憶さ

せておく。そして、ユーザ発話Ｕ（Ｎ）が「はい。」であれば、次発話選択手段２４により「最初に都道府県を教えてください。」を選択して発話生成手段２５によりそれを再生し、「はい、東京都です。」であれば、「市町村を教えてください。」を選択して再生し、「はい、東京都新宿区です。」であれば、「何丁目何番地ですか？」を選択して再生する。この際、ユーザ発話Ｕ（Ｎ）が「はい、東京都新宿区です。」の途中の「はい、東京都…」まで進行した段階で、入替準備手段４３Ｃにより、システム発話Ｓ（Ｎ＋１）＝「市町村を教えてください。」、「何丁目何番地ですか？」等への入替が行われる場合（「最初に都道府県を教えてください。」が次発話候補から除かれている場合）もある。このような場合、部分的にシナリオが形成されていると考えることができ、そのシナリオに従って次発話候補初期準備手段４３Ａによる準備処理が行われているが、入替準備手段４３Ｃによる役割も大きい。

【０２２５】

一方、シナリオデータがない場合は、次発話候補初期準備手段４３Ａによる準備処理よりも、入替準備手段４３Ｃによる準備処理が中心となる。最初はシナリオがあり、その後、フリートークに近い状態になる場合の後半の処理も同様である。そして、シナリオデータがない場合、次発話候補初期準備手段４３Ａは、対話状態管理手段４２からのＳ（Ｎ＋１）の準備開始指示情報を受け取ったときに、システム発話Ｓ（Ｎ＋１）の複数の候補を定めることができなければ、準備中のステータス（ステータス＝次発話候補検討中）を、ネットワーク１を介して再生装置２０へ送信してシステム状態記憶手段３１に記憶させ、入替準備手段４３Ｃに準備処理を任せることができる。このようにした場合は、入替準備手段４３Ｃによる準備処理は、入替の準備というより初期データの準備となるので、入替準備手段４３Ｃは、入替要否判断手段４３Ｂからの入替が必要であるという判断結果を受け取って準備処理を開始するのではなく、次発話候補初期準備手段４３Ａから転送されてくる準備開始指示情報を受け取って準備処理を開始することになる。この場合、入替要否判断手段４３Ｂによる判断処理は行われないので、入替準備手段４３Ｃは、対話状態管理手段４２から逐次送られてくる音声認識処理の結果を受け取り、入替要否判断手段４３Ｂによる重要度の高い単語の抽出処理に相当する処理を実行するが、この際の重要度判定用閾値は低く設定してもよい。なお、重要度判定用閾値を低く設定しても、重要度の高い単語が抽出されない場合には、進行中のユーザ発話Ｕ（Ｎ）の中に、未だシステム発話Ｓ（Ｎ＋１）の複数の候補の決定をするのに十分な情報（単語）が現れていないことになるので、ユーザ発話Ｕ（Ｎ）の進行を待つことになる。入替準備手段４３Ｃは、以上のような次発話候補初期準備手段４３Ａから転送されてくる準備開始指示情報を受け取った場合の初期データの準備処理を行い、複数の次発話候補の内容データを次発話候補記憶手段３０に記憶させた後には、通常通りの入替の準備処理（重要度判定用閾値も通常の設定とする。）を実行する。

【０２２６】

また、シナリオデータがない場合は、フリートークの状態に近いと考え、次発話候補初期準備手段４３Ａは、とりあえず、題材データ記憶手段５１や題材データ提供システム６０に記憶されている題材データの中からランダムに選択取得した複数の次発話候補の内容データを、ネットワーク１を介して再生装置２０へ送信して次発話候補記憶手段３０に記憶させてもよい。ランダムな選択取得を行っても、その後、ユーザ発話Ｕ（Ｎ）が進行すると、入替準備手段４３Ｃによる準備処理が行われ、複数の次発話候補の内容データが適切なものに入れ替えられる。仮に、ユーザ発話Ｕ（Ｎ）が進行しても、ランダムに選択取得した複数の次発話候補の内容データがそのまま維持されていたとすると、そのランダムな選択取得が適切であったということになる。なお、ランダムに選択取得する題材データが存在する（選択取得する範囲が定まっている）ということは、完全なフリートークではなく、システム発話の内容は、想定範囲外の情報を外部システムから取得しなければならない場合を除き、予め用意されていることになる。

【０２２７】

前述したように、入替要否判断手段４３Ｂは、対話状態管理手段４２から逐次送られて

くる新たな音声認識処理の結果を受け取り、この結果に含まれる単語のうち予め定められた重要度の高い単語を抽出する処理を実行するので、入替準備手段４３Ｃは、入替要否判断手段４３Ｂから、抽出された重要度の高い単語を受け取る。そして、入替準備手段４３Ｃは、この重要度の高い単語を用いて、予め定められた各単語と各分野との対応関係（単語帰属分野記憶手段（不図示）に記憶されている情報）から、ユーザの関心のある話題（分野）を決定し、題材データ記憶手段５１または題材データ提供システム６０に記憶されている題材データの中から、決定した話題（分野）に関連付けられて記憶されている題材データを選択し、次発話の候補となる別の複数の次発話候補の内容データを取得または生成する準備処理を実行することができる。

【０２２８】

例えば、システム発話Ｓ（Ｎ）が「早稲田太郎選手が４回転フリップを成功させたよ。」であり、システム発話Ｓ（Ｎ＋１）の複数の候補として、「グランプリシリーズのカナダ大会で跳んだそうだ。」（主計画要素）、「早稲田太郎選手は、…」という早稲田太郎の人物の説明データ（副計画要素）、「４回転フリップっていうのは、…」という４回転フリップの技の説明データ（副計画要素）が、次発話候補記憶手段３０に記憶されているとする。このとき、ユーザ発話Ｕ（Ｎ）が「フィギュアスケートは興味がないので、野球の話が聞きたいんだけど…」、「つまらない、野球の方がおもしろいから…」であった場合には、入替準備手段４３Ｃは、再生中のシナリオデータ（分野＝アイススケート、または、分野＝スポーツ、アイススケート）の中に野球の話は全くないので、シナリオデータ自体を別の分野（この例では、野球の分野）に入れ替え、その入替後のシナリオデータ内の先頭の構成要素を、次発話候補とすることができる。なお、この場合は、Ｓ（Ｎ＋１）の候補ではあるが、シナリオデータの先頭からの再生となるので、Ｓ（１）と同等であるから、次発話候補は１つでよい。また、この場合、「興味がない」、「つまらない」、「もう飽きた」、「くだらない」、「話題を変えてほしい」、「その話はもういい」、「その話はやめて」、「ところで」、「話は変わるけど」、「そういえば」等の話題転換要求を伴っているため、シナリオデータ自体を入れ替える話題転換処理を行っている。また、明確な話題転換要求が無くても、例えば、ユーザ発話Ｕ（Ｎ）が「来週、日米野球があるけど、高校野球の…」のように、再生中のシナリオデータ内にない単語が繰り返される場合にシナリオデータ自体を入れ替える話題転換処理を行ってもよい。

【０２２９】

一方、上記の例において、ユーザ発話Ｕ（Ｎ）が「早稲田次郎選手の方が好きなんだけど、早稲田次郎選手の成績は…」であった場合には、入替準備手段４３Ｃは、再生中のシナリオデータ内に早稲田次郎選手についての構成要素（主計画要素）も含まれているので、シナリオデータの入替は行わずに、同じシナリオデータ内での再生順序の変更を行う。例えば、「早稲田次郎」によるキーワードマッチングで「早稲田次郎選手は、４回転アクセルに挑戦したけど失敗したんだ。」（主計画要素）を選択するとともに、「早稲田次郎選手は、…」という早稲田次郎の人物の説明データ（副計画要素）、「４回転アクセルっていうのは、…」という４回転アクセルの技の説明データ（副計画要素）を選択し、次発話候補とする。

【０２３０】

また、次発話候補初期準備手段４３Ａや入替準備手段４３Ｃによる準備処理において、位置データや時刻データを反映させてもよい。例えば、博物館や遺跡等の案内を行うガイダンス対話では、ユーザの位置データ（例えば、再生装置２０に設置されたＧＰＳ受信機や、再生装置２０が本体と端末とに分割されている場合の端末に設置されたＧＰＳ受信機で得られる位置データ等）を用いて、複数の次発話候補の内容データが定まるようにしてもよい。例えば、博物館のガイダンス対話において、予め登録されて対話サーバ４０のメモリに記憶されている展示物Ｘの位置データと、再生装置２０からネットワーク１を介して送信されてきたユーザの位置データとを用いて、ユーザが展示物Ｘのそばに近づいたことを検出し、さらに時刻データを用いて１２時近くであることを検出した時点で、「そろそろ展示物Ｘが見えてきます。」というシステム発話Ｓ（Ｎ）を行い、次発話候補Ｓ（Ｎ

＋１）として「展示物Ｘは、・・・」、「食堂をご案内しましょうか。」等を用意して次発話候補記憶手段３０に記憶させておく。そして、ユーザ発話Ｕ（Ｎ）が「展示物Ｘの説明を聞きたいな。」であった場合には、「展示物Ｘは、・・・」を選択して再生し、「お腹すいた。」であった場合には、「食堂をご案内しましょうか。」を選択して再生する。また、時刻データを用いて、「外が暗くなってきたから、そろそろ帰り支度を始めましょう。」、「閉館時間が迫っているから、素早く回ろうね。」等を準備して次発話候補記憶手段３０に記憶させておくこともできる。

【０２３１】

さらに、次発話候補初期準備手段４３Ａや入替準備手段４３Ｃによる準備処理において、位置データや時刻データ以外の状態データ（変化の速度の大小の相違はあるが、原則として、時々刻々と変化するデータ）、例えば、温度データ、湿度データ、天候データ、高度データ等を用いて次発話候補の内容データを準備してもよい。例えば、「今日は暑いね。」、「今日は蒸すね。」、「今日は天気がいいね。」、「空気が薄くなってきたけど、大丈夫？」等の次発話候補を準備し、次発話候補記憶手段３０に記憶させておくことができる。なお、上記の例の天候データは、選択用判断材料としての天候データ（ユーザが操作する再生装置２０の所在地における晴・雨・曇り等のデータ）であるから、題材データとして用意されている「台風２８号が沖縄地方に接近しています。」、「ＸＸ地方に大雨洪水警報が出ていますので、ＹＹ川の氾濫に注意してください。」等の警報データとは異なる。つまり、例えば、雨という天候データに基づき、「今日は雨だけど、東京ドームは屋根があるから、野球の観戦はできるよ。」等の題材データが選択取得され、次発話候補記憶手段３０に記憶されることになる。

【０２３２】

また、入替準備手段４３Ｃによる準備処理は、必ずしも複数の次発話候補の全部を入れ替える必要はなく、少なくとも一部の入替が行われればよい。例えば、最初の複数の次発話候補の内容データ（初期データ）または前回の入替後のデータが、次発話候補Ａ，Ｂ，Ｃであったとすると、入替後の次発話候補は、Ｄ，Ｅ，Ｆのように全部が入れ替わっていてもよく、Ａ，Ｂ，Ｄのように一部が入れ替わっていてもよい。また、入替後の次発話候補は、Ａ，Ｂ，Ｃ，Ｄ，Ｅのように候補が追加されて増えた状態となっていてもよく、Ａ，Ｂのように一部削除された状態となっていてもよい。

