

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号

特許第7036015号

(P7036015)

(45)発行日 令和4年3月15日(2022.3.15)

(24)登録日 令和4年3月7日(2022.3.7)

(51)国際特許分類

F I

G 1 0 L	13/08	(2013.01)	G 1 0 L	13/08	1 2 4
G 1 0 L	13/00	(2006.01)	G 1 0 L	13/00	1 0 0 M
G 1 0 L	13/10	(2013.01)	G 1 0 L	13/08	1 2 2
G 1 0 L	15/22	(2006.01)	G 1 0 L	13/10	1 1 1 A
			G 1 0 L	13/10	1 1 1 C

請求項の数 16 (全35頁) 最終頁に続く

(21)出願番号 特願2018-534342(P2018-534342)  
 (86)(22)出願日 平成29年8月3日(2017.8.3)  
 (86)国際出願番号 PCT/JP2017/028292  
 (87)国際公開番号 WO2018/034169  
 (87)国際公開日 平成30年2月22日(2018.2.22)  
 審査請求日 令和2年6月25日(2020.6.25)  
 (31)優先権主張番号 特願2016-159864(P2016-159864)  
 (32)優先日 平成28年8月17日(2016.8.17)  
 (33)優先権主張国・地域又は機関  
 日本国(JP)

(73)特許権者 000002185  
 ソニーグループ株式会社  
 東京都港区港南1丁目7番1号  
 (74)代理人 100121131  
 弁理士 西川 孝  
 (74)代理人 100082131  
 弁理士 稲本 義雄  
 (72)発明者 岩瀬 広  
 東京都港区港南1丁目7番1号 ソニー  
 株式会社内  
 (72)発明者 斎藤 真里  
 東京都港区港南1丁目7番1号 ソニー  
 株式会社内  
 (72)発明者 河野 真一  
 東京都港区港南1丁目7番1号 ソニー  
 株式会社内  
 最終頁に続く

(54)【発明の名称】 対話制御装置および方法

## (57)【特許請求の範囲】

## 【請求項1】

1または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話を行わせ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話を制御し、前記発話テキストに基づく発話を途中で停止させた場合、前記発話テキストを含む通知情報の出力を制御する対話進行制御部を備える  
 対話制御装置。

## 【請求項2】

前記対話進行制御部は、前記ユーザによる前記理解行動が肯定的な行動であった場合、前記発話テキストのまだ発話がされていない次の文言を発話させる  
 請求項1に記載の対話制御装置。

## 【請求項3】

前記対話進行制御部は、前記ユーザによる前記理解行動が否定的な行動であった場合、直前に行われた発話を再度行わせる  
 請求項1または請求項2に記載の対話制御装置。

## 【請求項4】

前記対話進行制御部は、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど前記理解行動の誘発回数が少なくなるように、前記発話テキストに基づく発話を制御する

請求項 1 乃至請求項 3 の何れか一項に記載の対話制御装置。

【請求項 5】

前記対話進行制御部は、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど発話速度が速くなるように、前記発話テキストに基づく発話を制御する

請求項 1 乃至請求項 4 の何れか一項に記載の対話制御装置。

【請求項 6】

前記対話進行制御部は、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど発話のトーンが高くなるように、前記発話テキストに基づく発話を制御する

請求項 1 乃至請求項 5 の何れか一項に記載の対話制御装置。

【請求項 7】

前記対話進行制御部は、所定時間内に前記ユーザによる前記理解行動が検出されなかった場合、直前に行われた発話を再度行わせるか、または前記理解行動を要求する発話を行わせる

請求項 1 乃至請求項 6 の何れか一項に記載の対話制御装置。

【請求項 8】

前記対話進行制御部は、前記発話テキストに基づく文言に非完了文言を付加することで、前記理解行動を誘発する発話を行わせる

請求項 1 乃至請求項 7 の何れか一項に記載の対話制御装置。

【請求項 9】

前記対話進行制御部は、発話の語尾に抑揚を付加することで、前記理解行動を誘発する発話を行わせる

請求項 8 に記載の対話制御装置。

【請求項 10】

前記対話進行制御部は、前記理解行動要求位置において間をあけることで前記理解行動を誘発する発話を行わせる

請求項 1 乃至請求項 9 の何れか一項に記載の対話制御装置。

【請求項 11】

前記発話テキストの文章における文として非完結な位置を前記理解行動要求位置として検出する理解行動要求位置検出部をさらに備える

請求項 1 乃至請求項 10 の何れか一項に記載の対話制御装置。

【請求項 12】

前記理解行動要求位置検出部は、前記発話テキストの文章の述語節との係り受けに基づく位置を前記理解行動要求位置として検出する

請求項 11 に記載の対話制御装置。

【請求項 13】

前記理解行動要求位置検出部は、前記発話テキストの文章における同一の述語節に係る複数の対象格の文節または句のそれぞれの間の位置を前記理解行動要求位置として検出する

請求項 12 に記載の対話制御装置。

【請求項 14】

前記理解行動要求位置検出部は、前記発話テキストの文章における述語節に最初に係る文節または句の位置を前記理解行動要求位置として検出する

請求項 12 に記載の対話制御装置。

【請求項 15】

前記理解行動要求位置検出部は、前記発話テキストの文章における時間格、場所格、対象格、または原因格である文節または句の位置を前記理解行動要求位置として検出する

請求項 11 に記載の対話制御装置。

【請求項 16】

1 または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話を行わせ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話を制御し、前記発話テキストに基づく発話を途中

10

20

30

40

50

で停止させた場合、前記発話テキストを含む通知情報の出力を制御する

ステップを含む対話制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本技術は対話制御装置および方法に関し、特に、より適切な対話制御を行うことができるようにした対話制御装置および方法に関する。

【背景技術】

【0002】

従来、ユーザとの音声対話を制御する技術が知られている。

【0003】

例えばそのような技術として、ユーザに対して音声発話を要求するとともに、ユーザにより発話されたキーワードに基づいてユーザの理解度を判定し、その判定結果に応じて対話を進めていくものがある（例えば、特許文献1参照）。

【0004】

また、例えばユーザとの対話をするにあたり、ユーザの相槌行為を検出したり、ユーザに相槌を要求したりする技術も提案されている（例えば、特許文献2参照）。

【先行技術文献】

【特許文献】

【0005】

【文献】特開2005-234331号公報

特開平9-218770号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

ところが、上述した技術ではユーザとの音声対話を適切に制御することは困難であった。

【0007】

例えばキーワードに基づいてユーザの理解度を判定する技術では、ユーザの理解度に応じて音声入力を促すメッセージが出力される。しかし、ユーザへの伝達項目のリストが長かったり、伝達内容が長文であったりする場合など、システムが音声により提示する、ユーザに対して伝えたい情報の量が多い場合には、最適な粒度で、つまり最適なタイミングでユーザの理解度を測ることができない。

【0008】

TTS (Text To Speech) での合成音声は速度や抑揚が一定であり、情報量が多く音声発話する文章が長くなると、ユーザが発話内容を理解し、集中して聞いているにも関わらず遅い速度で発話が行われたり、頻繁に音声入力を促されたりするなど、ユーザの状況に合わない対話制御が行われてしまうことがある。このような場合、ユーザが発話内容を理解しにくいだけでなく、発話内容も記憶しにくくなってしまう。

【0009】

また、上述した技術では、音声対話中に、すなわち発話音声の文の途中で相槌や頷きなど、ユーザに対して負荷の低い理解行動自体を誘発することができない。

【0010】

そのため、例えば文よりもさらに細かい文節や意味の区切りで理解行動を要求する頻度をユーザの集中度や理解度に応じて制御するなど、最適な粒度でユーザに負荷の低い理解行動を要求することができなかつた。

【0011】

本技術は、このような状況に鑑みてなされたものであり、より適切な対話制御を行うことができるようにするものである。

【課題を解決するための手段】

【0012】

10

20

30

40

50

本技術の一側面の対話制御装置は、1または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話を行わせ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話を制御し、前記発話テキストに基づく発話を途中で停止させた場合、前記発話テキストを含む通知情報の出力を制御する対話進行制御部を備える。

【0013】

前記対話進行制御部には、前記ユーザによる前記理解行動が肯定的な行動であった場合、前記発話テキストのまだ発話されていない次の文言を発話させるようにすることができる。

【0014】

前記対話進行制御部には、前記ユーザによる前記理解行動が否定的な行動であった場合、直前に行われた発話を再度行わせるようにすることができる。

【0015】

前記対話進行制御部には、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど前記理解行動の誘発回数が少なくなるように、前記発話テキストに基づく発話を制御させることができる。

【0016】

前記対話進行制御部には、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど発話速度が速くなるように、前記発話テキストに基づく発話を制御させることができる。

【0017】

前記対話進行制御部には、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど発話のトーンが高くなるように、前記発話テキストに基づく発話を制御させることができる。

【0018】

前記対話進行制御部には、所定時間内に前記ユーザによる前記理解行動が検出されなかった場合、直前に行われた発話を再度行わせるか、または前記理解行動を要求する発話を行わせるようにすることができる。

【0020】

前記対話進行制御部には、前記発話テキストに基づく文言に非完了文言を付加することで、前記理解行動を誘発する発話を行わせるようにすることができる。

【0021】

前記対話進行制御部には、発話の語尾に抑揚を付加することで、前記理解行動を誘発する発話を行わせるようにすることができる。

【0022】

前記対話進行制御部には、前記理解行動要求位置において間をあけることで前記理解行動を誘発する発話を行わせるようにすることができる。

【0023】

対話制御装置には、前記発話テキストの文章における文として非完結な位置を前記理解行動要求位置として検出する理解行動要求位置検出部をさらに設けることができる。

【0024】

前記理解行動要求位置検出部には、前記発話テキストの文章の述語節との係り受けに基づく位置を前記理解行動要求位置として検出させることができる。

【0025】

前記理解行動要求位置検出部には、前記発話テキストの文章における同一の述語節に係る複数の対象格の文節または句のそれぞれの間の位置を前記理解行動要求位置として検出させることができる。

【0026】

前記理解行動要求位置検出部には、前記発話テキストの文章における述語節に最初に係る文節または句の位置を前記理解行動要求位置として検出させることができる。

10

20

30

40

50

## 【0027】

前記理解行動要求位置検出部には、前記発話テキストの文章における時間格、場所格、対象格、または原因格である文節または句の位置を前記理解行動要求位置として検出させることができる。

## 【0028】

本技術の一側面の対話制御方法は、1または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話を行わせ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話を制御し、前記発話テキストに基づく発話を途中で停止させた場合、前記発話テキストを含む通知情報の出力を制御するステップを含む。

10

## 【0029】

本技術の一側面においては、1または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話が行われ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話が制御される。また、前記発話テキストに基づく発話を途中で停止させた場合、前記発話テキストを含む通知情報の出力が制御される。

## 【発明の効果】

## 【0030】

本技術の一側面によれば、より適切な対話制御を行うことができる。

## 【0031】

なお、ここに記載された効果は必ずしも限定されるものではなく、本開示中に記載された何れかの効果であってもよい。

20

## 【図面の簡単な説明】

## 【0032】

【図1】音声対話システムの構成例を示す図である。

【図2】理解行動要求位置の検出について説明する図である。

【図3】音声発話処理を説明するフローチャートである。

【図4】理解行動判定処理を説明するフローチャートである。

【図5】対話進行制御処理を説明するフローチャートである。

【図6】対話シーケンスの例について説明する図である。

【図7】理解行動要求位置の検出について説明する図である。

【図8】理解行動要求位置の検出について説明する図である。

【図9】コンピュータの構成例を示す図である。

30

## 【発明を実施するための形態】

## 【0033】

以下、図面を参照して、本技術を適用した実施の形態について説明する。

## 【0034】

## 第1の実施の形態

## 音声対話システムの構成例

本技術は、発話テキストを、ユーザに対して理解行動を要求する区間ごとに区切り、その区切りの位置でユーザに理解行動を求めるとともに、ユーザの理解行動と発話テキストとに基づいて次の発話音声を生成することで、適切な対話制御を行うことができるようになるものである。

40

## 【0035】

特に、本技術は以下の2つの技術的な特徴を有している。

## 【0036】

まず、第1の特徴として、本技術では音声発話する文章である発話テキストに対する自然言語解析によりユーザの相槌や頷きによる理解行動を求めるポイントが発話文章から抽出される。そして、音声発話時に発話テキストの途中で非完了文言や抑揚（イントネーション）を付加したり、長い間の時間を挿入したりすることによりユーザの理解行動を誘発す