【０２３３】

さらに、対話状態管理手段４２は、ユーザ発話Ｕ（Ｎ）の進行中において、各時点において、それまでのＵ（Ｎ）の全部を保持しているので、対話状態管理手段４２から入替要否判断手段４３Ｂに送る入替要否判断に用いるための音声認識処理の結果の長さは、自在に調整することができる。従って、ショートポーズセグメンテーションの単位の最新の音声認識結果だけとしてもよく、最新の音声認識結果を含めたある程度の時間長の音声認識結果としてもよく、対話状態管理手段４２に保持されているＵ（Ｎ）の音声認識結果の全部としてもよい。

【０２３４】

＜対話サーバ４０／対話履歴記憶手段５０の構成＞

【０２３５】

対話履歴記憶手段５０は、システムとユーザとの間の対話履歴情報を記憶するものである。具体的には、図６の最上部に示すように、システム発話Ｓ（１）の内容データ（テキストデータ）、ユーザ発話Ｕ（１）の内容データ（テキストデータ）、同様に、Ｓ（２）、Ｕ（２）、Ｓ（３）、Ｕ（３）、…の各内容データ（テキストデータ）を、対話の順番に記憶する。ユーザ発話から始まっていてもよい。なお、進行中のユーザ発話Ｕ（Ｎ）は、本実施形態では、対話状態管理手段４２のメモリ（主メモリでよい）に記憶され、発話の終了後に、発話区間全体が対話履歴記憶手段５０に記憶される。

【０２３６】

＜対話サーバ４０／題材データ記憶手段５１の構成＞

【０２３７】

題材データ記憶手段５１は、題材データを記憶するものである。題材データは、例えば、シナリオデータ（主計画および副計画を有する複雑な分岐を行うシナリオに限らず、より単純なシナリオも含む。）、シナリオが形成されていない最近のトピックを集めた各種の話題データの集合（但し、話題データの１つ１つが、独立した題材データであり、それぞれ比較的短いデータである。）、辞書データ、事典データ、機器の使用方法や施設等のガイダンス用データ、アンケート調査用データ、機器や装置等の操作補助用データ、教育用データ等である。これらの題材データまたはその構成要素には、分野（例えば、ＩＴ・科学、政治・経済、国際、エンタメ、相撲、ゴルフ等）の識別情報が関連付けられている。なお、分野が定められていない題材データまたはその構成要素が混在していてもよいが、その場合は、キーワードマッチングにより、必要な情報を選択取得する。題材データ提供システム６０も同様である。

【０２３８】

＜対話サーバ４０／ユーザ情報記憶手段５２の構成＞

【０２３９】

ユーザ情報記憶手段５２は、ユーザ発話とシステム発話との衝突の発生情報、システムの交替潜時、およびユーザの発話速度を、ユーザ識別情報と関連付けて記憶するものである。このユーザ情報記憶手段５２に記憶される情報は、各ユーザとの複数回の対話を通じて得られたユーザ毎の蓄積情報であるから、ユーザの属性情報である。従って、ユーザとの対話中における一時的な情報ではないので、ユーザ状態記憶手段３２に記憶される情報とは異なる。これらの衝突の発生情報、システムの交替潜時、およびユーザの発話速度は、いずれも発話生成手段２５により得られて記録されたものである。

【０２４０】

＜ユーザからシステムへの話者交替時の処理の流れ：図５＞

【０２４１】

このような本実施形態においては、以下のようにしてユーザからシステムへの話者交替が行われる。

【０２４２】

図５において、先ず、対話開始前に、システム発話タイミング検出手段２２のユーザ発話権終了判定用閾値調整手段２２Ｇ（図２参照）により、ユーザ情報記憶手段５２から、対話相手のユーザについてのユーザ識別情報（ユーザＩＤ）を用いて、衝突の発生情報（蓄積情報）、交替潜時（蓄積情報）、および発話速度（蓄積情報）を取得し、衝突の発生情報（蓄積情報）によるユーザ発話権終了判定用閾値の事前調整（図１１参照）、発話速度（蓄積情報）による下方調整用閾値の事前調整およびその下方調整用閾値を用いた交替潜時（蓄積情報）によるユーザ発話権終了判定用閾値の事前調整（図１２参照）を行う（ステップＳ１）。

【０２４３】

次に、対話開始後においては、音声信号取得手段２１により取得したユーザの音声信号を用いて、システム発話タイミング検出手段２２の音響特徴量抽出手段２２Ａ（図２参照）により、周波数分析等を行って音響特徴量を抽出する（ステップＳ２）。また、必要な場合には、音声認識処理手段４１により得られた音声認識処理の結果である言語情報を用いて、言語特徴量抽出手段２２Ｂ（図２参照）により言語特徴量を抽出する。

【０２４４】

続いて、システム発話タイミング検出手段２２のユーザ発話権終了判定用閾値調整手段２２Ｇ（図２参照）により、ユーザ状態記憶手段３２からユーザ発話継続時間を取得し、ユーザ発話権終了判定用閾値のリアルタイム調整（図９参照）を行うとともに、システム状態記憶手段３１からシステム発話意欲度の指標値（目的データの残数および／または次発話候補の重要度）を取得し、ユーザ発話権終了判定用閾値のリアルタイム調整（図１０参照）を行う（ステップＳ３）。なお、これらの２種類のリアルタイム調整は、いずれか一方の調整を行ってもよく、双方の調整を行ってもよく、双方の調整を行う場合は、どちらの調整を先に行ってもよい。

【０２４５】
それから、システム発話タイミング検出手段２２のユーザ発話権終了判定用パターン認識器２２Ｃ（図２参照）により、ステップＳ２で抽出した音響特徴量、またはこれに加えて言語特徴量を用いて、ユーザ発話権の維持または終了を識別するパターン認識処理を実行し、このパターン認識処理で得られる尤度を用いたユーザ発話権終了判定用閾値による閾値判定を行い、維持または終了の識別結果を出力する（ステップＳ４）。
【０２４６】
また、これと並行して、次発話選択用情報生成手段２３により、ステップＳ２で得られた音響特徴量を用いた韻律分析を行い、ユーザ発話意図の識別処理を行う（ステップＳ５）。なお、図５中の２点鎖線で示すように、音声信号取得手段２１により取得したユーザの音声信号を用いて、ステップＳ２とは別途に韻律特徴量を抽出し、その韻律特徴量を用いた韻律分析を行い、ユーザ発話意図の識別処理を行ってもよい。
【０２４７】
そして、前述したステップＳ４の識別結果が維持または終了のいずれであるかを判断し（ステップＳ６）、維持であった場合には、ステップＳ２の処理に戻る。一方、識別結果が終了であった場合には、システム発話タイミング検出手段２２のシステム発話開始タイミング判断手段２２Ｆ（図２参照）により、システム状態記憶手段３１から、準備の状態を示すステータス（準備完了・各種の準備中の別）を取得し、図８に示す流れで、システム発話の開始タイミングであるか否かの判断結果を出力する（図５のステップＳ７）。
【０２４８】
ここで、出力された判断結果が、システム発話の開始タイミングであるか否かに応じ（ステップＳ８）、システム発話の開始タイミングではなかった場合には、ステップＳ２の処理に戻る。一方、システム発話の開始タイミングであった場合（当該タイミングが検出された場合）には、次発話選択手段２４により、前述したステップＳ５で次発話選択用情報生成手段２３により得られたユーザ発話意図の識別結果（質問、相槌等の別）および／または音声認識処理手段４１による音声認識処理の結果である言語情報（文字列）を用いて、次発話候補記憶手段３０に記憶されている複数（但し、１つである場合もある）の次発話候補の内容データの中から、次発話の内容データを選択するとともに、選択した次発話の内容データまたはその識別情報（シナリオＩＤ、発話節ＩＤ等）を、ネットワーク１を介して対話サーバ４０の対話状態管理手段４２へ送信する（ステップＳ９）。
【０２４９】
このステップＳ９の処理を行う際、次発話候補記憶手段３０には、図５の流れとは非同期で行われる次発話準備手段４３の準備処理により、次発話選択手段２４によるステップＳ９の処理に先んじて用意された複数（但し、１つである場合もある）の次発話候補の内容データが既に記憶されている状態である。なお、準備中の場合には、フィラーの挿入が行われる（図８のＰ１０参照）。
【０２５０】
また、次発話選択手段２４による選択結果が、ネットワーク１を介して対話状態管理手段４２へ送信されると、対話履歴記憶手段５０に記憶された対話履歴情報が更新されるとともに、次発話準備手段４３に対してさらにその次の次発話候補の準備開始指示情報が出され、対話履歴記憶手段５０の情報を用いたさらにその次の次発話候補の準備処理が開始されるので、この意味では、図５の流れと、次発話準備手段４３の処理とは一連の流れのように見える。しかし、図４に示すように、次発話準備手段４３は、図５の中心であるシステム発話タイミング検出手段２２の処理とは非同期で行われる音声認識処理手段４１による音声認識処理の結果（言語情報）を用いた準備処理を実行するので、結局、次発話準備手段４３の処理は、図５の一連の処理の流れの中に記載することはできない。
【０２５１】
その後、発話生成手段２５により、次発話選択手段２４により選択された次発話の内容データを用いて、システム発話の音声信号の再生処理が行われる（図５のステップＳ１０）。また、付随する映像データ、静止画データ、楽曲データがあれば、それらの再生処理

も行われる。なお、必要な場合には、ここでの音声合成処理も行われるが、次発話候補記憶手段３０に記憶された次発話候補の内容データには、音声合成処理で得られた音声データも含まれていることが好ましい。

【０２５２】

そして、対話終了であるか否かを判断し（ステップＳ１１）、対話終了でない場合には、ステップＳ２の処理に戻る。

【０２５３】

＜シナリオのデータ構成、シナリオの再生、および次発話候補の準備処理の流れ：図１３～図１７＞

【０２５４】

＜シナリオのデータ構成：図１３＞

【０２５５】

図１３には、主計画および副計画からなるシナリオのデータ構成の具体例が示されている。このようなシナリオデータは、非特許文献４に記載されたシナリオデータと同様のものであり、対話システム１０で利用することができる題材データの一種として題材データ記憶手段５１に記憶されている。

【０２５６】

より詳細には、このシナリオデータは、ニュースやコラムや歴史等の各種の話題を記載した記事データ（文書データ）から生成したものであり、元の文書データを構成する要素を、元の文書データの内容の要約となる主計画要素と、この主計画要素を補完する副計画要素と、これら以外の要素（省略要素）とに分割し、これらの３種類の要素のうちの主計画要素および副計画要素、並びに、発話計画情報（主計画要素の再生順序および副計画要素への分岐を定めた情報）を含むように構成したものである。なお、元の文書データからのシナリオ生成時に、結果的に省略要素が発生しなくてもよい。つまり、主計画要素を除いた残り全ての要素が、副計画要素に割り当てられてもよい。

【０２５７】

主計画要素は、元の文書データを要約し、口語化することにより生成される。文書の要約は、重要文抽出、整列、文圧縮の処理を経て行われる。先ず、重要文抽出で、文書の要点となる情報を文単位で大まかに抽出する。次に、整列を行い、抽出した重要文の提示順序を決定する。そして、文圧縮を行い、文自体を短く縮約する。最後に、口語化処理を行い、書き言葉から会話表現への書き換えを行う。なお、このシナリオ生成時における重要文抽出で考慮される文の重要度は、前述したシステム発話タイミング検出手段２２のユーザ発話権終了判定用閾値調整手段２２Ｇの説明で既に詳述した通り、システム状態記憶手段３１に記憶される次発話候補の内容データの重度度とは異なるものであり、防災関連情報の緊急性や日常生活への影響の大きさ等を加味した重要度ではない。

【０２５８】

副計画要素は、主計画要素の情報を補うためのシステム発話の計画要素である。この副計画要素には、主計画要素で省かれた内容に基づく補足説明データ、予想される質問に対する回答データが含まれる。ユーザ発話の内容に応じて、副計画要素が再生されることになる。副計画要素についても、文圧縮と口語化の処理を行う。

【０２５９】

図１３において、シナリオを構成するデータ（カラム）には、元の記事（文書）についての文書ＩＤ、元の文書を構成する段落についての段落ＩＤ、元の文書を構成する文の重要度（シナリオ生成時に考慮した文の重要度に、防災関連情報の緊急性や日常生活への影響の大きさ等を加味した重要度であり、システム状態記憶手段３１に記憶される対象となる重要度である。）、元の文書を構成する文の内容を伝達したか否かの情報（未伝達・伝達済の別）、元の文書を構成する段落内の文についての文ＩＤ、元の文書を構成する段落内の文を構成する文節についての文節ＩＤ、シナリオの構成要素として選択されたか否かを示す情報（選択文節）、元の文書を構成する文内での文節提示順序、シナリオ再生を行うための発話節ＩＤ、リンクする発話節の合成音声ファイル（ｗａｖファイル等）の名称

、口語表現、文節間の間（ま）、元の文節の内容、ユーザの定義型質問に対する応答用の定義の文字情報、リンクする定義の合成音声ファイル（ｗａｖファイル等）の名称、トリビアの文字情報等が含まれる。また、図１３での図示は省略されているが、リンクするトリビアの合成音声ファイル（ｗａｖファイル等）の名称も含まれ、さらに、口語表現は、複数段階の表現（例えば、伝聞口調・断定口調を使い分ける「標準」、伝聞口調だけの「伝聞」、断定口調だけの「断定」、ですます調だけの「敬体」等の口調の別を含む）が用意されている。

【０２６０】

なお、合成音声ファイルの名称は、上記のように、シナリオを構成するデータであるが、その他に、合成音声ファイル自体（自体とは、ファイルの名称を示す文字情報ではなく、音声データを記録しているファイルという意味）を、予め生成してシナリオデータに含めてもよく、そうすることにより、次発話準備手段４３による準備時や、発話生成手段２５による再生時に音声合成処理を行う必要がなくなるので、システムの応答性を向上させることができる。

【０２６１】

また、上記の例では、重要度（システム状態記憶手段３１に記憶される対象となる重要度）は、元の文書を構成する文を単位とする重要度とされているが、元の文書を構成する文の単位ではなく、より細かく発話節毎に設定してもよい。

【０２６２】

さらに、元の文書を構成する文の内容を伝達したか否かの情報（未伝達・伝達済の別）は、対話履歴記憶手段５０に記憶されている対話履歴情報（図６の最上部のＳ（１），Ｓ（２），Ｓ（３），…）と同期して更新されるが、この情報も、発話節毎に持たせてもよい。この未伝達・伝達済の別は、対話の進行に伴って逐次更新されるので、題材データ記憶手段５１に記憶されているシナリオデータを直接に書き換えるわけではなく、対話状態管理手段４２のメモリ（主メモリでよいが、不揮発性メモリでもよい。）にコピーされて保持されているシナリオデータを書き換える。題材データ記憶手段５１に記憶されている当該シナリオデータは、同時期に他のユーザとの対話で使用されることもあるからである。主計画および副計画を備えていない他のタイプのシナリオデータの場合も同様である。

【０２６３】

＜シナリオデータを用いた音声対話の進行の概要：図１４＞

【０２６４】

図１４には、図１３のシナリオデータを用いてシステムとユーザとの間で行われる音声対話の進行の概要が示されている。但し、次発話準備手段４３による次発話候補の準備処理等の詳細は、図１５および図１６を用いて後述するので、ここでは表面的に表れる発話だけで対話の進行を説明する。

【０２６５】

先ず、１番目の主計画要素である「α社が３ＤＳ向けにＳｕｉｃａとかと連携するゲームソフトを開発してるらしいよ」という発話節（文書ＩＤ＝１、段落ＩＤ＝１、文ＩＤ＝１における発話節ＩＤ＝１：合成音声ファイル＝１－１－１－１．ｗａｖ）が再生される。この発話節の途中で（例えば「α社が」の再生直後に）、ユーザから「α社って、どんな会社なの？」という定義型質問（割込み）があった場合には、「α社は」という元の文節に対して予め用意されている定義型質問応答の副計画要素「α社は、・・・」が再生される。また、発話節の途中における別の位置で（例えば「ゲームソフトを」の再生直後に）、「α社って、どんな会社なの？」という割込みがあった場合でも、「α社は」という元の文節に対して予め用意されている定義型質問応答の副計画要素「α社は、・・・」が再生される。ユーザの割込みを受けた後のシステム発話の戻りの再生開始位置は、図１４中の実線で示すように、割込みを受けた位置でもよく、図１４中の点線で示すように、幾つか前の文節からの再開でもよく、発話節の先頭からの再開でもよい。なお、発話節の再生の終了直後に、ユーザから「α社って、どんな会社なの？」という定義型質問（割込み）があった場合も、同様に定義型質問応答の副計画要素「α社は、・・・」が再生される

ので、途中であっても、終了直後であっても、同じシステム応答となる。
【０２６６】
次に、ユーザ発話が「楽しみだね。」であったとすると、２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」という発話節（文書ＩＤ＝１、段落ＩＤ＝１、文ＩＤ＝２における発話節ＩＤ＝１：合成音声ファイル＝１－１－２－１．ｗａｖ）が再生され、その再生の終了直後に、ユーザから「ＩＣカードって、何？」という定義型質問があった場合には、「ＩＣカードから」という元の文節に対して予め用意されている定義型質問応答の副計画要素「ＩＣカードっていうのは、・・・」が再生される。
【０２６７】
続いて、３番目の主計画要素である「ゲーム内で使えるポイントが手に入るんだって」という発話節（文書ＩＤ＝１、段落ＩＤ＝１、文ＩＤ＝２における発話節ＩＤ＝２：合成音声ファイル＝１－１－２－２．ｗａｖ）が再生され、その再生の終了直後に、ユーザから「へー、そうなんだ。」という反応があった場合には、４番目の主計画要素（図１３での図示は省略）が再生される。
【０２６８】
さらに、副計画要素の再生中に、ユーザからの割込みがあれば、別の副計画要素が再生されるので、副計画要素の再生は、ユーザの反応次第で階層的になることがある。また、ユーザ発話の内容次第では、副計画要素としてシナリオ内に用意していない計画外の回答を行うこともある。
【０２６９】
以上が対話の進行の概要であるが、以上のような対話を実現するために、対話システム１０は、具体的には、例えば、図１５～図１７に示すような各処理を実行する。但し、図１７は、図１３のシナリオデータではなく、主計画および副計画からなる同型のシナリオデータを用いている。
【０２７０】
＜次発話候補の準備処理の具体例（１）：図１５＞
【０２７１】
図１５において、次発話準備手段４３は、対話状態管理手段４２からのシステム発話Ｓ（１）の準備開始指示情報を受け取り、次発話候補（但し、ここでは最初の発話）の準備処理を行う。対話開始時であるから、複数の次発話候補を選択取得するのではなく、シナリオデータ内から１番目の主計画要素を選択取得する。従って、次発話準備手段４３は、図１３のシナリオデータから、Ｓ（１）＝「α社が３ＤＳ向けにＳｕｉｃａとかと連携するゲームソフトを開発してるらしいよ。」を選択取得し、これを次発話候補記憶手段３０に記憶させる。
【０２７２】
続いて、ユーザ発話は未だ無い状態なので、システム発話タイミング検出手段２２により、直ぐにシステム発話の開始タイミングが検出され、次発話選択手段２４により、次発話候補記憶手段３０に記憶されているＳ（１）＝「α社が３ＤＳ向けにＳｕｉｃａとかと連携するゲームソフトを開発してるらしいよ。」が選択されるとともに、その選択結果がネットワーク１を介して対話状態管理手段４２に送信される。対話状態管理手段４２は、受け取ったＳ（１）を対話履歴記憶手段５０に保存するとともに、Ｓ（２）の準備開始指示情報を次発話準備手段４３に送る。
【０２７３】
それから、発話生成手段２５により、次発話選択手段２４により選択されたＳ（１）＝「α社が３ＤＳ向けにＳｕｉｃａとかと連携するゲームソフトを開発してるらしいよ。」の再生が開始される。また、これと並行して、次発話準備手段４３により、Ｓ（２）の準備処理が進む。ここで準備されるＳ（２）の次発話候補は、２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」と、現在再生中の１番目の主計画要素に対するユーザの定義型質問に答えるために用意する定義型質問応答の副計画要素である「α社は、…」、「β社は、…」、「β社３ＤＳっていうのは、…」、「交通系ＩＣカードっ

ていうのは、…」、「Ｓｕｉｃａっていうのは、…」、「連携っていうのは、…」、「ゲームソフトっていうのは、…」、「開発っていうのは、…」、「発表っていうのは、…」と、現在再生中の１番目の主計画要素に対するユーザの補足要求に応えるために用意する補足説明用の副計画要素（トリビア）である「Ｓｕｉｃａの名称は…」、「開発は、もともと…」等であり、これらが次発話候補記憶手段３０に記憶される。なお、「β社は、…」という定義型質問応答の副計画要素は、１番目の主計画要素の発話節として選択されなかった元の文節「β社の」について用意された情報であるが、ユーザの連想が及ぶ範囲であるため、ここでは次発話候補としている。

【０２７４】

その後、ユーザ発話Ｕ（１）＝「楽しみだね。」であったとすると、この場合は、次発話選択用情報生成手段２３により、ユーザ発話意図として、例えば「理解」等の識別結果が得られるので、次発話選択手段２４により選択される次発話は、Ｓ（２）＝２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」となり、この選択結果がネットワーク１を介して対話状態管理手段４２に送信される。対話状態管理手段４２は、選択結果としてＳ（２）を受け取ると、自身のメモリに保持しているＵ（１）（音声認識処理手段４１による音声認識処理の結果として受け取り、保持している文字列）と、受け取ったＳ（２）とを対話履歴記憶手段５０に保存するとともに、Ｓ（３）の準備開始指示情報を次発話準備手段４３に送る。