50

るようになされる。

【0037】

ここで、ユーザによる理解行動とは、ユーザが発話内容を理解しているか否かを示す頷きや相槌などの行動をいう。また、ユーザに理解行動を求めるポイント、つまり位置は、例えばユーザに対して伝達する羅列情報の切れ目の位置や、係り受けの分岐位置などとされる。このように、自然言語解析の結果に基づいてユーザに理解行動を求める位置を抽出することで、発話テキストごとに適切な粒度で、ユーザの理解行動を誘発することができる。これにより、適切な対話制御を実現することができる。

【0038】

なお、以下では、ユーザに対して理解行動を求めるポイント（位置）を、特に理解行動要求位置とも称することとする。

10

【0039】

また、第2の特徴として、本技術ではユーザの「相槌」や「頷き」などの理解行動の種別、すなわちユーザの理解行動が肯定的な行動であるか、または否定的な行動であるかの判定と、ユーザの理解行動までのレスポンス時間の測定とが行われる。そして、それらの判定結果と測定結果とに基づいて、ユーザの理解度や音声発話に対する意識の集中度が推定、判定され、その理解度や集中度に応じて、適宜、理解行動誘発を行いながらの発話進行が動的に制御される。

【0040】

具体的には、例えば以下の制御C1乃至制御C4が行われる。

20

【0041】

（制御C1）

発話テキストにおける理解行動要求位置で発話音声を一旦停止させ、ユーザによる肯定的な理解行動（了解）があるまでは発話の間をあけて発話を先に進めないように対話制御が行われる。

【0042】

（制御C2）

理解行動の誘発後、一定時間が経過してもユーザによる理解行動がない場合には、直前の発話内容を繰り返したり、理解行動を催促する文言を挿入したり、それらの両方を行ったりするように対話制御が行われる。

30

【0043】

（制御C3）

ユーザによる理解行動の種別が否定的であった場合、直前の発話内容を繰り返すように対話制御が行われる。

【0044】

（制御C4）

ユーザの理解行動のレスポンス時間を測定し、そのレスポンス時間が短い場合にはユーザの理解度や集中度が高いと判定して、ユーザへの理解行動の誘発回数を減らし、発話音声の話速やトーンを上げていき、ユーザの理解度や集中度が低いと判定されたときには、誘発回数を増やし、発話音声の話速やトーンを下げていくような対話制御が行われる。

40

【0045】

本技術では、以上のような対話制御を行うことで、ユーザの理解度や集中度に応じて適切なタイミングで理解行動を促し、自然な音声対話を実現することができる。

【0046】

続いて、本技術を適用した具体的な実施の形態について説明する。

【0047】

図1は、本技術を適用した音声対話システムの一実施の形態の構成例を示す図である。

【0048】

図1に示す音声対話システムは、サーバ11およびクライアントデバイス12からなり、ユーザとの音声対話を行うシステムである。この例では、サーバ11と、スマートホンな

50

どの端末装置からなるクライアントデバイス 1 2 とが有線や無線などにより直接的に、または通信網を介して間接的に接続されている。

【 0 0 4 9 】

サーバ 1 1 は対話制御装置として機能し、ユーザからの理解行動というフィードバックと、発話内容を示すテキストデータである発話テキストとから音声発話の音声データを生成し、クライアントデバイス 1 2 に出力する。

【 0 0 5 0 】

クライアントデバイス 1 2 は、サーバ 1 1 から供給された音声データに基づいてユーザに対して発話音声を出力するとともに、適宜、ユーザからのフィードバックを受けて、そのフィードバックをサーバ 1 1 に供給する。

10

【 0 0 5 1 】

サーバ 1 1 は、自然言語解析部 2 1、理解行動要求位置検出部 2 2、音声出力対話進行制御部 2 3、音声合成エンジン 2 4、音声認識エンジン 2 5、行動認識エンジン 2 6、および理解行動判定部 2 7 を有している。また、クライアントデバイス 1 2 は、音声再生デバイス 3 1、音声入力デバイス 3 2、およびセンサ 3 3 を有している。

【 0 0 5 2 】

なお、ここでは音声合成エンジン 2 4 がサーバ 1 1 に設けられる例について説明するが、音声合成エンジン 2 4 はクライアントデバイス 1 2 に設けられるようにしてもよい。

【 0 0 5 3 】

サーバ 1 1 の自然言語解析部 2 1 には、音声発話によってユーザに対して提示する情報内容の全文章である発話テキストが入力される。自然言語解析部 2 1 は、入力された発話テキストに対して自然言語解析を行って、その解析結果と発話テキストを理解行動要求位置検出部 2 2 に供給する。

20

【 0 0 5 4 】

理解行動要求位置検出部 2 2 は、自然言語解析部 2 1 から供給された解析結果と発話テキストに基づいて、発話テキストにより示される文章内における、ユーザへの理解行動を要求する理解行動要求位置を検出する。

【 0 0 5 5 】

この理解行動要求位置は、発話テキストに基づく発話時にユーザの理解行動を誘発する発話を行う位置の候補となる位置である。理解行動要求位置検出部 2 2 が発話テキストの文章から理解行動要求位置を検出すると、1 または複数の理解行動要求位置で区切られた発話テキストが得られたことになる。

30

【 0 0 5 6 】

理解行動要求位置検出部 2 2 は、理解行動要求位置の検出結果と発話テキストを音声出力対話進行制御部 2 3 に供給する。

【 0 0 5 7 】

音声出力対話進行制御部 2 3 は、理解行動要求位置検出部 2 2 から供給された理解行動要求位置の検出結果および発話テキストと、理解行動判定部 2 7 から供給されたユーザの理解行動の判定結果とに基づいて、発話テキストに基づく音声発話を制御する。

【 0 0 5 8 】

すなわち、音声出力対話進行制御部 2 3 は、発話テキストに基づく発話文言を示すテキストデータを適切なタイミングで音声合成エンジン 2 4 に出力することで、ユーザとの対話制御を行う。このとき、音声出力対話進行制御部 2 3 は、必要に応じて、発話テキストにより示される発話文言に対して、ユーザの理解行動を誘発するための非完了文言を付加して出力したり、音声合成エンジン 2 4 に対して抑揚の付加を指示したり、発話文言の出力タイミングを制御することで発話の間をあけたりする。

40

【 0 0 5 9 】

なお、以下、発話文言に対して付加される、ユーザの理解行動を誘発するための非完了文言を理解行動誘発文言とも称することとする。

【 0 0 6 0 】

50

また、ここでは発話テキストにより示される発話文言に対して、適宜、理解行動要求位置において理解行動誘発文言が付加され、ユーザに対して理解行動を誘発する例について説明するが、ユーザ等による設定により、理解行動の誘発機能をオンまたはオフできるようにしてもよい。また、理解行動の誘発機能をオンするか、またはオフするかの設定結果が音声出力対話進行制御部 2 3 に記録できるようにしてもよい。

【 0 0 6 1 】

この場合、例えば理解行動の誘発機能がオフに設定されている場合には、音声出力対話進行制御部 2 3 は、特に理解行動誘発文言の付加等を行わず、発話テキストをそのまま音声合成エンジン 2 4 へと出力する。

【 0 0 6 2 】

音声合成エンジン 2 4 は、音声出力対話進行制御部 2 3 から供給された発話文言に基づいて、テキスト音声合成 (TTS) により発話文言を音声で再生するための音声データを生成し、音声再生デバイス 3 1 へと出力する。すなわち、音声合成エンジン 2 4 では、発話文言を示すテキストデータが、その発話文言の音声データへと変換される。

【 0 0 6 3 】

音声再生デバイス 3 1 は、例えばスピーカなどからなる音声出力部を有しており、音声合成エンジン 2 4 から供給された音声データに基づいて発話文言の音声を再生する。

【 0 0 6 4 】

発話文言の音声が再生されると、ユーザはその再生音を聞き、状況に応じて相槌や頷きなどの理解行動を行う。換言すれば、理解行動の誘発に対するフィードバックとして、相槌や頷きなどの理解行動がユーザにより行われる。

【 0 0 6 5 】

音声入力デバイス 3 2 は、例えばマイクロホンなどからなり、理解行動としてのユーザの相槌の音声を收音し、その結果得られた音声收音データを音声認識エンジン 2 5 に供給する。音声認識エンジン 2 5 は、音声入力デバイス 3 2 から供給された音声收音データに対して音声認識を行い、その音声認識の結果を理解行動判定部 2 7 に供給する。例えば音声認識によって音声收音データに基づく相槌音声がテキスト化され、得られたテキストデータが音声認識の結果として理解行動判定部 2 7 に出力される。

【 0 0 6 6 】

センサ 3 3 は、例えばユーザの頭部などに取り付けられたジャイロセンサや、ユーザ近傍に配置された、ユーザの頭部の動きを検出するセンサ、ユーザの頭部の画像を撮像するイメージセンサなどからなり、ユーザの理解行動としての頷き、すなわちユーザの頭部の動きを検出し、その検出結果を行動認識エンジン 2 6 に供給する。

【 0 0 6 7 】

行動認識エンジン 2 6 は、センサ 3 3 から供給されたユーザの頷きの検出結果に基づいて行動認識を行うことで、ユーザの頷きの種別を認識 (判定) し、その認識結果を理解行動判定部 2 7 に供給する。

【 0 0 6 8 】

理解行動判定部 2 7 は、音声認識エンジン 2 5 からの音声認識結果、および行動認識エンジン 2 6 からの頷きの種別の認識結果のうち少なくとも何れか一方に基づいて、ユーザの理解行動が肯定的な理解行動であるか、または否定的な理解行動であるかを判定し、その判定結果を音声出力対話進行制御部 2 3 に供給する。

【 0 0 6 9 】

このようにして得られた理解行動の判定結果は、音声出力対話進行制御部 2 3 において、理解行動の誘発を行った発話の次の発話の制御に用いられる。

【 0 0 7 0 】

なお、理解行動判定部 2 7 での理解行動の判定処理では、音声認識エンジン 2 5 からの音声認識結果、および行動認識エンジン 2 6 からの頷きの種別の認識結果のうち少なくとも何れか一方が用いられるようにすればよい。

【 0 0 7 1 】

10

20

30

40

50

例えば音声認識結果に基づいて理解行動の判定が行われる場合、理解行動判定部 27 では、音声認識結果に基づいて理解行動としてのユーザの相槌音声が決定的な相槌であるか、または否定的な相槌であるかが判定される。

【0072】

具体的には、例えば相槌音声として「うん」や「はい」などを示すテキストデータが音声認識の結果として得られた場合、理解行動判定部 27 は、ユーザの理解行動は肯定的な理解行動である、つまり相槌音声は肯定的な相槌であると判定する。

【0073】

これに対して、例えば相槌音声として「え」や「なに」、「もう 1 回」などを示すテキストデータが音声認識の結果として得られた場合、理解行動判定部 27 は、ユーザの理解行動は否定的な理解行動である、つまり相槌音声は否定的な相槌であると判定する。

10

【0074】

なお、音声認識エンジン 25 において、相槌音声を認識することができなかった場合には、例えば音声区間検出などで、ユーザから何らかの音声発話反応があったときに理解行動判定部 27 が、ユーザによる理解行動がなされたとし、理解行動の判定結果を出力するようにしてもよい。

【0075】

そのような場合、例えば設計段階において音声認識エンジン 25 の性能、すなわち性能信頼度に応じて、相槌音声を認識することができなかったがユーザから何らかの音声発話反応があったときに、ユーザの理解行動を肯定的なものとするか、または否定的なものとするかを定めておけばよい。

20

【0076】

例えば音声認識エンジン 25 の性能信頼度が低い場合には、ユーザから何らかの音声発話反応があったときに、そのユーザの理解行動を否定的なものとする、音声発話（対話）が先に進まなくなってしまうので、肯定的な理解行動（肯定的な相槌音声）であると判定されるようにすればよい。

【0077】

これに対して、音声認識エンジン 25 の性能信頼度が高い場合には、ユーザの再度の相槌を正しく認識できる可能性が高いので、ユーザから何らかの音声発話反応があったときには、その理解行動が否定的な理解行動（否定的な相槌音声）であると判定されるようにすればよい。

30

【0078】

また、例えばユーザの頷きの種別の認識結果に基づいて理解行動の判定が行われる場合、理解行動判定部 27 では、その認識結果に基づいて、理解行動としてのユーザの頷きが、肯定的な頷きであるか、または否定的な頷きであるかが判定される。

【0079】

例えばセンサ 33 がユーザの頭部などに取り付けられたジャイロセンサや、ユーザの頭部の動きを検出するセンサである場合、行動認識エンジン 26 はセンサ 33 からの出力に基づいて、ユーザが頭部（首）を縦方向に振る動きをしたか、またはユーザが頭部（首）を横方向に振る動きをしたかといった頷きの種別を認識する。