【０２７５】

それから、発話生成手段２５により、次発話選択手段２４により選択されたＳ（２）＝２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」の再生が開始される。また、これと並行して、次発話準備手段４３により、Ｓ（３）の準備処理が進む。ここで準備されるＳ（３）の次発話候補は、３番目の主計画要素である「ゲーム内で使えるポイントが手に入るんだって。」と、現在再生中の２番目の主計画要素に対するユーザの定義型質問に答えるために用意する定義型質問応答の副計画要素である「ＩＣカードっていうのは、…」、「基にっていうのは、…」と、現在再生中の２番目の主計画要素に対するユーザの補足要求に応えるために用意する補足説明用の副計画要素（トリビア）である「ＩＣカードは、国際…」等であり、これらが次発話候補記憶手段３０に記憶される。

【０２７６】

＜次発話候補の準備処理の具体例（２）：図１６＞

【０２７７】

図１６には、ユーザ発話Ｕ（１）が定義型質問となり、システム発話Ｓ（２）として定義型質問応答の副計画要素が再生される場合の具体例が示されている。システム発話Ｓ（１）の再生、Ｓ（２）の複数の候補の準備までは、前述した図１５の場合と同様である。

【０２７８】

図１６において、ユーザ発話Ｕ（１）＝「α社って、どんな会社なの？」（割込でもよい）であったとすると、この場合は、次発話選択用情報生成手段２３により、ユーザ発話意図として、例えば「質問」等の識別結果が得られるが、このユーザ発話意図だけでは、いずれの質問なのか判らないので、次発話選択手段２４は、音声認識処理手段４１による音声認識処理の結果（言語情報）を用いたキーワードマッチング等により、いずれの質問なのか判別し、α社についての質問であることを把握する。従って、次発話選択手段２４により選択される次発話は、Ｓ（２）＝定義型質問応答の副計画要素である「α社は、…」となり、この選択結果がネットワーク１を介して対話状態管理手段４２に送信される。対話状態管理手段４２は、選択結果としてＳ（２）＝「α社は、…」を受け取ると、自身のメモリに保持しているＵ（１）＝「α社って、どんな会社なの？」（音声認識処理手段４１による音声認識処理の結果として受け取り、保持している文字列）と、受け取ったＳ（２）＝「α社は、…」とを対話履歴記憶手段５０に保存するとともに、Ｓ（３）の準備開始指示情報を次発話準備手段４３に送る。

【０２７９】

それから、発話生成手段２５により、次発話選択手段２４により選択されたＳ（２）＝定義型質問応答の副計画要素である「α社は、…」の再生が開始される。また、これと並行して、次発話準備手段４３により、Ｓ（３）の準備処理が進む。この準備処理では、次発話準備手段４３は、対話履歴記憶手段５０を参照し、未だ２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」が再生されていないことを確認することができる。従って、ここで準備されるＳ（３）の次発話候補は、２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」と、再生を終えている１番目の主計画要素に対するユーザの定義型質問に答えるために用意する定義型質問応答の副計画要素である「α社は、…」、「β社は、…」、「β社３ＤＳっていうのは、…」、「交通系ＩＣカードっていうのは、…」、「Ｓｕｉｃａっていうのは、…」、「連携っていうのは、…」、「ゲームソフトっていうのは、…」、「開発っていうのは、…」、「発表っていうのは、…」と、再生を終えている１番目の主計画要素に対するユーザの補足要求に応えるために用意する補足説明用の副計画要素（トリビア）である「Ｓｕｉｃａの名称は…」、「開発は、もともと…」等であり、これらが次発話候補記憶手段３０に記憶される。従って、結果的に、副計画を再生した場合は、次発話候補を維持することになる。

【０２８０】

この際、再生を終えている１番目の主計画要素に対するユーザの質問を想定した準備を行うのは、Ｕ（１）＝「α社って、どんな会社なの？」というユーザの定義型質問に対し、システムが、Ｓ（２）＝定義型質問応答の副計画要素である「α社は、…」を再生した後に、さらにユーザが、「Ｓｕｉｃａって、何？」という定義型質問をする場合があるからである。また、上述したように、結果的に次発話候補を維持するだけでもよいが、Ｓ（２）＝定義型質問応答の副計画要素である「α社は、…」を再生すると、その後、ユーザから「α社は、…」の中の用語について、さらに定義型質問が行われる場合があるので、シナリオデータ内に、「α社は、…」という定義型質問応答の副計画要素の中の用語について、更なる定義型質問応答の副計画要素が用意されていれば、それをＳ（３）の次発話候補に含めて準備してもよい。

【０２８１】

続いて、ユーザ発話Ｕ（２）＝「なるほど。」であったとすると、この場合は、次発話選択用情報生成手段２３により、ユーザ発話意図として、例えば「理解」、「相槌」等の識別結果が得られるので、次発話選択手段２４により選択される次発話は、Ｓ（３）＝２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」となり、この選択結果がネットワーク１を介して対話状態管理手段４２に送信される。対話状態管理手段４２は、選択結果としてＳ（３）を受け取ると、自身のメモリに保持しているＵ（２）＝「なるほど。」（音声認識処理手段４１による音声認識処理の結果として受け取り、保持している文字列）と、受け取ったＳ（３）とを対話履歴記憶手段５０に保存するとともに、Ｓ（４）の準備開始指示情報を次発話準備手段４３に送る。

【０２８２】

それから、発話生成手段２５により、次発話選択手段２４により選択されたＳ（３）＝２番目の主計画要素である「ＩＣカードから読み取った乗車履歴を基に、」の再生が開始される。また、これと並行して、次発話準備手段４３により、Ｓ（４）の準備処理が進む。ここで次発話準備手段４３により準備されるＳ（４）の次発話候補は、３番目の主計画要素である「ゲーム内で使えるポイントが手に入るんだって。」と、現在再生中の２番目の主計画要素に対するユーザの定義型質問に答えるために用意する定義型質問応答の副計画要素である「ＩＣカードっていうのは、…」、「基にっていうのは、…」と、現在再生中の２番目の主計画要素に対するユーザの補足要求に応えるために用意する補足説明用の副計画要素（トリビア）である「ＩＣカードは、国際…」等であり、これらが次発話候補記憶手段３０に記憶される。

【０２８３】

＜次発話候補の準備処理の具体例（３）：図１７＞

【０２８４】

図１７には、次発話候補の入替が行われる具体例が示されている。但し、図１７は、図１３のシナリオデータではなく、同型の別のシナリオデータ（不図示）を用いている。

【０２８５】

図１７において、次発話準備手段４３は、対話状態管理手段４２からのシステム発話Ｓ（１）の準備開始指示情報を受け取り、次発話候補（但し、ここでは最初の発話）の準備処理を行う。対話開始時であるから、複数の次発話候補を選択取得するのではなく、シナリオデータ内から１番目の主計画要素を選択取得する。ここでは、次発話準備手段４３は、シナリオデータから、Ｓ（１）＝「早稲田太郎選手が１００ｍ平泳ぎで優勝したんだ。」を選択取得し、これを次発話候補記憶手段３０に記憶させる。

【０２８６】

続いて、ユーザ発話は未だ無い状態なので、システム発話タイミング検出手段２２により、直ぐにシステム発話の開始タイミングが検出され、次発話選択手段２４により、次発話候補記憶手段３０に記憶されているＳ（１）＝「早稲田太郎選手が１００ｍ平泳ぎで優勝したんだ。」が選択されるとともに、その選択結果がネットワーク１を介して対話状態管理手段４２に送信される。対話状態管理手段４２は、受け取ったＳ（１）を対話履歴記憶手段５０に保存するとともに、Ｓ（２）の準備開始指示情報を次発話準備手段４３に送る。

【０２８７】

それから、発話生成手段２５により、次発話選択手段２４により選択されたＳ（１）＝「早稲田太郎選手が１００ｍ平泳ぎで優勝したんだ。」の再生が開始される。また、これと並行して、次発話準備手段４３により、Ｓ（２）の準備処理が進む。ここで準備されるＳ（２）の次発話候補は、２番目の主計画要素である「オーストラリアで開催された国際大会での快挙なんだ。」と、現在再生中の１番目の主計画要素に対するユーザの定義型質問に答えるために用意する定義型質問応答の副計画要素である「早稲田太郎選手は、…」と、現在再生中の１番目の主計画要素に対するユーザの補足要求に応じるために用意する補足説明用の副計画要素（トリビア）である「去年の優勝者は、…」等であり、これらが次発話候補記憶手段３０に記憶される。

【０２８８】

その後、ユーザ発話Ｕ（１）＝「僕は、平泳ぎよりも、バタフライの方が興味あるんだよ。バタフライの・・・」であったとすると、次発話準備手段４３は、例えば、Ｕ（１）＝「僕は、平泳ぎよりも、バタフライの方が興味あるんだよ。バタフライの」という途中までの情報（音声認識処理手段４１による音声認識処理の結果である文字列）を、対話状態管理手段４２から得ることになる。従って、次発話準備手段４３は、このようなＵ（１）の途中までの音声認識処理の結果に基づき、平泳ぎからバタフライへの話題の転換要求（２回のバタフライという単語の出現、あるいは、「…よりもバタフライの方が興味ある」）を捉え、次発話候補の入替が必要であると判断し、次発話候補の入替のための準備処理を実行する。そして、次発話準備手段４３は、使用中のシナリオデータ内にバタフライのデータ（構成要素であり、主計画要素でも副計画要素でもよい）が存在する場合には、それを選択取得する。また、使用中のシナリオデータ内にバタフライのデータが存在しない場合には、題材データ記憶手段５１に記憶されている別のシナリオデータや、シナリオになっていない別の題材データの中からバタフライのデータを探し、それでも該当データが見つからない場合には、ネットワーク１を介して外部システムである題材データ提供システム６０にアクセスし、該当データを探す。その間は、ステータス＝準備中となる。この際、次発話準備手段４３は、題材データ記憶手段５１および題材データ提供システム６０のいずれを検索する場合でも、分野が関係付けられている題材データについては、先ず分野（例えば、スポーツおよび／または水泳）を用いた絞り込み検索を行うことができ、分野が関係付けられていない題材データについては、バタフライの語によるキーワード検索を行う。

【０２８９】

そして、バタフライの結果が見つかった場合には、次発話準備手段４３により準備され

るＳ（２）の次発話候補は、Ｓ（２）＝「１００ｍバタフライでは、早稲田次郎選手が６位入賞だったんだ。」、「２００ｍバタフライでは、早稲田三郎選手が残念ながら予選落ちしたんだ。」等となり、これらが次発話候補記憶手段３０に記憶される。一方、見つからなかった場合には、例えば、Ｓ（２）＝「ごめんね。バタフライの結果は知らないんだ。」、「バタフライじゃなくて、背泳ぎの結果なら知ってるよ。」等を準備し、次発話候補記憶手段３０に記憶させる。

【０２９０】

最終的に、ユーザ発話Ｕ（１）が終了し、Ｕ（１）＝「僕は、平泳ぎよりも、バタフライの方が興味あるんだよ。バタフライの結果を教えてくれないかな。早稲田次郎選手はどうだったの？」であった場合には、システム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出された後、次発話選択手段２４は、次発話候補記憶手段３０に記憶されているＳ（２）の複数の候補の中からの選択処理を行う。例えば、Ｕ（１）に含まれる単語である「早稲田次郎」によるキーワードマッチングを行う。従って、ここで次発話選択手段２４により選択される次発話は、Ｓ（２）＝「１００ｍバタフライでは、早稲田次郎選手が６位入賞だったんだ。」となる。そして、次発話選択手段２４により、その選択結果がネットワーク１を介して対話状態管理手段４２に送信される。対話状態管理手段４２は、選択結果としてＳ（２）を受け取ると、自身のメモリに保持しているＵ（１）＝「僕は、平泳ぎよりも、バタフライの方が興味あるんだよ。バタフライの結果を教えてくれないかな。早稲田次郎選手はどうだったの？」（音声認識処理手段４１による音声認識処理の結果として受け取り、保持している文字列）と、受け取ったＳ（２）とを対話履歴記憶手段５０に保存するとともに、Ｓ（３）の準備開始指示情報を次発話準備手段４３に送る。