40

【0080】

そして、理解行動判定部 27 は、その頷きの種別の認識結果に基づいて、ユーザが頭部を縦方向に振ったとの認識結果が得られた場合には、ユーザの理解行動は肯定的な理解行動であると判定する。これに対して、理解行動判定部 27 は、ユーザが頭部を横方向に振ったとの認識結果が得られた場合には、ユーザの理解行動は否定的な理解行動であると判定する。

【0081】

また、例えばセンサ 33 がユーザの頭部の画像を撮像するイメージセンサなどからなる場合には、行動認識エンジン 26 において、センサ 33 で得られた画像に対して画像認識が行われ、ユーザの縦方向や横方向への頷き、つまり頷きの種別が認識される。

50

## 【 0 0 8 2 】

さらに、例えば音声認識エンジン 2 5 による音声認識結果と、行動認識エンジン 2 6 による頷きの種別の認識結果との両方が得られた場合には、判定エンジン、つまり音声認識エンジン 2 5 や行動認識エンジン 2 6 の認識の信頼度（認識精度）に応じて認識結果に優先順位をつけて理解行動の判定を行うようにしてもよい。

## 【 0 0 8 3 】

そのような場合、例えば行動認識エンジン 2 6 よりも音声認識エンジン 2 5 の信頼度が高いときには、理解行動判定部 2 7 は音声認識エンジン 2 5 による音声認識結果に基づいて理解行動の判定の処理を行う。また、理解行動判定部 2 7 は、音声認識エンジン 2 5 において音声認識が正しく行うことができなかつた場合には、行動認識エンジン 2 6 による頷きの種別の認識結果に基づいて理解行動の判定の処理を行う。

10

## 【 0 0 8 4 】

さらに、例えば音声認識の結果として「うーん」などの曖昧な認識結果、すなわち肯定的とも否定的とも判定できる認識結果が得られた場合には、理解行動判定部 2 7 がユーザの相槌音声の抑揚やユーザの動き、ユーザの視線方向、ユーザの視線の動きなども用いて理解行動の判定を行うようにしてもよい。この場合、どのような音声認識結果が曖昧な認識結果とされるかは、予め曖昧な認識結果とされる単語を登録するなどして定めておけばよい。

## 【 0 0 8 5 】

例えば曖昧な音声認識結果が得られた場合、理解行動判定部 2 7 は行動認識エンジン 2 6 による頷きの種別の認識結果に基づいて、ユーザの理解行動を判定する。

20

## 【 0 0 8 6 】

また、例えば音声対話システムにおいて「うーん」などの曖昧な相槌音声について、その相槌音声を発しているときのユーザの頷きや視線方向、視線の動きなどのユーザの動き、またはユーザの音声の抑揚をフィードバック付き学習し、その学習結果を示す辞書を生成するようにしてもよい。

## 【 0 0 8 7 】

そのような場合、フィードバック付き学習時にはユーザは相槌音声を発するとともに、その相槌音声が肯定的なものであるか、または否定的なものであるかの入力を行う。そして、音声対話システムでは、その入力結果と、音声認識により得られた相槌音声の認識結果と、検出されたユーザの頷き等の動きまたは音声の抑揚とから辞書が生成される。

30

## 【 0 0 8 8 】

理解行動判定部 2 7 は、実際の音声の発話時に曖昧な相槌音声であるとの音声認識結果が得られたときには、音声入力デバイス 3 2 で収録された音声収録データから得られた音声の抑揚や、センサ 3 3 等で得られたユーザの動きなどと、予め記録している辞書とに基づいてユーザの理解行動を判定する。

## 【 0 0 8 9 】

なお、以上においては、ユーザの理解行動を検出するための構成として、クライアントデバイス 1 2 に音声入力デバイス 3 2 およびセンサ 3 3 が設けられている例について説明したが、その他、どのような方法によりユーザの理解行動が検出されるようにしてもよい。

40

## 【 0 0 9 0 】

例えばクライアントデバイス 1 2 に特定のボタンやタッチパネルを設け、ユーザがそのボタンを押下したり、タッチパネルをタッチしたりするようにしてもよい。すなわち、ユーザによるボタンやタッチパネル等に対する操作が理解行動として行われ、ユーザの音声対話システムに対するフィードバックが行われるようにしてもよい。

## 【 0 0 9 1 】

また、クライアントデバイス 1 2 に視線検出センサを設け、その視線検出センサの出力に基づいて、ユーザの理解行動の判定が行われてもよい。そのような場合、例えばユーザが対象機器などの特定の方向に視線を向けたことが検出されたときに、ユーザの理解行動が肯定的な理解行動であると判定される。

50

## 【 0 0 9 2 】

さらに、クライアントデバイス 1 2 においては、音声入力デバイス 3 2 やセンサ 3 3 は、音声再生デバイス 3 1 と一体的に設けられているようにしてもよいし、音声再生デバイス 3 1 が設けられた装置とは別に設けられていてもよい。

## 【 0 0 9 3 】

具体的には、例えば音声再生デバイス 3 1 がスマートホンに設けられており、音声入力デバイス 3 2 やセンサ 3 3 がそのスマートホンに接続されているようにしてもよい。そのような場合には、例えばスマートホンが音声入力デバイス 3 2 やセンサ 3 3 から、音声収音データやユーザの動きの検出結果を取得して、音声認識エンジン 2 5 や行動認識エンジン 2 6 に送信するようにしてもよい。

10

## 【 0 0 9 4 】

自然言語解析と理解行動要求位置について

続いて、理解行動要求位置検出部 2 2 による理解行動要求位置の検出方法の具体的な例について説明する。

## 【 0 0 9 5 】

例えば発話テキストとして、図 2 の矢印 A 1 1 に示すように「今日のお出かけは、財布と携帯電話と鍵と弁当と時計を持って行ってください。夕方から雨が降るかもしれないので傘を持って行ってください。」というテキストデータが自然言語解析部 2 1 に入力されたとする。

## 【 0 0 9 6 】

このような場合に、自然言語解析部 2 1 において発話テキストに対する自然言語解析として例えば形態素解析が行われると、矢印 A 1 2 に示す解析結果が得られる。形態素解析では、入力された発話テキストの文章が文節単位に区切られ、各文節の言語的な格の情報と、各文節の係り受けの構造を示す情報とが解析結果として得られる。

20

## 【 0 0 9 7 】

矢印 A 1 2 に示す部分では各四角形が 1 つの文節を表しており、その四角形内には発話テキストの文節に区切られた文言、つまり文節を構成する文言が記されている。また、各文節を表す四角形同士を結ぶ線分は、文節同士の係り受けを表している。

## 【 0 0 9 8 】

すなわち、文節を表す四角形から延びる線分の終点は、その文節の係り先の文節となっている。したがって、例えば図中、一番上側の文節「今日の」は、文節「お出かけは」に係っていることが分かる。

30

## 【 0 0 9 9 】

また、矢印 A 1 2 に示す部分では、各文節を表す四角形の図中、右側には、それらの文節の言語的な格が記されている。例えば図中、一番上側の文節「今日の」の言語的な格は連体修飾節であることが記されている。

## 【 0 1 0 0 】

このように自然言語解析部 2 1 では、自然言語解析の結果として、矢印 A 1 2 に示すように文節の言語的な格の情報と、各文節の係り受けの構造を示す情報とが得られる。

## 【 0 1 0 1 】

理解行動要求位置検出部 2 2 は、自然言語解析部 2 1 から供給されたそれらの情報に基づいて、ユーザに対して理解行動を要求する位置、つまり、理解行動を誘発するための理解行動誘発文言を挿入（付加）する理解行動要求位置を検出する処理を行う。

40

## 【 0 1 0 2 】

具体的には、理解行動要求位置検出部 2 2 は、1 つの文の中で同一の述語節の文節に係る複数の対象格の文節が存在するかを検出する。

## 【 0 1 0 3 】

例えば矢印 A 1 2 に示した発話テキストの 1 文目の例では、複数の文節「財布と」、「携帯電話と」、「鍵と」、「弁当と」、および「時計を」がそれぞれ対象格となっており、それらの文節が、述語節である同一の文節「持って行ってください。」に係っている。

50

## 【0104】

この例では、発話テキストの1文目は、お出かけの際にユーザが持っていくべきもの（アイテム）、つまりお出かけの際の持ち物リストを示す文章となっており、その持ち物リストにより示される持ち物（アイテム）の文節が同一の述語節の文節に係っている。すなわち、持ち物が羅列されている文章となっている。

## 【0105】

このように、1つの文の中で同一の述語節の文節に係る複数の対象格の文節は、この例で示したお出かけ時の持ち物リストのようなユーザに対して提示される複数の羅列情報であることを意味している。換言すれば、同一の述語節の文節に複数の対象格の文節に係る文の構造は、ユーザに対して複数の羅列情報を提示していることを意味している。

10

## 【0106】

そこで、このような同一の述語節の文節に係る複数の対象格の文節、すなわち羅列情報が検出された場合、理解行動要求位置検出部22は、それらの同一の述語節の文節に係る各対象格の文節の間の位置、つまり各羅列情報の文節の間の位置を理解行動要求位置とする。

## 【0107】

これは、各羅列情報の文節の間の位置は文として完結していない非完結な位置であり、まだ続きがあることをユーザに連想させるため、これらの文節の間の位置では理解行動が誘発されやすく、かつ各アイテム（持ち物）に対する了解をユーザから得ることができるので、理解行動を求める位置として有効であるからである。

## 【0108】

図2の例では、発話テキストの1文目の矢印W11乃至矢印W14に示される位置が理解行動要求位置検出部22により検出された理解行動要求位置となっている。具体的には、例えば矢印W11により示される理解行動要求位置は、文節「財布と」の直後の位置となっている。

20

## 【0109】

また、理解行動要求位置検出部22は、発話テキストに複数の文が含まれる場合は、文と文の区切りの位置を理解行動要求位置とは判定しない。すなわち、2つ目以降の文では、その文の直前にある文の終了位置は、理解行動要求位置とはされず、文が続けて発話される。これは、文と文の間は、文として完結した位置であるから理解行動を誘発することが困難であるからである。例えば、文として完結している箇所でユーザの理解行動を求めた場合、ユーザは音声対話システムによる発話が終了したと受け取ってしまうことがある。

30

## 【0110】

さらに、理解行動要求位置検出部22は、発話テキストに複数の文が含まれる場合、2つ目以降の文において、文の述語節の文節に対する最初の係り受けが発生する文節を検出し、その文節の直後の位置を理解行動要求位置とする。

## 【0111】

これは、文の述語節に対する最初の係り受けが発生する文節の位置は、前述した例と同様に文として非完結の状態であり、さらに2つ目以降の文でも音声対話システムによる発話に対するユーザの集中度が持続しているかを確認するのに有効な位置（タイミング）となるからである。

40

## 【0112】

図2の例では、2つ目の文において述語節である文節「持って行ってください。」に係る最初の文節「降るかもしれないので」の直後の位置、すなわち矢印W15に示す位置が理解行動要求位置とされている。ここでは、矢印W15に示すように連用修飾節の直後の位置が理解行動要求位置とされるので、文として完結しない位置でユーザに対して理解行動が要求されることになる。

## 【0113】

なお、2つ目以降の文においても、羅列情報が検出された場合、つまり同一の述語節の文節に係る複数の対象格の文節が検出された場合、それらの対象格の文節の間の位置が理解行動要求位置とされるようにしてもよい。

50

## 【 0 1 1 4 】

以上の例では、理解行動要求位置検出部 2 2 は、発話テキストの文章内の述語節の文節に係る文節位置に基づいて、理解行動要求位置を検出（決定）している。すなわち、述語節との係り受けに基づく位置が理解行動要求位置として検出される。

## 【 0 1 1 5 】

音声発話処理および理解行動判定処理の説明

次に、図 1 に示した音声対話システムの動作について説明する。

## 【 0 1 1 6 】

例えば音声対話システムに発話テキストが供されると、音声対話システムは、音声発話を行ってユーザとの対話を行う音声発話処理と、音声対話システムの音声発話に対するユーザからのフィードバックに基づいて理解行動を判定する理解行動判定処理を行う。これらの音声発話処理と理解行動判定処理は同時に行われる。

10

## 【 0 1 1 7 】

まず、図 3 のフローチャートを参照して、音声対話システムによる音声発話処理について説明する。

## 【 0 1 1 8 】

ステップ S 1 1 において、自然言語解析部 2 1 は、供給された発話テキストに対して自然言語解析を行い、その解析結果と発話テキストを理解行動要求位置検出部 2 2 に供給する。例えばステップ S 1 1 では、図 2 を参照して説明したように形態素解析が行われる。

## 【 0 1 1 9 】

ステップ S 1 2 において、理解行動要求位置検出部 2 2 は、自然言語解析部 2 1 から供給された解析結果と発話テキストに基づいて理解行動要求位置を検出し、その検出結果と発話テキストを音声出力対話進行制御部 2 3 に供給する。

20

## 【 0 1 2 0 】

例えばステップ S 1 2 では、図 2 を参照して説明したように羅列情報である対象格の文節の間の位置や、2 つ目以降の文における述語節に対する最初の係り受けが発生する文節の直後の位置などが理解行動要求位置として検出される。

## 【 0 1 2 1 】

ステップ S 1 3 において、音声出力対話進行制御部 2 3 は、理解行動要求位置検出部 2 2 から供給された理解行動要求位置の検出結果および発話テキストと、理解行動判定部 2 7 から供給されたユーザの理解行動の判定結果とに基づいて、次に発話させる発話文言のテキストデータを音声合成エンジン 2 4 に出力する。このとき、音声出力対話進行制御部 2 3 は、音声合成エンジン 2 4 に対して、必要に応じて理解行動誘発文言部分での抑揚の付加も指示する。また、音声出力対話進行制御部 2 3 は、適宜、発話文言のテキストデータの出力タイミングを制御することで発話の間をあける。

30

## 【 0 1 2 2 】

ここで、ステップ S 1 3 において用いられる理解行動の判定結果は、図 4 を参照して後述する理解行動判定処理により得られたものとされる。

## 【 0 1 2 3 】

音声出力対話進行制御部 2 3 は、理解行動要求位置ごとに次に発話させる発話文言を決定し、決定した発話文言のテキストデータを出力することで、ユーザとの音声対話の進行を制御する。また、音声出力対話進行制御部 2 3 は、理解行動誘発文言が付加された発話文言のテキストデータを出力したときには、次の発話文言のテキストデータの出力までに、予め定めた長さの時間である間を設けて理解行動の誘発を行う。

40

## 【 0 1 2 4 】

なお、より詳細には、次に発話させる発話文言を決定するにあたっては、理解行動の判定結果だけでなく、ユーザの理解度および集中度を示す理解・集中度も用いられる。この理解・集中度は、ユーザが音声対話システムによる音声発話に対して、どれだけ意識を集中させ、発話内容を理解できているかを示すパラメータである。

## 【 0 1 2 5 】

50

ステップS 14において、音声合成エンジン24は、音声出力対話進行制御部23から供給された発話文言のテキストデータに基づいて、テキスト音声合成により発話文言を音声で再生するための音声データを生成し、音声再生デバイス31へと出力する。

【0126】

ステップS 15において、音声再生デバイス31は、音声合成エンジン24から供給された音声データに基づいて発話文言の音声を再生する。

【0127】

ユーザは、このようにして発話された音声に対して、適宜、理解行動によりフィードバックを行い、音声対話システムとの音声対話を行う。このときユーザにより行われた理解行動、つまりフィードバックに対して、図4を参照して後述する理解行動判定処理が行われる。

10

【0128】

ステップS 16において、音声出力対話進行制御部23は、ユーザとの音声対話を終了するか否かを判定する。例えば入力された発話テキストの内容が全て発話された場合、音声対話を終了すると判定される。

【0129】

ステップS 16において、音声対話を終了しないと判定された場合、まだ発話していない内容が残っているので、処理はステップS 13に戻り、上述した処理が繰り返し行われる。すなわち、次の文言の発話が行われる。

【0130】

これに対して、ステップS 16において音声対話を終了すると判定された場合、音声発話処理は終了する。

20

【0131】

以上のようにして音声対話システムは、発話テキストに対する自然言語解析の結果に基づいて理解行動要求位置を検出し、その検出結果と理解行動の判定結果とに基づいて次の文言を発話する。このようにすることで、適切なタイミングでユーザに対して相槌や頷きなどの負荷の低い理解行動を誘発し、対話を進めることができる。すなわち、より適切な対話制御を行うことができる。

【0132】

続いて、図4のフローチャートを参照して、図3を参照して説明した音声発話処理と同時に、音声対話システムによる理解行動判定処理について説明する。

30

【0133】

図3のステップS 15で音声対話システムにより理解行動を誘発する音声発話が行われると、ユーザはその音声発話に対して理解行動を行う。例えばユーザは理解行動として、相槌を打ったり頷いたりするという否定的または肯定的な行動を行う。

【0134】

すると、ステップS 41において、音声入力デバイス32は、理解行動としてのユーザの相槌の音声を収録し、その結果得られた音声収録データを音声認識エンジン25に供給する。

【0135】

ステップS 42において、音声認識エンジン25は、音声入力デバイス32から供給された音声収録データに対して音声認識を行い、その音声認識の結果を理解行動判定部27に供給する。

40

【0136】

また、ステップS 43において、センサ33は、ユーザの頭部の動き、つまり頷きをユーザの理解行動として検出し、その検出結果を行動認識エンジン26に供給する。

【0137】

ステップS 44において、行動認識エンジン26は、センサ33から供給されたユーザの理解行動の検出結果に基づいて行動認識を行うことで、ユーザの頷きの種別を認識(判定)し、その認識結果を理解行動判定部27に供給する。

50

## 【 0 1 3 8 】

なお、より詳細には、ステップ S 4 1 およびステップ S 4 2 の処理と、ステップ S 4 3 およびステップ S 4 4 の処理とは並行して行われる。また、ステップ S 4 1 およびステップ S 4 2 の処理と、ステップ S 4 3 およびステップ S 4 4 の処理とのうちの何れか一方のみが行われるようにしてもよい。

## 【 0 1 3 9 】

ステップ S 4 5 において、理解行動判定部 2 7 は、音声認識エンジン 2 5 からの音声認識結果、および行動認識エンジン 2 6 からの領きの種別の認識結果のうちの少なくとも何れか一方に基づいて理解行動判定を行い、その判定結果を音声出力対話進行制御部 2 3 に供給する。

10

## 【 0 1 4 0 】

理解行動判定では、ユーザの理解行動が肯定的な理解行動であるか、または否定的な理解行動であるかが判定され、その判定結果が次に行われる図 3 のステップ S 1 3 の処理で用いられる。換言すれば、理解行動判定によって、ユーザによる肯定的または否定的な理解行動が検出されることになる。

## 【 0 1 4 1 】

例えば理解行動判定では、音声認識エンジン 2 5 での音声認識の結果として、ユーザの相槌音声「うん」や「はい」などが検出された場合、ユーザの理解行動は肯定的な理解行動であると判定され、肯定的な理解行動である旨の判定結果が出力される。

## 【 0 1 4 2 】

また、例えば行動認識エンジン 2 6 での領きの種別の認識結果として、ユーザが頭部を縦方向に振ったとの認識結果が得られた場合には、ユーザの理解行動は肯定的な理解行動であると判定され、肯定的な理解行動である旨の判定結果が出力される。

20

## 【 0 1 4 3 】

なお、上述したように音声認識エンジン 2 5 からの音声認識結果、および行動認識エンジン 2 6 からの領きの種別の認識結果の何れか一方のみが用いられて理解行動判定が行われてもよいし、それらの両方が用いられて理解行動判定が行われてもよい。

## 【 0 1 4 4 】

ステップ S 4 6 において、理解行動判定部 2 7 は、理解行動判定を終了するか否かを判定する。例えば図 3 を参照して説明した音声発話処理において、音声対話を終了するとされた場合、理解行動判定を終了すると判定される。

30

## 【 0 1 4 5 】

ステップ S 4 6 において理解行動判定を終了しないと判定された場合、処理はステップ S 4 1 に戻り、上述した処理が繰り返し行われる。すなわち、ユーザによる次の理解行動について理解行動判定が行われる。

## 【 0 1 4 6 】

これに対して、ステップ S 4 6 において理解行動判定を終了すると判定された場合、理解行動判定処理は終了する。

## 【 0 1 4 7 】

以上のようにして音声対話システムは、ユーザとの音声対話が始まると、その対話時におけるユーザの理解行動に対して、肯定的な理解行動であるか、または否定的な理解行動であるかの理解行動判定を行う。このようにして理解行動判定を行うことで、その判定結果を用いて、より適切な対話制御を行うことができるようになる。

40

## 【 0 1 4 8 】

## 対話進行制御処理の説明

続いて、図 3 を参照して説明した音声発話処理のステップ S 1 3 において、音声出力対話進行制御部 2 3 が理解行動要求位置ごとに、次に発話させる発話文言を決定して音声発話、つまり音声対話の進行を制御する処理である対話進行制御処理の具体例について説明する。すなわち、以下、図 5 のフローチャートを参照して、音声出力対話進行制御部 2 3 により行われる対話進行制御処理について説明する。

50

## 【 0 1 4 9 】

この対話進行制御処理は、図 3 のステップ S 1 3 において、発話テキストの文章における理解行動要求位置ごとに行われる。すなわち、発話テキストの文章の理解行動要求位置の直前の文節を音声発話するタイミングで対話進行制御処理が行われる。

## 【 0 1 5 0 】

以下では、特に発話テキストの文章における理解行動要求位置の直前の文節の文言、つまりこれから発話される文言を特に処理対象文言とも称することとする。

## 【 0 1 5 1 】

ステップ S 7 1 において、音声出力対話進行制御部 2 3 は、保持している理解・集中度が予め定められた所定の閾値  $t_{h1}$  以上であるか否かを判定する。

10

## 【 0 1 5 2 】

ここで、理解・集中度は、ユーザが音声対話システムによる音声発話に対して、どれだけ意識を集中させ、発話内容を理解できているかを示すパラメータであり、例えば音声対話開始時には、予め定められた初期値とされる。このとき、理解・集中度の初期値は、例えば閾値  $t_{h1}$  未満の値となるように定められる。

## 【 0 1 5 3 】

理解・集中度の値は、例えば理解行動要求位置で文言を発した直後のタイミングなど、ユーザとの音声対話を行っていく過程で累積的に更新されていき、更新された理解・集中度の値は次の理解行動要求位置等で使用される。

## 【 0 1 5 4 】

さらに、ここでは理解・集中度の値が大きいほどユーザが意識を集中させ、高い理解度で対話を行っているものとする。すなわち、理解・集中度の値が大きいほど、音声対話に対するユーザの集中度および理解度が高いものとする。

20

## 【 0 1 5 5 】

ステップ S 7 1 において理解・集中度が閾値  $t_{h1}$  以上であると判定された場合、つまりユーザの集中度および理解度が十分に高い状態であると判定された場合、処理はステップ S 7 2 へと進む。

## 【 0 1 5 6 】

ステップ S 7 2 において、音声出力対話進行制御部 2 3 は発話時における発話速度の設定を行う。

30

## 【 0 1 5 7 】

例えばステップ S 7 2 が行われる状態では、ユーザが十分集中および理解している状態であるので、ある程度速い速度で処理対象文言が発話されるように発話速度が定められる。具体的には、例えば発話速度が現在の理解・集中度の値に対して定められた速度となるように発話速度が決定される。この場合、理解・集中度の値が大きいほど、発話速度が速くなるようになされる。また、例えば発話速度がこれまでの値から所定値だけ上がるように発話速度が決定されるようにしてもよい。

## 【 0 1 5 8 】

ステップ S 7 3 において、音声出力対話進行制御部 2 3 は、処理対象文言のテキストデータを音声合成エンジン 2 4 に出力し、理解行動の誘発をせずに、ステップ S 7 2 で定められた発話速度で処理対象文言が発話されるように処理対象文言の音声が発話させる。

40

## 【 0 1 5 9 】

この場合、音声合成エンジン 2 4 は、音声出力対話進行制御部 2 3 の制御に従って、供給された処理対象文言が指定された発話速度で再生される音声データを生成し、音声再生デバイス 3 1 に供給する。このとき、理解行動の誘発は行われないので、処理対象文言には、理解行動誘発文言は付加されない。また、音声再生デバイス 3 1 は音声合成エンジン 2 4 からの音声データに基づいて、処理対象文言の音声を再生する。このようにして処理対象文言の音声データを生成し、再生する処理は、図 3 のステップ S 1 4 およびステップ S 1 5 の処理に対応する。

## 【 0 1 6 0 】

50

このような場合、処理対象文言に理解行動誘発文言が付加されない状態で、かつ比較的速い速度で処理対象文言の音声が発話される。このように、ユーザの集中度や理解度が高い状態では、理解行動要求位置であっても特に理解行動の誘発をせずに、比較的速い速度で音声発話を行うことで、ユーザの集中度や理解度、つまりユーザの状況に応じた適切な音声対話を実現することができる。

【0161】

なお、ここでは理解・集中度が閾値  $t_{h1}$  以上であるか否か、つまり理解・集中度に応じて発話速度が変更（決定）される場合について説明するが、発話速度だけでなく発話音声のトーンも変更（決定）されるようにしてもよい。この場合、例えば理解・集中度の値が大きく、ユーザの集中度および理解度が高いほど、発話音声のトーンが高くなるように、処理対象文言のトーンが決定されるようにすることができる。