【０２９１】

＜本実施形態の効果＞

【０２９２】

このような本実施形態によれば、次のような効果がある。すなわち、対話システム１０は、システム発話タイミング検出手段２２を備えているので、ユーザが自己の発話権を維持しているか、または、譲渡若しくは放棄により終了させたかをパターン認識処理により逐次推定することができる。また、次発話準備手段４３を備えているので、システム発話タイミング検出手段２２によるパターン認識処理とは非同期で、かつ、システム発話タイミング検出手段２２によりシステム発話の開始タイミングが検出される前に（すなわち、ユーザ発話の進行中に、または、それよりも前の段階であるユーザ発話の開始前に）、ユーザ発話に対するシステムの次発話の内容データを準備することができる。

【０２９３】

このため、対話相手であるユーザが自己の発話権を譲渡若しくは放棄することによりユーザ発話権が終了し、システム発話タイミング検出手段２２により、このユーザ発話権の終了が捉えられ、システム発話の開始タイミングが検出された場合には、その検出直後に、発話生成手段２５により、タイミングよくシステム発話を開始させることができるので、システムの応答性を向上させることができる。

【０２９４】

また、システム発話タイミング検出手段２２は、音声認識処理手段４１による音声認識処理とは非同期で、ユーザ発話権の維持または終了を識別するパターン認識処理を繰り返し実行する構成とされているので、音声区間検出処理（ＶＡＤ処理）を前提としない処理を実現することができるため、ＶＡＤ処理による遅延なしに早期に、システム発話の開始タイミングを決定することができるとともに、ユーザ発話とシステム発話との衝突も回避または抑制することができる。

【０２９５】

以上より、対話システム１０では、次発話準備手段４３により、システムが発話すべき内容（本実施形態では、複数の次発話候補の内容）を早期に確定したうえで、システム発話タイミング検出手段２２により、ユーザ発話権の終了が推定され、システム発話の開始

タイミングが検出されるのを待って、発話生成手段２５により、システム応答を行うことができる。このため、ユーザ発話の終了後、システム発話の開始までに、長い間（ま）が空くことを避けることができるうえ、両者の発話の衝突の発生も回避または抑制することができる。

【０２９６】

また、対話システム１０は、次発話準備手段４３による準備処理で取得または生成した複数の次発話候補の内容データ中から、次発話選択手段２４が、発話生成手段２５で用いる次発話の内容データを選択する構成とされているので、様々な種別の対話に柔軟に対応することができる。すなわち、各種の対話の中には、ユーザ発話の内容が確定する前に、そのユーザ発話に対するシステムの次発話の内容が１つに定まらない種別の対話も多いが、そのような場合でも、システムの応答性の向上を図ることができる。

【０２９７】

そして、次発話選択手段２４は、異なる処理で得られた複数の種類の情報を用いて、次発話の選択処理を行うことができるで、この点でも、様々な種別の対話に柔軟に対応することができる。

【０２９８】

具体的には、次発話選択手段２４は、音声認識処理手段４１による音声認識処理の結果として得られた言語情報（文字列）と、次発話選択用情報生成手段２３により得られた次発話選択用情報（主としてユーザ発話意図の識別結果であるが、ユーザの顔画像から得られた表情の識別結果や、ユーザのジェスチャー画像から得られた身振り・手振りの意図の識別情報を加えてもよい。）とのうちのいずれか一方の情報を用いて、次発話の選択処理を行うことができ、また、これらの情報を組み合わせて用いて、次発話の選択処理を行うこともできる。さらに、システム発話タイミング検出手段２２で得られたユーザ発話意図の識別結果を用いることができる場合もある。従って、様々な種別の対話において、ユーザ発話の内容（必ずしも言語情報に限らず、ユーザ発話意図等も含めた内容）に応じて、システムの次発話の内容データを選択することができる。

【０２９９】

また、上記において、次発話選択手段２４が、韻律分析で推定したユーザ発話意図だけを用いて、複数の次発話候補の内容データの中から、次発話の内容データを選択することができる場合は、音声認識処理の結果を得る必要はないので、システムの応答性を、より一層向上させることができる。

【０３００】

また、システム発話タイミング検出手段２２は、システム状態記憶手段３１に記憶されている準備完了・準備中の別を示すステータスを参照する構成とされているので（図８、図２参照）、システム状態を考慮し、より適切なシステム発話の開始タイミングを検出することができる。

【０３０１】

さらに、システム発話タイミング検出手段２２は、ユーザ状態記憶手段３２に記憶されているユーザ発話継続時間を用いて、ユーザ発話権終了判定用閾値の調整を行うことができるので（図９参照）、ユーザ発話継続時間の長短に応じ、システム発話の開始タイミングを調整することができる。

【０３０２】

また、システム発話タイミング検出手段２２は、システム状態記憶手段３１に記憶されているシステム発話意欲度の指標値（目的データの残数、次発話候補の重要度）を用いてユーザ発話権終了判定用閾値を動的に調整することができるので（図１０参照）、システム発話意欲度が強いときには、ユーザ発話権が終了したという識別結果が出やすくなる設定状態とし、システム発話意欲度が弱いときには、ユーザ発話権が終了したという識別結果が出にくい設定状態とすることができる。

【０３０３】

さらに、システム発話タイミング検出手段２２は、ユーザ情報記憶手段５２に記憶され

ている衝突の発生情報（蓄積情報）やシステムの交替潜時（蓄積情報）を用いて、ユーザ発話権終了判定用閾値を事前調整することができるので（図１１、図１２参照）、各ユーザについて、衝突の発生が起きる傾向にあるときには、ユーザ発話権が終了したという識別結果が出にくい設定状態とし、システムの交替潜時が長い傾向にあるときには、ユーザ発話権が終了したという識別結果が出やすくなる設定状態とすることができる。このため、ユーザ属性に応じたユーザ発話権終了判定用閾値の調整を実現することができる。

【０３０４】

この際、システム発話タイミング検出手段２２は、ユーザ情報記憶手段５２に記憶されているユーザの発話速度（蓄積情報）を用いて、ユーザ発話権終了判定用閾値を下方調整することを決めるための下方調整用閾値を、ユーザの発話速度の関数として設定することができるので（図１２参照）、各ユーザの発話速度の傾向に応じ、下方調整用閾値の設定を変更することができる。このため、ユーザ属性に応じたユーザ発話権終了判定用閾値の下方調整を実現することができる。すなわち、システムの交替潜時が長い傾向にあるときには、ユーザ発話権終了判定用閾値を下方調整することにより、ユーザ発話権が終了したという識別結果が出やすくなる設定状態とし、システムの交替潜時が短くなるようにすることができるが、この際、システムの交替潜時が長い傾向にあるか否かは、ユーザ毎に異なり、各ユーザの発話速度の傾向と関係するので、下方調整用閾値をユーザの発話速度の関数とすることで、ユーザ属性に応じてユーザ発話権終了判定用閾値の下方調整を行うか否かを決めることができる。

【０３０５】

また、次発話準備手段４３は、入替要否判断手段４３Ｂおよび入替準備手段４３Ｃを備えているので（図４参照）、進行中のユーザ発話の内容を逐次反映させ、既に準備されている複数の次発話候補の内容データの入替を行うことができる。このため、ユーザ発話の内容に応じた適切な次発話候補の内容データを準備することができる。

【０３０６】

例えば、次発話準備手段４３は、逐次得られる音声認識処理の結果に含まれる重要度の高い単語を用いて、ユーザの関心のある話題を決定し、題材データ記憶手段５１または題材データ提供システム６０に記憶された題材データの中から、決定した話題に関連付けられて記憶されている題材データを選択し、次発話の候補となる別の複数の次発話候補の内容データを取得または生成する準備処理を実行することができる。従って、次発話により提示する話題を変更することができる。

【０３０７】

＜変形の形態＞

【０３０８】

なお、本発明は前記実施形態に限定されるものではなく、本発明の目的を達成できる範囲内での変形等は本発明に含まれるものである。

【０３０９】

例えば、前記実施形態の対話システム１０は、次発話準備手段４３により、複数の次発話候補の内容データを準備して次発話候補記憶手段３０に記憶させ、次発話選択手段２４により、複数の次発話候補の内容データの中から、次発話の内容データを選択する構成とされていたが、本発明の対話システムは、このような構成に限定されるものではなく、次発話準備手段４３により、次発話の内容データを１つだけ準備し、次発話選択手段２４を設置しない構成としてもよい。但し、様々な種別の対話に対応できるようにするという観点で、前記実施形態のように、次発話選択手段２４を設け、次発話準備手段４３により複数の次発話候補の内容データを準備する構成とすることが好ましい。

【０３１０】

また、前記実施形態では、システム発話タイミング検出手段２２は、ユーザ情報記憶手段５２に記憶されているユーザの発話速度（蓄積情報）を用いて、ユーザ発話権終了判定用閾値を下方調整することを決めるための下方調整用閾値を、ユーザの発話速度の関数として設定する構成とれていたが（図１２参照）、蓄積されたユーザの発話速度から得られ

るユーザ属性（発話速度の傾向）を用いるのではなく、ユーザ状態記憶手段３２に記憶されているユーザのリアルタイムの発話速度（その時々の発話速度）を用いて、ユーザ発話権の維持または終了を識別するパターン認識処理を行う構成としてもよい。

【０３１１】

このような構成とする場合、ユーザ発話権の維持または終了を識別する識別器を構築する際には、ユーザ発話の音声信号から抽出した音響特徴量と、音声認識処理手段４１による音声認識処理の結果として得られた言語情報から抽出した言語特徴量（但し、省略してもよい）と、ユーザ発話における対応する各時点でのユーザの発話速度とを入力して識別器の学習を行う。そして、運用時には、音響特徴量と、言語特徴量（但し、省略してもよい）と、逐次得られるリアルタイムの発話速度とを、識別器に入力する。これにより、ユーザのリアルタイムの発話速度を加味した識別結果を得ることができる。このため、ユーザ毎に異なる発話速度の傾向（蓄積された発話速度から得られるユーザ属性）に応じてユーザ発話権終了判定用閾値を調整する必要がなくなる。なお、閾値調整と併用してもよく、その場合には、閾値調整量が少なくなる。

【産業上の利用可能性】

【０３１２】

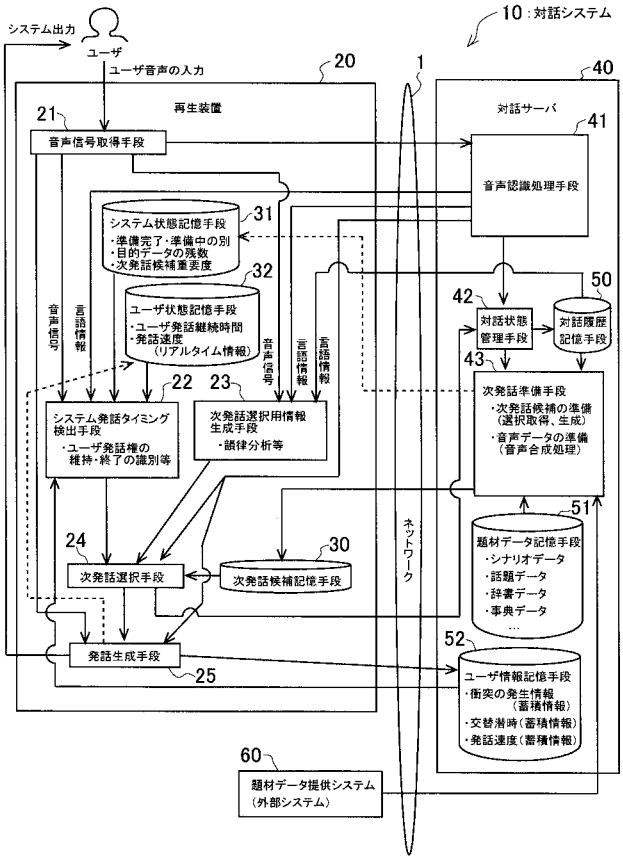
以上のように、本発明の対話システムおよびプログラムは、例えば、ニュースやコラムや歴史等の各種の話題を記載した記事データから生成したシナリオデータを用いてユーザに対して記事の内容を伝達するニュース対話システム、ユーザに対して機器の使用方法の説明や施設の案内等を行うガイダンス対話システム、選挙情勢や消費者志向等の各種のユーザの動向調査を行うアンケート対話システム、ユーザが店舗・商品・旅行先・聞きたい曲等の情報検索を行うための情報検索対話システム、ユーザが家電機器や車等の各種の機器や装置等を操作するための操作対話システム、子供や学生や新入社員等であるユーザに対して教育を行うための教育対話システム、システムがユーザ属性等の情報を特定するための情報特定対話システム等に用いるのに適している。

【符号の説明】

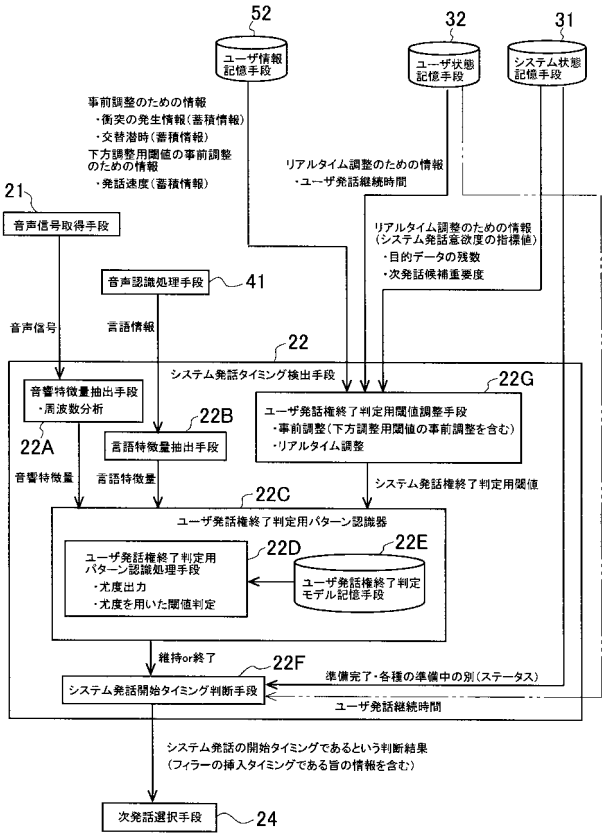
【０３１３】

１　ネットワーク
１０　対話システム
２１　音声信号取得手段
２２　システム発話タイミング検出手段
２３　次発話選択用情報生成手段
２４　次発話選択手段
２５　発話生成手段
３０　次発話候補記憶手段
３１　システム状態記憶手段
３２　ユーザ状態記憶手段
４１　音声認識処理手段
４２　対話状態管理手段
４３　次発話準備手段
５０　対話履歴記憶手段
５１　題材データ記憶手段
５２　ユーザ情報記憶手段
６０　外部システムである題材データ提供システム

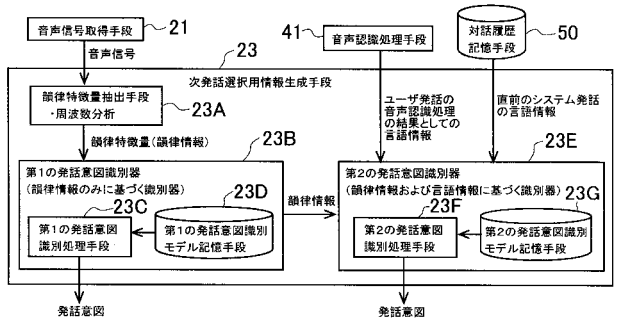
【図１】


システム出力
ユーザ
ユーザ音声の入力
10：対話システム
20
1
40
再生装置
対話サーバ
21
音声信号取得手段
41
音声認識処理手段
システム状態記憶手段
・準備完了・準備中の別
・目的データの残数
・次発話候補重要度
31
32
ユーザ状態記憶手段
・ユーザ発話継続時間
・発話速度
（リアルタイム情報）
50
42
対話状態管理手段
対話履歴記憶手段
43
次発話準備手段
・次発話候補の準備（選択取得、生成）
・音声データの準備（音声合成処理）
22
システム発話タイミング検出手段
・ユーザ発話権の維持・終了の識別等
23
次発話選択用情報生成手段
・韻律分析等
音声信号
言語情報
51
題材データ記憶手段
・シナリオデータ
・話題データ
・辞書データ
・事典データ
…
24
次発話選択手段
30
次発話候補記憶手段
ネットワーク
52
ユーザ情報記憶手段
・衝突の発生情報（蓄積情報）
・交替潜時（蓄積情報）
・発話速度（蓄積情報）
発話生成手段
25
60
題材データ提供システム（外部システム）

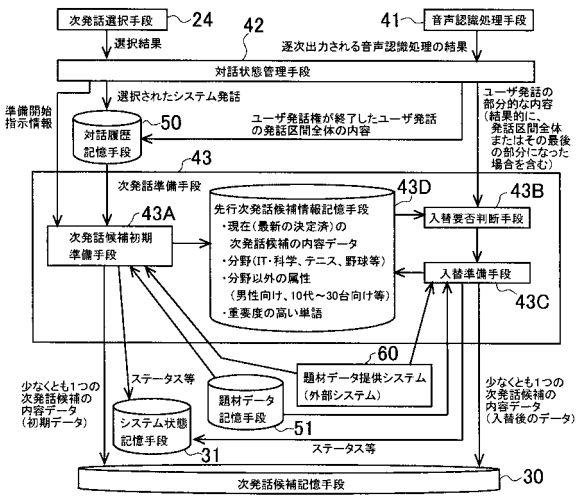
【図２】


52
ユーザ情報記憶手段
32
ユーザ状態記憶手段
31
システム状態記憶手段
事前調整のための情報
・衝突の発生情報（蓄積情報）
・交替潜時（蓄積情報）
下方調整用閾値の事前調整のための情報
・発話速度（蓄積情報）
リアルタイム調整のための情報
・ユーザ発話継続時間
リアルタイム調整のための情報（システム発話意欲度の指標値）
・目的データの残数
・次発話候補重要度
21
音声信号取得手段
音声認識処理手段
41
音声信号
言語情報
22
システム発話タイミング検出手段
音響特徴量抽出手段
・周波数分析
22A
言語特徴量抽出手段
22B
ユーザ発話権終了判定用閾値調整手段
・事前調整（下方調整用閾値の事前調整を含む）
・リアルタイム調整
22G
音響特徴量
言語特徴量
システム発話権終了判定用閾値
ユーザ発話権終了判定用パターン認識器
22C
ユーザ発話権終了判定用パターン認識処理手段
・尤度出力
・尤度を用いた閾値判定
22D
ユーザ発話権終了判定モデル記憶手段
22E
維持or終了
22F
システム発話開始タイミング判断手段
準備完了・各種の準備中の別（ステータス）
ユーザ発話継続時間
システム発話の開始タイミングであるという判断結果
（フィラーの挿入タイミングである旨の情報を含む）
次発話選択手段
24

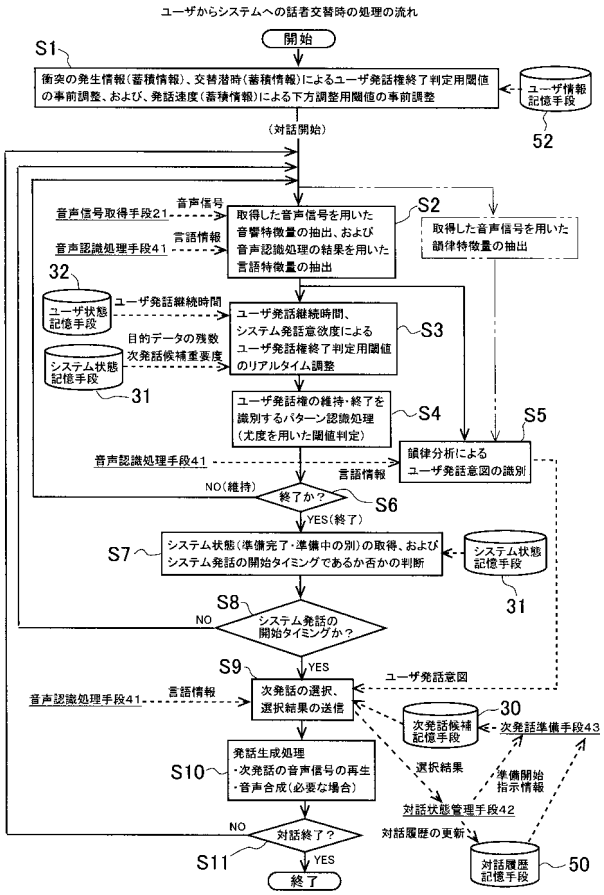
【図３】


音声信号取得手段
21
音声信号
41
音声認識処理手段
対話履歴記憶手段
50
23
次発話選択用情報生成手段
韻律特徴量抽出手段
・周波数分析
23A
韻律特徴量（韻律情報）
ユーザ発話の音声認識処理の結果としての言語情報
直前のシステム発話の言語情報
23B
23E
第1の発話意図識別器（韻律情報のみに基づく識別器）
第2の発話意図識別器（韻律情報および言語情報に基づく識別器）
23G
23D
23C
第1の発話意図識別処理手段
第1の発話意図識別モデル記憶手段
韻律情報
23F
第2の発話意図識別処理手段
第2の発話意図識別モデル記憶手段
発話意図
発話意図