10

【0162】

ステップ S74 において、音声出力対話進行制御部 23 は、保持している理解・集中度の値を予め定められた微小値だけ下げた後、処理対象文言を発話テキストにおける、発話が終了した文言の次の文言、つまり次の文節の文言として対話進行制御処理は終了する。

【0163】

ステップ S74 において、理解・集中度が微小値だけ下げられる（減少させる）のは、その後の対話において、理解行動誘発が1度も行われなくなってしまうことを防止するためである。このようにユーザの集中度や理解度が十分に高い場合でも、適宜、理解行動が誘発されるように理解・集中度の更新を行うことで、ユーザの集中を維持させ、より適切な対話制御を実現することができる。

20

【0164】

また、ステップ S71 において、理解・集中度が閾値  $t_{h1}$  以上でないと判定された場合、すなわちユーザの集中度や理解度が十分に高いとはいえない場合、処理はステップ S75 へと進む。

【0165】

ステップ S75 において、音声出力対話進行制御部 23 は、処理対象文言に対して理解行動誘発文言を付加するとともに、その理解行動誘発文言の部分、つまり理解行動誘発文言が付加された文言（発話）の語尾の抑揚が上がるように抑揚の付加を行う。

【0166】

例えば理解行動誘発文言として「ね」や「ですね」などが付加されて、その語尾の抑揚が上げられる。具体的には、例えば処理対象文言が「財布と」である場合、理解行動誘発文言として「ね」が付加されて発話文言が「財布とね」とされるとともに、その語尾部分の「とね」の抑揚が上がるようになされる。

30

【0167】

ステップ S76 において、音声出力対話進行制御部 23 は、発話時における発話速度の設定を行う。

【0168】

ステップ S76 ではステップ S72 と同様にして発話速度の設定が行われるが、例えばステップ S76 の処理が行われる場合には、ユーザの集中度や理解度が十分に高いとはいえない状態であるので、処理対象文言が比較的遅い速度で発話されるように、発話速度が下げられる。すなわち、例えば発話速度が現時点での理解・集中度に対して定められた値となるように変更される。その他、例えば現時点での発話速度の値が予め定められた所定値だけ下げられるようにしてもよい。

40

【0169】

また、ステップ S76 においても、発話速度だけでなく発話のトーンも変更されるようにしてもよい。そのような場合、例えばユーザの集中度や理解度が低い時には、発話音声のトーンが低くなるように発話時のトーンが定められる。すなわち、理解・集中度に基づいて発話時のトーンの高さが定められる。

【0170】

50

ステップS 77において、音声出力対話進行制御部23は、理解行動誘発文言が付加された処理対象文言のテキストデータを音声合成エンジン24に出力し、理解行動の誘発が行われるように、理解行動誘発文言が付加された処理対象文言の音声を発話させる。すなわち、音声出力対話進行制御部23は、理解行動要求位置においてユーザの理解行動を誘発する発話を行わせる。

【0171】

この場合、音声出力対話進行制御部23は、音声合成エンジン24に対して理解行動誘発文言の付加された処理対象文言がステップS 76で定めた発話速度で発話されるように指示するとともに、語尾への抑揚の付加も指示する。また、音声出力対話進行制御部23は発話に間があげられるようにし、理解行動誘発文言と、発話の語尾の抑揚と、発話の間によって、ユーザの理解行動が誘発されるようにする。

10

【0172】

音声合成エンジン24は、音声出力対話進行制御部23の制御に従って、供給された文言、つまり処理対象文言と理解行動誘発文言を、指定された発話速度および抑揚で再生するための音声データを生成し、音声再生デバイス31に供給する。また、音声再生デバイス31は音声合成エンジン24からの音声データに基づいて、処理対象文言とそれに付加された理解行動誘発文言の音声を再生する。このようにして処理対象文言等の音声データを生成し、再生する処理は図3のステップS 14およびステップS 15の処理に対応する。

【0173】

音声の再生により、処理対象文言に理解行動誘発文言が付加され、語尾の抑揚が上がるように、かつ比較的遅い速度で処理対象文言の音声が発話される。このように、ユーザの集中度や理解度が十分でない状態では、ユーザに対して適宜、理解行動の誘発を行い、比較的遅い速度で音声発話を行うことで、ユーザの集中度や理解度に合わせた適切な音声対話を実現することができる。

20

【0174】

また、音声発話が行われると、ユーザによる理解行動の待ち状態となる。すなわち、音声出力対話進行制御部23による発話制御によって、ユーザの理解行動誘発のために間があげられる。

【0175】

この場合、音声出力対話進行制御部23は、音声合成エンジン24に音声発話の文言を出力したタイミング、つまり音声発話を指示したタイミングでレスポンス時間の計測を開始する。

30

【0176】

ここで、レスポンス時間とは、音声発話を指示してから、その音声発話に対してユーザが何らかの理解行動を行うまでの時間、すなわちユーザの応答時間である。

【0177】

処理対象文言と理解行動誘発文言が発話された後の理解行動の待ち状態では、音声対話システムによる発話の文は完結していない状態であることと、理解行動誘発文言と語尾に付加された抑揚と発話の間とから、ユーザによる理解行動が誘発されることになる。ユーザは、発話された音声に対して、相槌や頷きなどの理解行動を行って、自身の理解の状態をフィードバックする。

40

【0178】

なお、ステップS 77での音声発話時には、クライアントデバイス12においてLED (Light Emitting Diode) の点灯や他の視覚的な提示が可能である場合には、LEDの点灯やアイコン等の提示などにより、音声対話システムが停止状態でないことをユーザに提示するようにしてもよい。また、停止状態でないことのユーザへの提示は、その他、例えば定期的な効果音の再生などにより行われてもよい。さらに、ユーザに対して理解行動が要求されている旨を視覚的または聴覚的に提示(通知)するようにしてもよい。

【0179】

以上のようにしてユーザの理解行動の待ち状態となると、ステップS 78において、音声

50

出力対話進行制御部 23 は、レスポンス時間の計測を開始してから規定時間内に、ユーザによる理解行動があったか否かを判定する。すなわち、規定時間内にユーザの理解行動が検出されたかが判定される。

【0180】

例えばステップ S78 では、レスポンス時間の計測を開始してから規定時間内に、理解行動判定部 27 から理解行動の判定結果が供給された場合に、規定時間内に理解行動があったと判定される。

【0181】

なお、ここでの規定時間は予め定められた固定の時間とされてもよいし、理解・集中度などに応じて動的に定められる時間とされるようにしてもよい。例えば理解・集中度により規定時間が定められるときには、理解・集中度が高いほど規定時間を短くするなどすることで、ユーザの集中度および理解度に応じた適切な時間を定めることができる。

10

【0182】

ステップ S78 において、規定時間内に理解行動がなかったと判定された場合、音声発話がなされた後、しばらくしてもユーザは理解行動をしなかったので、ステップ S79 において、音声出力対話進行制御部 23 は保持している理解・集中度の値を下げる。

【0183】

ステップ S79 では、例えばステップ S74 の処理で理解・集中度が下げられるよりも、より大きく理解・集中度の値が下がるように、理解・集中度が更新される。これは、ユーザが理解行動をしなかったということは、ユーザが音声発話を聞いていなかったり、発話内容を十分に理解していなかったりしている状況であるからである。

20

【0184】

ステップ S80 において、音声出力対話進行制御部 23 は、ステップ S79 での更新後の理解・集中度が予め定められた所定の閾値  $t_{h2}$  以上であるか否かを判定する。

【0185】

ここで、閾値  $t_{h2}$  の値は、例えば上述した閾値  $t_{h1}$  の値よりも小さい、予め定められた値とされる。

【0186】

ステップ S80 において、理解・集中度が閾値  $t_{h2}$  以上でないと判定された場合、ステップ S81 において、音声出力対話進行制御部 23 は、ユーザは音声対話システムによる音声発話を全く聞いていないものとして、ユーザとの音声対話を終了（停止）させ、対話進行制御処理は終了する。この場合、その後の図 3 の音声発話処理のステップ S16 では音声対話を終了すると判定されることになる。

30

【0187】

なお、ステップ S81 で音声対話を終了させる場合、音声出力対話進行制御部 23 は、発話テキストを未発話状態であるとして、その発話テキスト等を記録しておき、しばらく時間が経過してから、再度、発話テキストの内容の音声発話を行うようにしてもよい。この場合、音声発話が行われるタイミングは、例えば一定時間が経過した後、再度、センサ 33 等によりユーザの存在が認識されたときや、他の発話テキストの音声発話の終了時など、ユーザの理解・集中度が高い状態のときなどとすることができる。

40

【0188】

未発話状態の発話テキストの音声発話を行う際には、発話テキストの文章の最初から音声発話を行うようにすればよい。

【0189】

また、ステップ S81 で音声対話を終了させる場合、すなわち音声対話を途中で停止させた場合、音声出力対話進行制御部 23 は、例えば音声対話を終了させた旨のメッセージと、発話テキストとが含まれる通知情報を生成し、通知情報の出力を制御してもよい。これにより、音声対話以外の方法でユーザに対して発話テキストの内容を通知することができる。

【0190】

50

そのような場合、例えば音声出力対話進行制御部 2 3 は、通知情報をサーバ 1 1 に設けられた図示せぬ通信部によりクライアントデバイス 1 2 としてのスマートフォン等の端末装置に送信し、通知情報により示される発話テキストの内容等をクライアントデバイス 1 2 に表示させる。例えば通知情報の送信方法、つまり通知方法は、電子メールや、クライアントデバイス 1 2 にインストールされたアプリケーションプログラムの通知機能など、どのような方法であってもよい。

【 0 1 9 1 】

これにより、例えばノーティフィケーション機能を利用して、クライアントデバイス 1 2 のステータスバーに通知情報を受信した旨等を表示させたり、通知情報をクライアントデバイス 1 2 の表示画面にポップアップ表示させたりすることができる。その結果、ユーザに発話テキストの内容を通知することができる。

10

【 0 1 9 2 】

また、発話テキストの内容を含む通知情報を送信するのではなく、発話テキストの内容の確認を促すメッセージのみが含まれる通知情報を送信するようにしてもよい。

【 0 1 9 3 】

一方、ステップ S 8 0 において、理解・集中度が閾値  $t_h 2$  以上であると判定された場合、ユーザは集中度および理解度は低いが、まだ音声発話を聞いているとし、処理はステップ S 8 2 へと進む。

【 0 1 9 4 】

ステップ S 8 2 において、音声出力対話進行制御部 2 3 は、ユーザに対して理解行動要求を行う音声を発話させる。

20

【 0 1 9 5 】

すなわち、音声出力対話進行制御部 2 3 は、例えば「いいですか？」など、ユーザに対して理解行動を直接的に促す（要求する）旨の文言である理解行動要求文言のテキストデータを音声合成エンジン 2 4 に出力し、音声発話を指示する。

【 0 1 9 6 】

すると、音声合成エンジン 2 4 は、音声出力対話進行制御部 2 3 の指示に従って、供給された理解行動要求文言のテキストデータから、その理解行動要求文言の音声データを生成し、音声再生デバイス 3 1 に供給する。また、音声再生デバイス 3 1 は音声合成エンジン 2 4 からの音声データに基づいて、理解行動要求文言の音声を再生する。これにより、ユーザに対して理解行動を要求する発話が行われる。このようにして理解行動要求文言の音声を再生することで、ユーザに理解行動を促すことができる。

30

【 0 1 9 7 】

なお、理解行動要求文言は「いいですか？」に限らず、他のどのような文言であってもよい。例えばユーザの理解・集中度が著しく低い場合には、理解行動要求文言として「聞いてる？」などの文言を発話するようにしてもよい。

【 0 1 9 8 】

また、ステップ S 8 2 において、理解行動要求文言を発話させるのではなく、現在の処理対象文言と、その処理対象文言に付加された理解行動誘発文言とが再度、発話されるようにしてもよい。すなわち、ステップ S 7 7 の処理での理解行動を誘発する音声発話を再度行うことで、直前の発話が繰り返されるようにしてもよい。

40

【 0 1 9 9 】

さらに、ステップ S 8 2 において、理解行動要求文言を発話させるとともに、現在の処理対象文言とその処理対象文言に付加された理解行動誘発文言とを再度、発話させるようにしてもよい。その他、ステップ S 7 8 で規定時間内に理解行動がなかったと判定された場合に、直前に行った処理対象文言と理解行動誘発文言の発話が繰り返し行われるようにしてもよい。