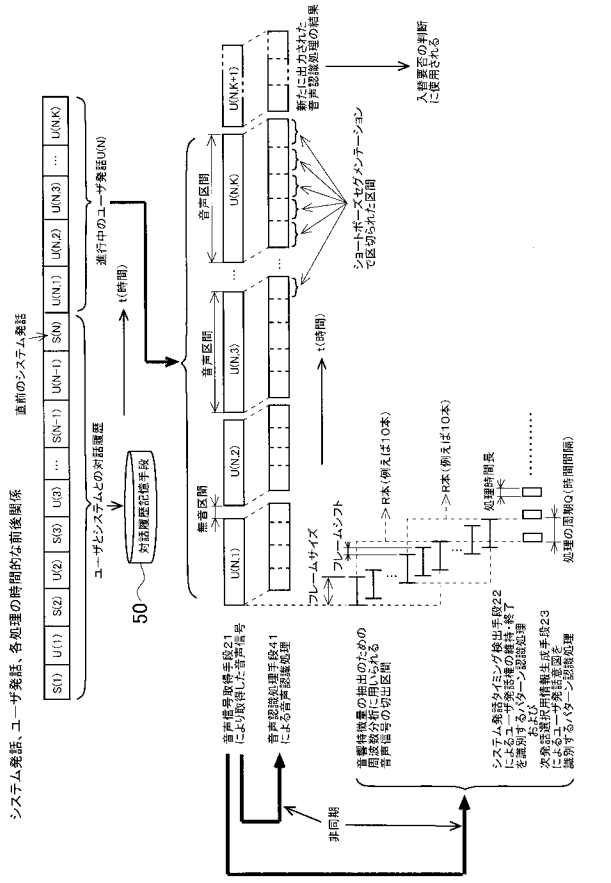
【図４】


次発話選択手段
24
選択結果
42
41
音声認識処理手段
逐次出力される音声認識処理の結果
対話状態管理手段
準備開始指示情報
選択されたシステム発話
50
対話履歴記憶手段
ユーザ発話権が終了したユーザ発話の発話区間全体の内容
ユーザ発話の部分的な内容（結果的に、発話区間全体またはその最後の部分になった場合を含む）
43
次発話準備手段
43D
43B
先行次発話候補情報記憶手段
・現在（最新の決定済）の次発話候補の内容データ
・分野（IT・科学、テニス、野球等）
・分野以外の属性（男性向け、10代～30台向け等）
・重要度の高い単語
入替要否判断手段
43A
次発話候補初期準備手段
入替準備手段
43C
60
題材データ提供システム（外部システム）
少なくとも1つの次発話候補の内容データ（初期データ）
ステータス等
題材データ記憶手段
51
システム状態記憶手段
31
ステータス等
少なくとも1つの次発話候補の内容データ（入替後のデータ）
次発話候補記憶手段
30

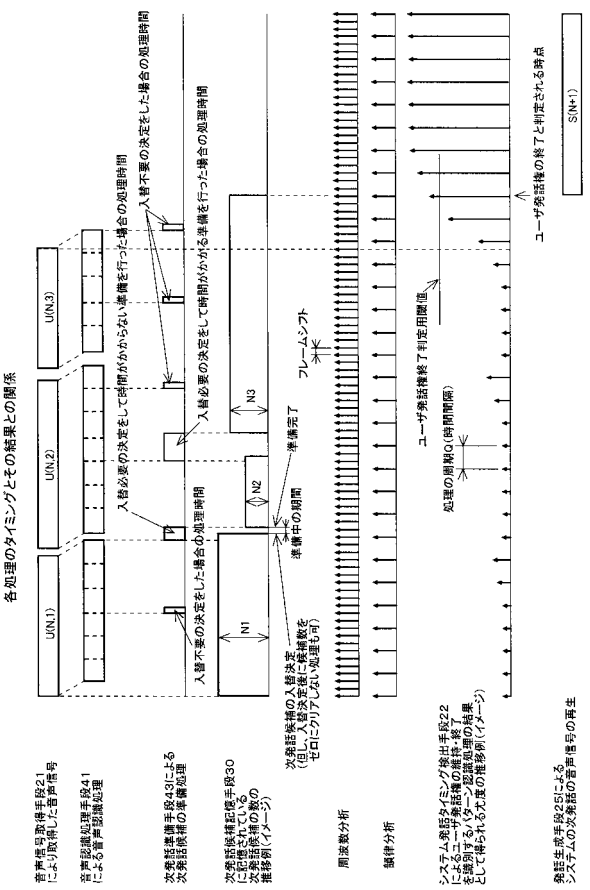
【図５】


ユーザからシステムへの話者交替時の処理の流れ
開始
S1
衝突の発生情報（蓄積情報）、交替潜時（蓄積情報）によるユーザ発話権終了判定用閾値の事前調整、および、発話速度（蓄積情報）による下方調整用閾値の事前調整
ユーザ情報記憶手段
52
（対話開始）
音声信号取得手段21
音声信号
音声認識処理手段41
言語情報
取得した音声信号を用いた音響特徴量の抽出、および音声認識処理の結果を用いた言語特徴量の抽出
S2
取得した音声信号を用いた韻律特徴量の抽出
32
ユーザ状態記憶手段
ユーザ発話継続時間
目的データの残数
次発話候補重要度
システム状態記憶手段
31
ユーザ発話継続時間、システム発話意欲度によるユーザ発話権終了判定用閾値のリアルタイム調整
S3
ユーザ発話権の維持・終了を識別するパターン認識処理（尤度を用いた閾値判定）
S4
S5
韻律分析によるユーザ発話意図の識別
音声認識処理手段41
言語情報
NO（維持）
終了か？
S6
YES（終了）
S7
システム状態（準備完了・準備中の別）の取得、およびシステム発話の開始タイミングであるか否かの判断
システム状態記憶手段
31
S8
NO
システム発話の開始タイミングか？
YES
S9
ユーザ発話意図
音声認識処理手段41
言語情報
次発話の選択、選択結果の送信
30
次発話候補記憶手段
次発話準備手段43
発話生成処理
・次発話の音声信号の再生
・音声合成（必要な場合）
S10
選択結果
準備開始指示情報
対話状態管理手段42
NO
対話終了？
対話履歴の更新
S11
YES
終了
対話履歴記憶手段
50