【 0 2 0 0 】

ステップ S 8 2 において理解行動要求文言の発話が行われると、その後、処理はステップ S 7 8 に戻り、上述した処理が繰り返される。

50

## 【0201】

また、ステップS78において規定時間内に理解行動があったと判定された場合、つまり、規定時間内に理解行動判定部27から音声出力対話進行制御部23に理解行動の判定結果が供給された場合、音声出力対話進行制御部23はレスポンス時間の計測を停止し、処理はステップS83へと進む。

## 【0202】

このようにしてレスポンス時間の計測を停止させることで、処理対象文言の音声発話が指示されてから、理解行動の判定結果が供給されるまでの時間であるレスポンス時間が得られる。このレスポンス時間は、音声発話が行われてから、ユーザがその音声発話に対する理解行動を行うまでの応答時間を示しているということができる。

10

## 【0203】

また、規定時間内に理解行動があったと判定された場合、つまり規定時間内にユーザの理解行動が検出された場合、音声出力対話進行制御部23は、そのユーザの理解行動の検出結果と発話テキストに基づいて、次の発話を制御する。

## 【0204】

すなわち、ステップS83において、音声出力対話進行制御部23は、理解行動判定部27から供給された理解行動の判定結果が肯定的な理解行動を示すものであるか否かを判定する。

## 【0205】

ステップS83において、肯定的な理解行動を示すものでない、つまり否定的な理解行動を示すものであると判定された場合、ユーザは発話内容を理解しておらず理解度が低いので、ステップS84において、音声出力対話進行制御部23は保持している理解・集中度の値を下げる。

20

## 【0206】

ステップS84では、例えばステップS79の処理で理解・集中度が下げられるときと同じだけ、またはステップS79の処理で理解・集中度が下げられるときよりも小さい値だけ、理解・集中度の値が下がるように理解・集中度が更新される。なお、ステップS84においても、ステップS74の処理で理解・集中度が下げられるよりも、より大きく理解・集中度の値が下がるように、理解・集中度が更新される。

## 【0207】

ステップS84で理解・集中度が更新されると、その後、処理はステップS76へと戻り、上述した処理が繰り返し行われる。

30

## 【0208】

この場合、音声発話時の発話速度が再度設定されて、より遅い発話速度で直前の発話内容が繰り返し発話されることになる。すなわち、ユーザが否定的な理解行動を行った場合、音声出力対話進行制御部23は、直前に行われた発話を再度行わせる。

## 【0209】

ステップS84の処理が行われたときには、ユーザの理解度および集中度が低下しているので、より遅い発話速度で直前の発話を繰り返すことで、ユーザが発話内容を理解しやすくすることができる。しかも、この場合、ユーザが発話内容を理解するまで対話が次に進まないため、ユーザが途中で理解を諦めたり、発話内容を理解できなくなったりしてしまうようなこともない。

40

## 【0210】

このようにユーザの集中度および理解度が低い場合には、理解・集中度の値が下がるように更新が行われるので、理解行動の誘発回数が増加していくとともに音声発話の発話速度やトーンも徐々に下がっていくことになる。

## 【0211】

これに対して、ステップS83において肯定的な理解行動を示すものであると判定された場合、ステップS85において、音声出力対話進行制御部23は、得られたレスポンス時間に基づいて理解・集中度を更新する。

50

## 【 0 2 1 2 】

具体的には、例えば音声出力対話進行制御部 2 3 は、レスポンス時間が短いほど理解・集中度の値が大きくなるように、理解・集中度の更新を行う。

## 【 0 2 1 3 】

これにより、ユーザによる肯定的な理解行動のレスポンス時間が短く、ユーザの集中度や理解度が高いときには累積的に理解・集中度が上がるため、音声発話の文章が進むにつれて次第に発話速度が速くなり、理解行動の誘発回数も減少することになる。また、理解・集中度に応じて発話のトーンも変更する場合には、音声発話の文章が進むにつれて次第に発話のトーンが高くなっていくことになる。

## 【 0 2 1 4 】

このように、レスポンス時間が短いほど理解行動の誘発回数がより少なくなるとともに、発話速度が速くなり、発話のトーンが高くなるように発話を制御することで、ユーザの集中度および理解度に合わせた、より適切な対話制御を実現することができる。

## 【 0 2 1 5 】

ステップ S 8 5 において理解・集中度が更新されると対話進行制御処理は終了する。そして、その後、適宜、それ以降の文節の文言の発話が行われ、次の理解行動要求位置の直前の文節が新たな処理対象文言とされて、次の対話進行制御処理が行われる。

## 【 0 2 1 6 】

すなわち、ユーザが肯定的な理解行動を行ったときには、音声出力対話進行制御部 2 3 は、発話テキストのまだ発話がされていない次の文言を発話させることで、発話文言が次の文言へと進むことになる。

## 【 0 2 1 7 】

以上のようにして音声出力対話進行制御部 2 3 は、理解・集中度およびユーザの理解行動に応じて、適宜、発話速度を設定したり、理解行動を誘発したり、理解行動を要求したりしながら発話テキストに示される発話内容での音声対話を進めていく。このようにすることで、ユーザの集中度および理解度に応じて、より適切な対話制御を行うことができる。

## 【 0 2 1 8 】

なお、ここでは理解行動要求位置において、適宜、ユーザの理解行動を誘発し、ユーザの理解行動に応じて対話を進める例について説明した。

## 【 0 2 1 9 】

しかし、これに限らず、例えば理解行動要求位置以外のタイミング（時間）においても音声出力対話進行制御部 2 3 がユーザによる理解行動を常時監視しておき、その理解行動の判定結果に応じて理解・集中度の値を更新するようにしてもよい。

## 【 0 2 2 0 】

そのような場合、肯定的な理解行動があったときには、理解・集中度の値を上げるように更新が行われ、その後において理解行動の誘発が必要以上に多く行われることが抑制される。逆に、否定的な理解行動があったときには、理解・集中度の値を下げるように更新が行われ、理解行動の誘発の機会が増やされてユーザが発話内容を理解しやすくされる。

## 【 0 2 2 1 】

また、以上のような対話進行制御処理により、例えば図 6 に示すような対話が行われることになる。この例は、音声対話システムにより買い物リストを提示する音声発話が行われる対話シーケンス例となっている。

## 【 0 2 2 2 】

対話シーケンスの開始時には、理解・集中度がある程度低い状態から対話制御が開始される。矢印 Q 1 1 に示すように、まず音声対話システムにより処理対象文言に理解行動誘発文言「ね」が付加され、語尾の抑揚が上げられた文言「今日の買い物はニンジンとね」が、低速の発話速度で発話される。ここでは、理解行動誘発文言「ね」を付加することで、発話文言は続きがあることを示す言い方となっている。なお、図中、上向きの矢印は抑揚が上げられていることを表している。

## 【 0 2 2 3 】

10

20

30

40

50

文言「今日の買い物はニンジンとね」が発話されると理解行動の待ち状態となり、最初は長めの間が設けられ、相槌（理解行動）が促される。

【0224】

これに対して、ユーザが矢印Q12に示すように発話「うん」を行うと、音声対話システムは、その発話「うん」に基づいて肯定的な相槌を検出して、ユーザが肯定的な理解行動を行ったと判定し、次の音声発話を開始する。

【0225】

すなわち、音声対話システムは、矢印Q13に示すように処理対象文言に理解行動誘発文言が付加された文言「玉ねぎ3個とね」を発話する。このとき、語尾の抑揚が上げられて矢印Q11における場合と同様に低速の発話速度での発話が行われる。

10

【0226】

この発話に対してユーザが規定時間以上、理解行動を行わなかったため理解・集中度が下げられ、音声対話システムは、矢印Q14に示すように理解行動要求文言「いいですか？」を低速の発話速度で発話する。

【0227】

ここでは、文言「玉ねぎ3個とね」の発話後、ユーザの理解行動がなかったことから、ユーザの対話への意識、つまり集中度が下がったものとされて理解・集中度が下げられ、理解行動要求文言「いいですか？」の発話が行われている。

【0228】

すると、ユーザにより矢印Q15に示すように否定的な相槌の発話「え？」が行われたので、音声対話システムでは、この発話、つまり理解行動の判定が行われ、否定的な理解行動であるとの判定結果が得られる。

20

【0229】

このように否定的な理解行動がなされるとさらに理解・集中度が下げられ、音声対話システムは、矢印Q16に示すように直前の処理対象文言とそれに付加された理解行動誘発文言である文言「玉ねぎ3個とね」を再度、発話する。このとき、語尾の抑揚が上げられるように発話が行われる。すなわち、矢印Q13に示した発話と同様の発話が繰り返される。このような状態では、ユーザの理解・集中度が低いので低速の発話速度で発話が行われる。

【0230】

矢印Q16に示す文言「玉ねぎ3個とね」の発話後、矢印Q17に示すように規定時間内にユーザが理解行動としての発話「うん」を行うと、音声対話システムでは、この理解行動は肯定的な理解行動であるとの判定がなされる。

30

【0231】

そして、矢印Q18に示すように、音声対話システムによって、次の処理対象文言に理解行動誘発文言が付加された文言「じゃがいもとね」が低速の発話速度で、語尾の抑揚が上げられて発話される。

【0232】

これに対して、矢印Q19に示すように規定時間内にユーザが理解行動としての発話「うん」を行うと、音声対話システムは、矢印Q20に示すように次の処理対象文言に理解行動誘発文言が付加された文言「牛肉300gとね」を低速の発話速度で、語尾の抑揚を上げて発話する。

40

【0233】

次に、これに対してユーザにより矢印Q21に示すようにこれまでよりも短いレスポンス時間で、肯定的な理解行動を示す相槌の文言「うん」が発話された。

【0234】

すると、音声対話システムでは、ユーザの理解・集中度が上げられて、矢印Q22に示すように、次の処理対象文言に理解行動誘発文言が付加された文言「塩とね」が中速の発話速度で、語尾の抑揚が上げられて発話される。つまり、ユーザの理解・集中度が高まったことに伴い、これまでよりも速い発話速度で発話が行われる。

50

## 【 0 2 3 5 】

そして、これに対してユーザにより矢印Q 2 3 に示すように、矢印Q 2 1 に示した場合よりもさらに短いレスポンス時間で、肯定的な理解行動を示す相槌の文言「うん」が発話された。

## 【 0 2 3 6 】

その結果、音声対話システムでは、ユーザの理解・集中度がさらに上げられて、矢印Q 2 4 に示すように、次の処理対象文言に理解行動誘発文言が付加された文言「胡椒とね」が中速の発話速度で、語尾の抑揚が上げられて発話される。

## 【 0 2 3 7 】

さらに、これに対してユーザにより矢印Q 2 5 に示すように、矢印Q 2 3 に示した場合よりもさらに短いレスポンス時間で、肯定的な理解行動を示す相槌の文言「うん」が発話された。そうすると、音声対話システムでは、ユーザの理解・集中度がさらに上げられて、矢印Q 2 6 に示すように、次の処理対象文言「醤油とビールです。」が高速の発話速度で発話され、対話シーケンスが終了する。

10

## 【 0 2 3 8 】

この場合、矢印Q 2 5 に示した理解行動によって理解・集中度が十分高い値となったので、文言「醤油と」と、その次の文言「ビールです。」との間には理解行動誘発文言が付加されず、つまり理解行動の誘発が行われずに発話が行われている。また、ここでは理解・集中度が十分高い値となったので、発話速度も最も速い発話速度となっている。

## 【 0 2 3 9 】

このように、図 6 に示した例では、矢印Q 1 6 に示した発話以降において、理解行動の誘発に対するユーザのレスポンス時間が短くなっていき、それに伴ってユーザの理解・集中度が徐々に大きくなっている様子が示されている。

20

## 【 0 2 4 0 】

この例では、理解・集中度が大きくなるのに伴って音声対話システムによる音声発話の発話速度も速くなっていくとともに、理解行動の誘発回数も減少している。このようにユーザの集中度および理解度に応じて理解行動を誘発したり、発話速度を変化させたりすることで、適切な対話制御を実現することができる。

## 【 0 2 4 1 】

以上のように、本技術を適用した音声対話システムによれば、より適切な対話制御を行うことができる。

30

## 【 0 2 4 2 】

特にテキスト音声合成が苦手とする長い文章や羅列情報を含む文章の音声発話において、ユーザの集中度および理解度に応じた発話速度で音声発話が行われるため、音声対話システムがユーザに対して確実に伝達したい発話内容について、ユーザの理解や記憶のしやすさを向上させることができる。

## 【 0 2 4 3 】

また、一般的なテキスト音声合成の音声発話性能により、すなわち抑揚と発話速度が一定であることにより理解しにくい文章の内容を、本技術を適用した音声対話システムでは理解行動を利用したインタラクションにより補うことができる。すなわち、理解行動の判定結果やレスポンス時間に応じて発話速度を変化させたり、抑揚を変化させたりすることで、ユーザが発話内容を理解しやすいように、また発話内容を記憶しやすいように対話を行うことができる。

40

## 【 0 2 4 4 】

さらに、ユーザの集中度および理解度が高い場合、つまりユーザが集中して対話音声を聞いている場合には、発話速度が上げられるだけでなく、理解行動誘発回数が減らされてユーザの理解行動への負荷が下がるので、ユーザはより短い時間かつ少ない負荷で対話を行うことができる。

## 【 0 2 4 5 】

理解行動要求位置の他の検出例 1

50

なお、以上においては、例えば図 2 を参照して説明したように、発話テキストの文章内の述語節の文節に係る文節位置に基づいて、理解行動要求位置を検出する例について説明した。しかし、理解行動を誘発したいアイテム等の文言が含まれる文節位置を検出することができれば、どのような方法により理解行動要求位置を検出するようにしてもよい。

【 0 2 4 6 】

例えば理解行動要求位置の他の検出方法の例として、いつ (When)、どこで (Where)、誰が (Who)、何を (What)、なぜ (Why)、どのように (How) という 6 つの要素、すなわち文節の格としていわゆる 5W1H を示す時間格、場所格、対象格、原因格が検出されたときに、それらの格の文節の直後の位置を理解行動要求位置と判定してもよい。

【 0 2 4 7 】

このような場合においても、文として非完結な状態の位置が理解行動要求位置とされることに加えて、5W1H の各情報粒度でユーザの理解を確認することができる。この例は、5W1H の各情報粒度で理解確認を行いたいときに特に有効である。

【 0 2 4 8 】

具体的には、例えば発話テキストとして、「今日の予定として、打ち合わせのため 10 時に品川駅で山田さんと待ち合わせがあります。」が入力されたとする。

【 0 2 4 9 】

そのような場合、理解行動要求位置検出部 2 2 では、原因格である文節「打ち合わせのため」の直後と、時間格である文節「10 時に」の直後と、場所格である文節「品川駅で」の直後と、対象格である文節「山田さんと」の直後のそれぞれの位置が、理解行動要求位置として検出されることになる。

【 0 2 5 0 】

以上のように発話テキストの文章内の時間格や、場所格、対象格、原因格の文節の位置に基づいて、理解行動要求位置を検出するようにすることもできる。

【 0 2 5 1 】

また、理解行動要求位置の他の検出方法の例として、予め単語の重要度が登録された単語辞書を用いるようにしてもよい。そのような場合、単語と、単語の重要度とが対応付けられた単語辞書が予め理解行動要求位置検出部 2 2 に記録されている。

【 0 2 5 2 】

理解行動要求位置検出部 2 2 は、単語辞書を参照して、発話テキストの文章に含まれる各単語の重要度を特定し、重要度が高い単語を含む文節の直後の位置を、理解行動要求位置として検出する。このとき、例えば重要度が所定の閾値以上である単語を含む文節の直後が全て理解行動要求位置とされるようにしてもよいし、発話テキストの文章に含まれる単語の中から、重要度が高い順に所定個数の単語を選択し、選択したそれらの単語が含まれる文節の直後が理解行動要求位置とされるようにしてもよい。

【 0 2 5 3 】

理解行動要求位置の他の検出例 2

さらに、以上においては発話テキストが日本語の文章である場合について説明したが、これに限らず、発話テキストが英語など、日本語以外の他の言語である場合においても本技術は適用可能である。すなわち、以上において説明した例と同様に理解行動要求位置を検出することができる。

【 0 2 5 4 】

例えば発話テキストが英語の文章である場合について考える。

【 0 2 5 5 】

英語では、文頭に主語 + 述語 (動詞) が現れるが、その述語 (動詞) の後から述語に係る目的語 (対象格) が複数あった場合にそれらの目的語が羅列情報であるとされる。そして、日本語と同様に非完結の状態である羅列情報の間の位置、つまり「,」(カンマ) の直後の位置や「and」の直後の位置が理解行動要求位置とされ、その理解行動要求位置で、適宜、相槌や頷きなどの理解行動が誘発される。

【 0 2 5 6 】

10

20

30

40

50

また、2つ目以降の文では主語 + 述語（動詞）の直後の位置が理解行動要求位置とされる。この場合においても文章が非完結な状態で、適宜、相槌や頷きなどの理解行動が誘発されることになる。

【0257】

具体例として、例えば図7の矢印A41に示すように英語の文章「Please buy carrots, onions, potatoes, beef, and salt. Please don't forget to post the letter.」というテキストデータが自然言語解析部21に入力されたとする。

【0258】

このような場合に、自然言語解析部21において発話テキストに対する自然言語解析が行われると、矢印A42に示す解析結果が得られる。

【0259】

矢印A42に示す部分では、図2における場合と同様に、各四角形が文の区切られた1つの区間、つまり句を表しており、その四角形内には発話テキストの句に区切られた文言が記されている。また、各句を表す四角形同士を結ぶ線分は、句同士の係り受けを表している。これらの係り受けの表記方法は図2における場合と同様である。

【0260】

さらに、矢印A42に示す部分では、各句を表す四角形の図中、右側には、それらの句の言語的な格と品詞が記されている。例えば図中、一番上側の句「Please buy」の言語的な格は述語節で品詞は動詞であることが記されている。

【0261】

このような場合、理解行動要求位置検出部22は、自然言語解析の結果から同一の述語節（動詞）に係る複数の対象格（目的格）を検出し、検出された複数の対象格の句の間の位置を理解行動要求位置とする。

【0262】

但し、最も後ろに位置する対象格の句については、その句に含まれる「and」の直後の位置、すなわち単語「and」と、その直後の対象語との間の位置が理解行動要求位置とされる。

【0263】

この例では、発話テキストの1文目は、ユーザが買ってくるべきもの（アイテム）、つまり買い物リストを示す文章となっており、買うべき物（アイテム）を含む対象格の句が同一の述語節の句に係っている。すなわち、買うべきものが羅列されている文章となっている。

【0264】

図7に示す例では、発話テキストの1文目の矢印W41乃至矢印W44に示される位置が理解行動要求位置検出部22により検出された理解行動要求位置となっている。具体的には、例えば矢印W41により示される理解行動要求位置は、句「carrots,」と「onions,」の間の位置となっている。また、例えば矢印W44により示される理解行動要求位置は、句「and salt.」における単語「and」と、対象語「salt」との間の位置となっている。

【0265】

また、理解行動要求位置検出部22は、発話テキストに複数の文が含まれる場合は、文と文の区切りの位置を理解行動要求位置とは判定しない。

【0266】

さらに、理解行動要求位置検出部22は、発話テキストに複数の文が含まれる場合、2つ目以降の文において、文の最初の述語節を検出し、その最初の述語節の句の直後の位置を理解行動要求位置とする。

【0267】

この例では、2つ目の文における矢印W45に示される位置が理解行動要求位置とされている。ここでは、矢印W45に示される位置は、述語節の句に係る最初の句の直前の位置となっている。

【0268】

10

20

30

40

50

音声出力対話進行制御部 23 は、以上のようにして検出された理解行動要求位置に基づいて、例えば図 5 を参照して説明した対話進行制御処理と同様の処理を行うことで、対話制御を行う。

【0269】

このとき、音声出力対話進行制御部 23 は、例えば理解行動の誘発を行う理解行動要求位置の直前の句が対象格（目的語）であり、その句の次の句に単語「and」が含まれていないときには、理解行動要求位置の直前の句の最後の単語（目的語）の次に単語「and」を付加して音声発話を行わせる。また、この場合、付加した単語「and」の直前に「,」（カンマ）があるときには、その「,」が削除される。このように発話文言の句に付加される単語「and」は、理解行動誘発文言として機能する。

10

【0270】

さらに、音声発話時には、理解行動の誘発を行う理解行動要求位置においては、ある程度の間が設けられて、つまりある程度長い時間があけられて音声の発話が行われる。

【0271】

したがって、矢印 W 4 1 乃至矢印 W 4 5 のそれぞれの理解行動要求位置において理解行動の誘発が行われる場合には、例えば矢印 A 4 3 に示すように音声発話が行われる。

【0272】

すなわち、まず最初の文言「Please buy carrots and」が発話されて理解行動の誘発が行われ、続いて文言「onions and」が発話されて理解行動の誘発が行われる。ここで、理解行動の誘発は例えば発話の間をあけたり、単語「and」を付加したり、発話の語尾に抑揚を付加したりすることにより行われる。

20

【0273】

さらに文言「potatoes and」が発話されて理解行動の誘発が行われてから、文言「beef and」が発話されて理解行動の誘発が行われ、文言「salt. Please don't forget」が発話されて理解行動の誘発が行われる。そして、最後に文言「to post the letter.」が発話されて音声対話が終了する。

【0274】

以上のように発話テキストが日本語以外の言語の文章であるときでも、発話テキストの文章内の述語節に係る句（文節）の位置に基づいて、理解行動要求位置を検出することが可能である。

30

【0275】

理解行動要求位置の他の検出例 3

さらに、発話テキストが英語の文章である場合、いわゆる 5W1H で前置詞により場所や時間、原因を表すときには、その前置詞の直後の位置が理解行動要求位置とされて、その理解行動要求位置で、適宜、相槌や頷きなどの理解行動が誘発されるようにしてもよい。この場合においても文章が非完結な状態で、適宜、相槌や頷きなどの理解行動が誘発されることになる。

【0276】

具体的には、例えば図 8 の矢印 A 7 1 に示すように英語の文章「You have an appointment for the meeting at 3pm at Shinagawa office.」というテキストデータが自然言語解析部 21 に入力されたとする。

40

【0277】

このような場合に、自然言語解析部 21 において発話テキストに対する自然言語解析が行われると、矢印 A 7 2 に示す解析結果が得られる。

【0278】

矢印 A 7 2 に示す部分では、図 2 における場合と同様に、各四角形が文の区切られた 1 つの区間、つまり句を表しており、その四角形内には発話テキストの句に区切られた文言が記されている。また、各句を表す四角形同士を結ぶ線分は、句同士の係り受けを表している。これらの係り受けの表記方法は図 2 における場合と同様である。

【0279】

50

さらに、矢印 A 7 2 に示す部分では、各句を表す四角形の図中、右側には、それらの句の言語的な格が記されている。また、必要に応じて格とともに品詞も記されている。例えば図中、一番上側の句「You」の言語的な格は主格であることが記されている。

【0280】

このような場合、理解行動要求位置検出部 2 2 は、自然言語解析の結果から原因格、時間格、および場所格の句を検出し、検出された句内（句中）の前置詞の直後の位置を理解行動要求位置とする。

【0281】

したがって、図 8 に示す例では、発話テキストの原因格の句の中の前置詞「for」の直後の位置、すなわち矢印 W 7 1 により示される位置と、時間格の句の中の前置詞「at」の直後の位置、すなわち矢印 W 7 2 により示される位置と、場所格の句の中の前置詞「at」の直後の位置、すなわち矢印 W 7 3 により示される位置とが理解行動要求位置検出部 2 2 により検出される。そして、それらの矢印 W 7 1 乃至矢印 W 7 3 のそれぞれにより示される位置が理解行動要求位置とされている。

10

【0282】

音声出力対話進行制御部 2 3 は、以上のようにして検出された理解行動要求位置に基づいて、例えば図 5 を参照して説明した対話進行制御処理と同様の処理を行うことで、対話制御を行う。このとき、理解行動の誘発を行う理解行動要求位置においては、ある程度の間が設けられて音声の発話が行われる。

【0283】

したがって、矢印 W 7 1 乃至矢印 W 7 3 のそれぞれの理解行動要求位置において理解行動の誘発が行われる場合には、例えば矢印 A 7 3 に示すように音声発話が行われる。

20

【0284】

すなわち、まず最初の文言「You have an appointment for」が発話されて理解行動の誘発が行われ、続いて文言「the meeting at」が発話されて理解行動の誘発が行われる。ここで、理解行動の誘発は、例えば発話の間をあけたり、発話の語尾に抑揚を付加したりすることにより行われる。

【0285】

さらに文言「3pm at」が発話されて理解行動の誘発が行われてから、最後に文言「Shinagawa office.」が発話されて音声対話が終了する。

30

【0286】

以上のように発話テキストが日本語以外の言語の文章であるときでも、発話テキストの文章内の時間格や、場所格、対象格、原因格の句（文節）の位置に基づいて、理解行動要求位置を検出することが可能である。

【0287】

コンピュータの構成例

ところで、上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェアに組み込まれているコンピュータや、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどが含まれる。