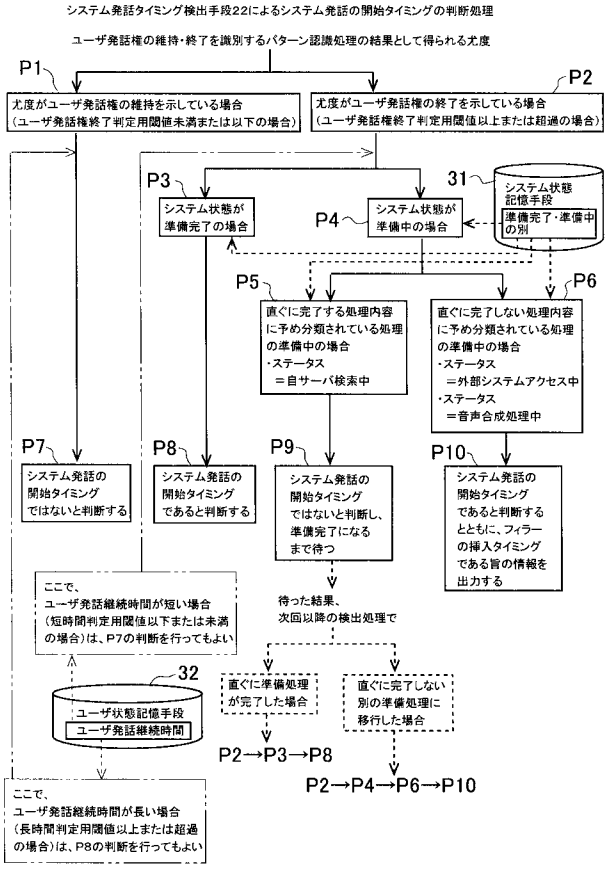
【図６】



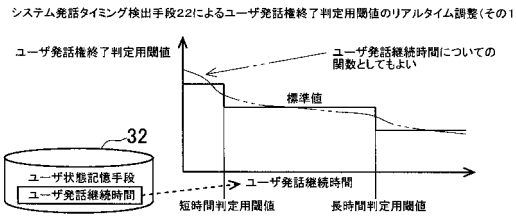

【図７】



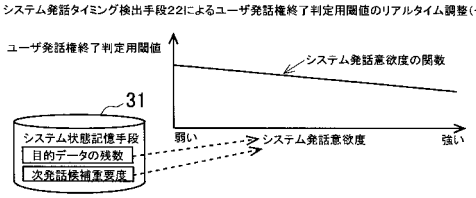

【図８】



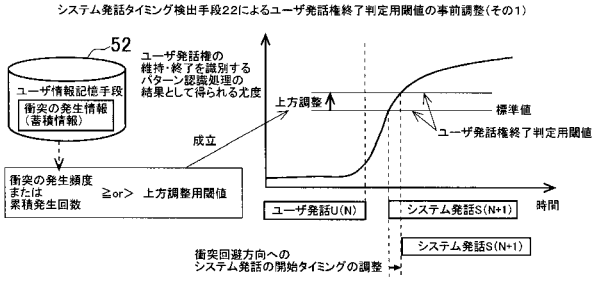

【図９】



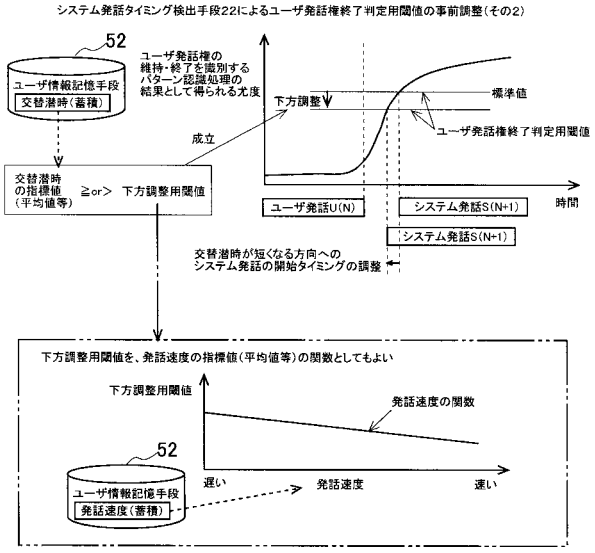

【図１０】



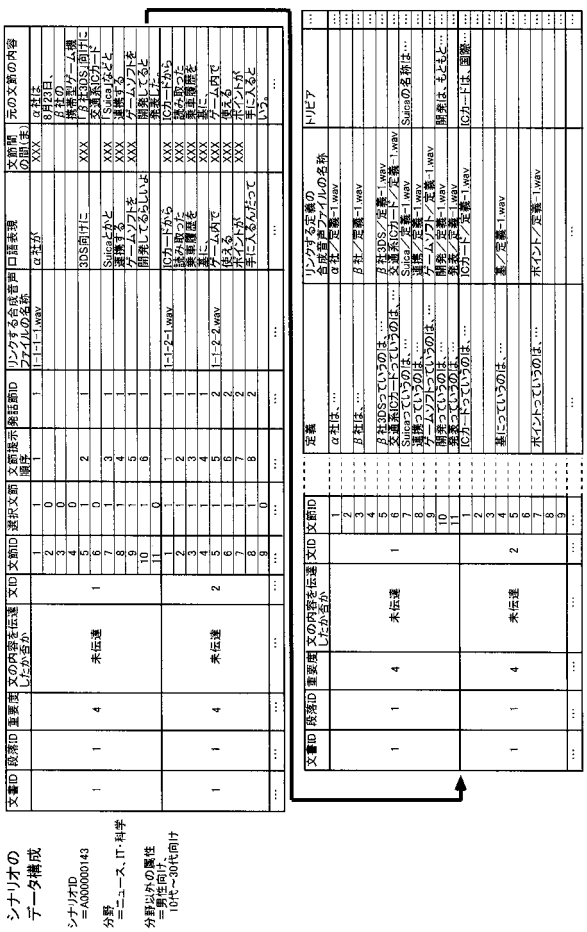

【図１１】



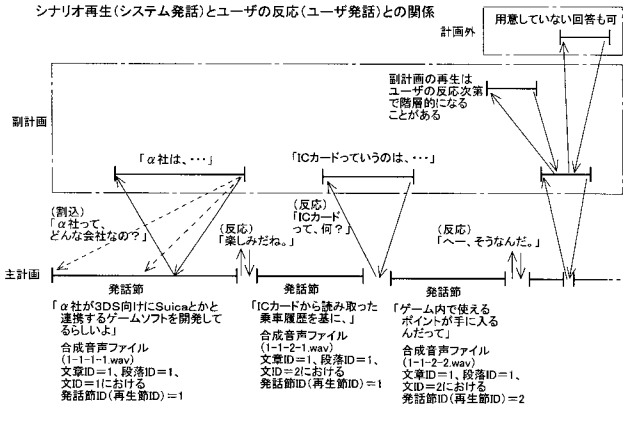
【図１２】




【図１３】




【図１４】

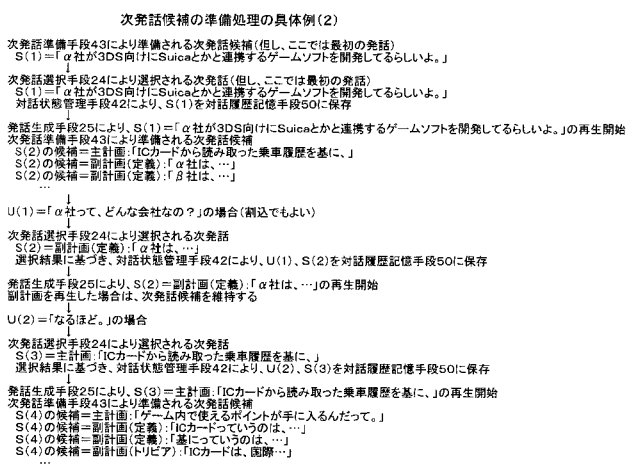



【図１５】

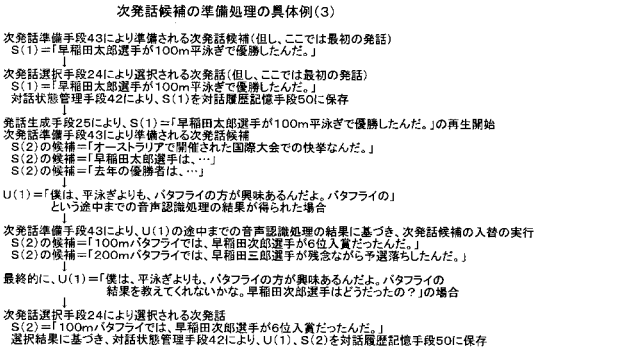
次発話候補の準備処理の具体例（1）

次発話準備手段43により準備される次発話候補（但し、ここでは最初の発話）
S(1)＝「α社が3DS向けにSuicaとかと連携するゲームソフトを開発してるらしいよ。」
↓
次発話選択手段24により選択される次発話（但し、ここでは最初の発話）
S(1)＝「α社が3DS向けにSuicaとかと連携するゲームソフトを開発してるらしいよ。」
対話状態管理手段42により、S(1)を対話履歴記憶手段50に保存
↓
発話生成手段25により、S(1)＝「α社が3DS向けにSuicaとかと連携するゲームソフトを開発してるらしいよ。」の再生開始
次発話準備手段43により準備される次発話候補
S(2)の候補＝主計画：「ICカードから読み取った乗車履歴を基に、」
S(2)の候補＝副計画（定義）：「α社は、…」
S(2)の候補＝副計画（定義）：「β社は、…」
S(2)の候補＝副計画（定義）：「β社3DSっていうのは、…」
S(2)の候補＝副計画（定義）：「交通系ICカードっていうのは、…」
S(2)の候補＝副計画（定義）：「Suicaっていうのは、…」
S(2)の候補＝副計画（定義）：「連携っていうのは、…」
S(2)の候補＝副計画（定義）：「ゲームソフトっていうのは、…」
S(2)の候補＝副計画（定義）：「開発っていうのは、…」
S(2)の候補＝副計画（定義）：「発表っていうのは、…」
S(2)の候補＝副計画（トリビア）：「Suicaの名称は…」
S(2)の候補＝副計画（トリビア）：「開発は、もともと…」
…
↓
U(1)＝「楽しみだね。」の場合
↓
次発話選択手段24により選択される次発話
S(2)＝主計画：「ICカードから読み取った乗車履歴を基に、」
選択結果に基づき、対話状態管理手段42により、U(1)、S(2)を対話履歴記憶手段50に保存
↓
発話生成手段25により、S(2)＝主計画：「ICカードから読み取った乗車履歴を基に、」の再生開始
次発話準備手段43により準備される次発話候補
S(3)の候補＝主計画：「ゲーム内で使えるポイントが手に入るんだって。」
S(3)の候補＝副計画（定義）：「ICカードっていうのは、…」
S(3)の候補＝副計画（定義）：「基にっていうのは、…」
S(3)の候補＝副計画（トリビア）：「ICカードは、国際…」
…

【図１６】


次発話候補の準備処理の具体例(2)
次発話準備手段43により準備される次発話候補(但し、ここでは最初の発話)
S(1)＝「α社が3DS向けにSuicaとかと連携するゲームソフトを開発してるらしいよ。」
次発話選択手段24により選択される次発話(但し、ここでは最初の発話)
S(1)＝「α社が3DS向けにSuicaとかと連携するゲームソフトを開発してるらしいよ。」
対話状態管理手段42により、S(1)を対話履歴記憶手段50に保存
発話生成手段25により、S(1)＝「α社が3DS向けにSuicaとかと連携するゲームソフトを開発してるらしいよ。」の再生開始
次発話準備手段43により準備される次発話候補
S(2)の候補＝主計画:「ICカードから読み取った乗車履歴を基に、」
S(2)の候補＝副計画(定義):「α社は、…」
S(2)の候補＝副計画(定義):「β社は、…」
…
U(1)＝「α社って、どんな会社なの？」の場合(割込でもよい)
次発話選択手段24により選択される次発話
S(2)＝副計画(定義):「α社は、…」
選択結果に基づき、対話状態管理手段42により、U(1)、S(2)を対話履歴記憶手段50に保存
発話生成手段25により、S(2)＝副計画(定義):「α社は、…」の再生開始
副計画を再生した場合は、次発話候補を維持する
U(2)＝「なるほど。」の場合
次発話選択手段24により選択される次発話
S(3)＝主計画:「ICカードから読み取った乗車履歴を基に、」
選択結果に基づき、対話状態管理手段42により、U(2)、S(3)を対話履歴記憶手段50に保存
発話生成手段25により、S(3)＝主計画:「ICカードから読み取った乗車履歴を基に、」の再生開始
次発話準備手段43により準備される次発話候補
S(4)の候補＝主計画:「ゲーム内で使えるポイントが手に入るんだって。」
S(4)の候補＝副計画(定義):「ICカードっていうのは、…」
S(4)の候補＝副計画(定義):「基にっていうのは、…」
S(4)の候補＝副計画(トリビア):「ICカードは、国際…」
…

【図１７】


次発話候補の準備処理の具体例(3)
次発話準備手段43により準備される次発話候補(但し、ここでは最初の発話)
S(1)＝「早稲田太郎選手が100m平泳ぎで優勝したんだ。」
次発話選択手段24により選択される次発話(但し、ここでは最初の発話)
S(1)＝「早稲田太郎選手が100m平泳ぎで優勝したんだ。」
対話状態管理手段42により、S(1)を対話履歴記憶手段50に保存
発話生成手段25により、S(1)＝「早稲田太郎選手が100m平泳ぎで優勝したんだ。」の再生開始
次発話準備手段43により準備される次発話候補
S(2)の候補＝「オーストラリアで開催された国際大会での快挙なんだ。」
S(2)の候補＝「早稲田太郎選手は、…」
S(2)の候補＝「去年の優勝者は、…」
U(1)＝「僕は、平泳ぎよりも、バタフライの方が興味あるんだよ。バタフライの」
という途中までの音声認識処理の結果が得られた場合
次発話準備手段43により、U(1)の途中までの音声認識処理の結果に基づき、次発話候補の入替の実行
S(2)の候補＝「100mバタフライでは、早稲田次郎選手が6位入賞だったんだ。」
S(2)の候補＝「200mバタフライでは、早稲田三郎選手が残念ながら予選落ちしたんだ。」
最終的に、U(1)＝「僕は、平泳ぎよりも、バタフライの方が興味あるんだよ。バタフライの
結果を教えてくれないかな。早稲田次郎選手はどうだったの？」の場合
次発話選択手段24により選択される次発話
S(2)＝「100mバタフライでは、早稲田次郎選手が6位入賞だったんだ。」
選択結果に基づき、対話状態管理手段42により、U(1)、S(2)を対話履歴記憶手段50に保存

フロントページの続き

| (51)Int.Cl. | FI | | テーマコード（参考） |
|---|---|---|---|
| | G06F | 3/16 | 690 |

特許法第３０条第２項適用申請有り　（１）令和１年７月１２日掲載　（目次）　ｈｔｔｐｓ：／／ｉｐｓｊ．ｉｘｓｑ．ｎｉｉ．ａｃ．ｊｐ／ｅｊ／ｉｎｄｅｘ．ｐｈｐ？ａｃｔｉｏｎ＝ｐａｇｅｓ＿ｖｉｅｗ＿ｍａｉｎ＆ａｃｔｉｖｅ＿ａｃｔｉｏｎ＝ｒｅｐｏｓｉｔｏｒｙ＿ｖｉｅｗ＿ｍａｉｎ＿ｉｔｅｍ＿ｓｎｉｐｐｅｔ＆ｉｎｄｅｘ＿ｉｄ＝９８５９＆ｐｎ＝１＆ｃｏｕｎｔ＝２０＆ｏｒｄｅｒ＝７＆ｌａｎｇ＝ｊａｐａｎｅｓｅ＆ｐａｇｅ＿ｉｄ＝１３＆ｂｌｏｃｋ＿ｉｄ＝８　（論文へのアクセス用記事）　ｈｔｔｐ：／／ｉｄ．ｎｉｉ．ａｃ．ｊｐ／１００１／００１９７９５４／　を通じて論文抄録及び論文を発表（２）令和１年７月１９日発表　一般社団法人情報処理学会の第１２８回音声言語情報処理研究会（ＳＩＧ－ＳＬＰ）の研究発表会、新潟県月岡温泉、風鈴屋にて、スライドを用いて発表するとともに、発表当日に来場者にダウンロード形式で論文を電子的に配布