40

【0288】

図 9 は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

【0289】

コンピュータにおいて、CPU（Central Processing Unit）5 0 1，ROM（Read Only Memory）5 0 2，RAM（Random Access Memory）5 0 3 は、バス 5 0 4 により相互に接続されている。

【0290】

50

バス504には、さらに、入出力インターフェース505が接続されている。入出力インターフェース505には、入力部506、出力部507、記録部508、通信部509、及びドライブ510が接続されている。

【0291】

入力部506は、キーボード、マウス、マイクロホン、撮像素子などよりなる。出力部507は、ディスプレイ、スピーカなどよりなる。記録部508は、ハードディスクや不揮発性のメモリなどよりなる。通信部509は、ネットワークインターフェースなどよりなる。ドライブ510は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブル記録媒体511を駆動する。

【0292】

以上のように構成されるコンピュータでは、CPU501が、例えば、記録部508に記録されているプログラムを、入出力インターフェース505及びバス504を介して、RAM503にロードして実行することにより、上述した一連の処理が行われる。

【0293】

コンピュータ(CPU501)が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブル記録媒体511に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

【0294】

コンピュータでは、プログラムは、リムーバブル記録媒体511をドライブ510に装着することにより、入出力インターフェース505を介して、記録部508にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部509で受信し、記録部508にインストールすることができる。その他、プログラムは、ROM502や記録部508に、あらかじめインストールしておくことができる。

【0295】

なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

【0296】

また、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

【0297】

例えば、本技術は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

【0298】

また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

【0299】

さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

【0300】

さらに、本技術は、以下の構成とすることも可能である。

【0301】

(1)

1または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話を行わせ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話を制御する対話進行制御部を備える対話制御装置。

(2)

前記対話進行制御部は、前記ユーザによる前記理解行動が肯定的な行動であった場合、前

10

20

30

40

50

記発話テキストのまだ発話がされていない次の文言を発話させる

( 1 ) に記載の対話制御装置。

( 3 )

前記対話進行制御部は、前記ユーザによる前記理解行動が否定的な行動であった場合、直前に行われた発話を再度行わせる

( 1 ) または ( 2 ) に記載の対話制御装置。

( 4 )

前記対話進行制御部は、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど前記理解行動の誘発回数が少なくなるように、前記発話テキストに基づく発話を制御する

( 1 ) 乃至 ( 3 ) の何れか一項に記載の対話制御装置。

( 5 )

前記対話進行制御部は、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど発話速度が速くなるように、前記発話テキストに基づく発話を制御する

( 1 ) 乃至 ( 4 ) の何れか一項に記載の対話制御装置。

( 6 )

前記対話進行制御部は、前記ユーザによる肯定的な前記理解行動のレスポンス時間が短いほど発話のトーンが高くなるように、前記発話テキストに基づく発話を制御する

( 1 ) 乃至 ( 5 ) の何れか一項に記載の対話制御装置。

( 7 )

前記対話進行制御部は、所定時間内に前記ユーザによる前記理解行動が検出されなかった場合、直前に行われた発話を再度行わせるか、または前記理解行動を要求する発話を行わせる

( 1 ) 乃至 ( 6 ) の何れか一項に記載の対話制御装置。

( 8 )

前記対話進行制御部は、前記発話テキストに基づく発話を途中で停止させた場合、前記発話テキストを含む通知情報の出力を制御する

( 1 ) 乃至 ( 7 ) の何れか一項に記載の対話制御装置。

( 9 )

前記対話進行制御部は、前記発話テキストに基づく文言に非完了文言を付加することで、前記理解行動を誘発する発話を行わせる

( 1 ) 乃至 ( 8 ) の何れか一項に記載の対話制御装置。

( 10 )

前記対話進行制御部は、発話の語尾に抑揚を付加することで、前記理解行動を誘発する発話を行わせる

( 9 ) に記載の対話制御装置。

( 11 )

前記対話進行制御部は、前記理解行動要求位置において間をあけることで前記理解行動を誘発する発話を行わせる

( 1 ) 乃至 ( 10 ) の何れか一項に記載の対話制御装置。

( 12 )

前記発話テキストの文章における文として非完結な位置を前記理解行動要求位置として検出する理解行動要求位置検出部をさらに備える

( 1 ) 乃至 ( 11 ) の何れか一項に記載の対話制御装置。

( 13 )

前記理解行動要求位置検出部は、前記発話テキストの文章の述語節との係り受けに基づく位置を前記理解行動要求位置として検出する

( 12 ) に記載の対話制御装置。

( 14 )

前記理解行動要求位置検出部は、前記発話テキストの文章における同一の述語節に係る複

10

20

30

40

50

数の対象格の文節または句のそれぞれの間の位置を前記理解行動要求位置として検出する  
( 1 3 ) に記載の対話制御装置。

( 1 5 )

前記理解行動要求位置検出部は、前記発話テキストの文章における述語節に最初に係る文節または句の位置を前記理解行動要求位置として検出する

( 1 3 ) に記載の対話制御装置。

( 1 6 )

前記理解行動要求位置検出部は、前記発話テキストの文章における時間格、場所格、対象格、または原因格である文節または句の位置を前記理解行動要求位置として検出する

( 1 2 ) に記載の対話制御装置。

( 1 7 )

1 または複数の理解行動要求位置で区切られた発話テキストに基づいて、前記理解行動要求位置においてユーザの理解行動を誘発する発話を行わせ、前記理解行動の検出結果と前記発話テキストとに基づいて、次の発話を制御するステップを含む対話制御方法。

【符号の説明】

【 0 3 0 2 】

1 1 サーバ， 1 2 クライアントデバイス， 2 1 自然言語解析部， 2 2 理解行動要求位置検出部， 2 3 音声出力対話進行制御部， 2 4 音声合成エンジン， 2 7 理解行動判定部

10

20

30

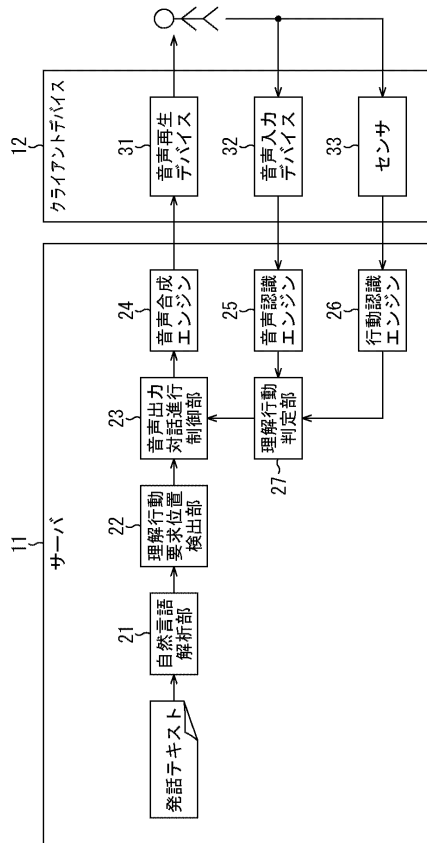
40

50

【図面】

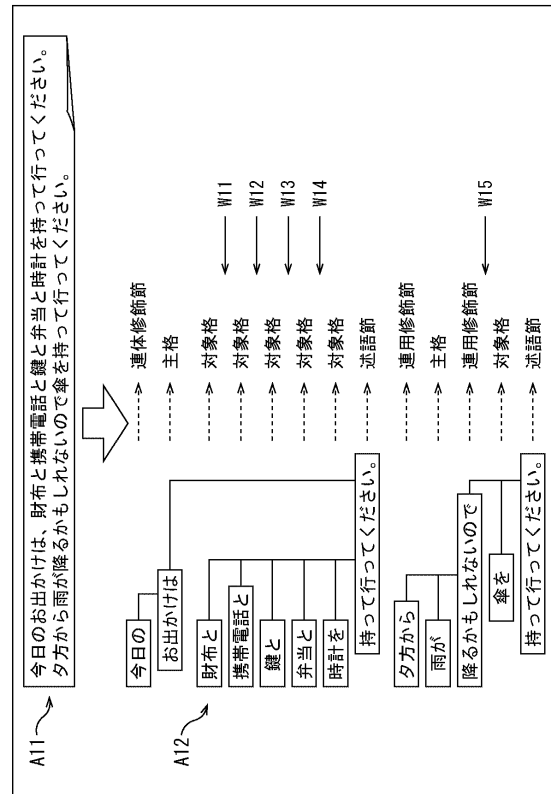
【図 1】

FIG. 1



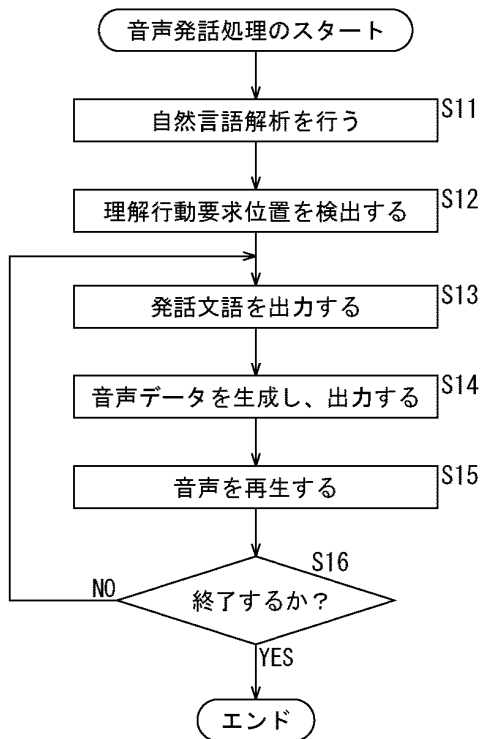
【図 2】

FIG. 2



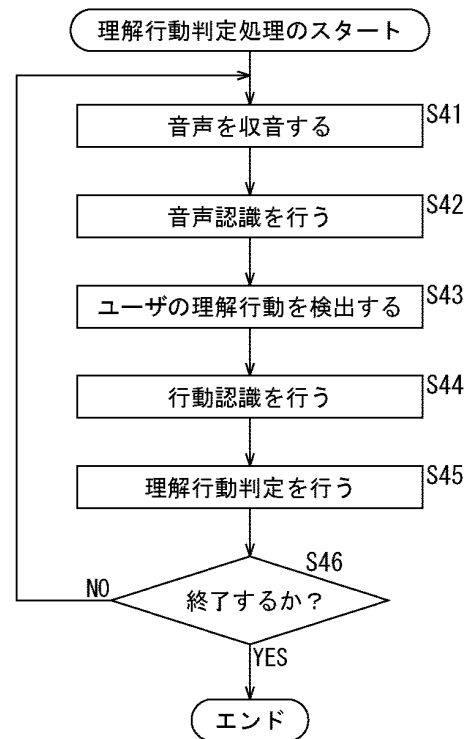
【図 3】

FIG. 3



【図 4】

FIG. 4



10

20

30

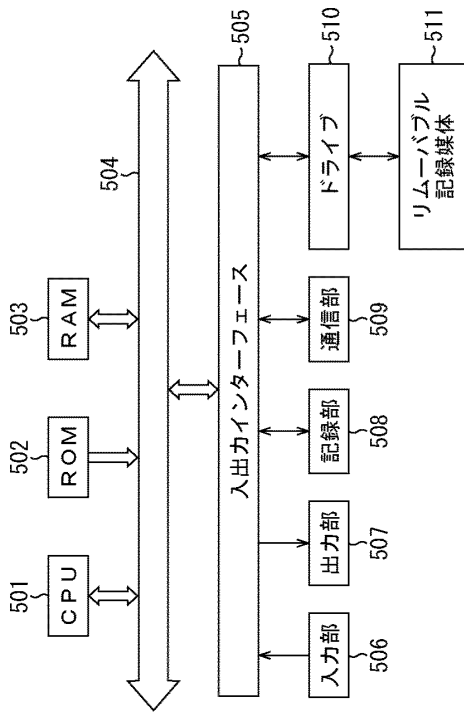
40

50



【図 9】

FIG. 9



10

20

30

40

50

---

フロントページの続き

(51)国際特許分類

F I

G 1 0 L 15/22 3 0 0 Z  
G 1 0 L 13/10 1 1 4

株式会社内

審査官 富澤 直樹

(56)参考文献 国際公開第2008/001549(WO, A1)  
特開2010-008854(JP, A)  
特開昭60-247697(JP, A)  
特開平10-083196(JP, A)  
特開平06-342297(JP, A)  
特開2005-202076(JP, A)  
特開2010-157081(JP, A)  
特開平10-116177(JP, A)  
米国特許出願公開第2007/0255566(US, A1)

(58)調査した分野 (Int.Cl., DB名)  
G 1 0 L 13/00 - 13/10  
G 1 0 L 15/00 - 15/34