

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4680714号
(P4680714)

(45) 発行日 平成23年5月11日(2011.5.11)

(24) 登録日 平成23年2月10日(2011.2.10)

(51) Int.Cl.

F 1

G 1 O L	15/08	(2006.01)	G 1 O L	15/08	300B
G 1 O L	15/22	(2006.01)	G 1 O L	15/22	300U
G 1 O L	15/06	(2006.01)	G 1 O L	15/06	300J

請求項の数 10 (全 25 頁)

(21) 出願番号	特願2005-225877 (P2005-225877)
(22) 出願日	平成17年8月3日(2005.8.3)
(65) 公開番号	特開2007-41319 (P2007-41319A)
(43) 公開日	平成19年2月15日(2007.2.15)
審査請求日	平成20年5月27日(2008.5.27)

(73) 特許権者	000005821 パナソニック株式会社 大阪府門真市大字門真1006番地
(74) 代理人	100109210 弁理士 新居 広守
(72) 発明者	井上 剛 大阪府門真市大字門真1006番地 松下電器産業株式会社内
(72) 発明者	沖本 純幸 大阪府門真市大字門真1006番地 松下電器産業株式会社内
(72) 発明者	九津見 洋 大阪府門真市大字門真1006番地 松下電器産業株式会社内

最終頁に続く

(54) 【発明の名称】音声認識装置および音声認識方法

(57) 【特許請求の範囲】

【請求項 1】

入力された音声を認識し、認識結果に応じて、ユーザとの対話に関するシステムの状態であるシステム状態を遷移させて、対話をを行う音声認識装置であって、

音声認識辞書を用いて、入力された音声を認識して、認識結果を出力する音声認識手段と、

前記音声認識手段の認識結果により前記システム状態を遷移させて応答を行う対話制御手段と、

今回の認識結果で前記システム状態が先に進まず停滞している状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定手段と、

前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定としてリジェクトの閾値の変更と、音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更との少なくとも1つを行なう変更制御手段と

を備えることを特徴とする音声認識装置。

【請求項 2】

入力された音声を認識し、認識結果に応じて、ユーザとの対話に関するシステムの状態であるシステム状態を遷移させて、対話をを行う音声認識装置であって、

音声認識辞書を用いて、入力された音声を認識して、認識結果を出力する音声認識手段

と、

前記音声認識手段の認識結果により前記システム状態を遷移させて応答を行う対話制御手段と、

今回の認識結果により、前記システム状態が、前回の認識結果によるシステム状態と同一の状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定手段と、

前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定としてリジェクトの閾値の変更と、音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更との少なくとも1つを行う変更制御手段と

を備えることを特徴とする音声認識装置。

【請求項3】

前記システム状態の停滞状態は、前記音声認識結果のリジェクトによる同一システム状態が続く状態であり、

前記停滞脱出判定手段は、今回の認識結果が前回の認識結果と同一単語である場合、言い直しであると判定し、今回の認識結果が前回の認識結果と同一単語では無いが、あらかじめ定められた同じシステム動作を実行する認識単語である場合、言い換えであると判定する

ことを特徴とする請求項1または2に記載の音声認識装置。

【請求項4】

前記システム状態の停滞状態は、2つのシステム状態の往復が繰り返し続く状態であり、

前記停滞脱出判定手段は、今回の認識結果が前々回の認識結果と同一単語である場合、言い直しであると判定し、今回の認識結果が前々回の認識結果と同一単語では無いが、あらかじめ定められた同じシステム動作を実行する認識単語である場合、言い換えであると判定する

ことを特徴とする請求項1または2に記載の音声認識装置。

【請求項5】

前記変更制御手段は、前記リジェクトの閾値を認識対象単語ごとに設定し変更する

ことを特徴とする請求項1または2に記載の音声認識装置。

【請求項6】

前記変更制御手段は、前記リジェクトの閾値、および、前記音声認識辞書への新規追加または変更を、ユーザごとに設定する

ことを特徴とする請求項1または2に記載の音声認識装置。

【請求項7】

前記音声認識装置は、さらに、

前記停滞状態から脱出した際に、今回の認識結果が前回の認識結果と同一単語では無く、かつあらかじめ定められた同じシステム動作を実行する認識単語でない場合、今回の認識対象語彙の省略語を作成する省略語作成手段を備え、

前記音声認識手段は、前記省略語を用いて前回の認識結果を再認識し、

前記変更制御手段は、前記音声認識手段の再認識結果に応じて前記省略語を前記音声認識辞書へ新規追加する

ことを特徴とする請求項1または2に記載の音声認識装置。

【請求項8】

入力された電子番組表に関する音声を認識し、認識結果に応じて、ユーザとの対話に関するシステムの状態であるシステム状態を遷移させて、対話を行う電子番組表用音声認識装置であって、

電子番組表に対応する音声認識辞書を用いて、入力された電子番組表に関する音声を認識して、認識結果を出力する音声認識手段と、

前記音声認識手段の認識結果により前記システム状態を遷移させて応答を行う対話制御

10

20

30

40

50

手段と、

今回の認識結果で前記システム状態が先に進まず停滞している状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定手段と、

前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定としてリジェクトの閾値の変更と、音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更との少なくとも1つを行う変更制御手段と

を備えることを特徴とする電子番組表用音声認識装置。

【請求項9】

10

入力された音声を認識し、認識結果に応じて、ユーザとの対話に関するシステムの状態であるシステム状態を遷移させて、対話をを行う音声認識方法であって、

音声認識辞書を用いて、入力された音声を認識して、認識結果を出力する音声認識ステップと、

前記音声認識ステップにおける認識結果により前記システム状態を遷移させて応答を行う対話制御ステップと、

今回の認識結果で前記システム状態が先に進まず停滞している状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定ステップと、

20

前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定としてリジェクトの閾値の変更と、音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更との少なくとも1つを行う変更制御ステップと

を含むことを特徴とする音声認識方法。

【請求項10】

入力された音声を認識し、認識結果に応じて、ユーザとの対話に関するシステムの状態であるシステム状態を遷移させて、対話をを行うためのプログラムであって、

音声認識辞書を用いて、入力された音声を認識して、認識結果を出力する音声認識ステップと、

前記音声認識ステップにおける認識結果により前記システム状態を遷移させて応答を行う対話制御ステップと、

30

今回の認識結果で前記システム状態が先に進まず停滞している状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定ステップと、

前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定としてリジェクトの閾値の変更と、音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更との少なくとも1つを行う変更制御ステップとをコンピュータに実行させる

ことを特徴とするプログラム。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された音声を音声認識辞書を用いて認識し、認識結果によりシステム状態を遷移させて対話をを行う音声認識装置に関する。

【背景技術】

【0002】

一般的な音声認識の方法の一つとして、予め認識辞書内に登録された語彙を表す種々の音響パターンとユーザから入力された音声信号とを比較することでスコアを計算し、最も類似したパターンを示す認識辞書内語彙を認識結果の候補とする手法がある。

50

【 0 0 0 3 】

このような一般的な音声認識方法では、多くのユーザに対して高い認識精度を実現するために作成された音響モデルを用いて音声認識を行っているため、ユーザによっては一般的な音響モデルが適応していないために、認識精度が低くなり、誤認識を多く起こしてしまう場合が生じる。

【 0 0 0 4 】

また、従来においては、例えば認識スコアを利用して認識信頼度を計算し、最も類似した認識候補が得られた場合も、認識信頼度に基づいてリジェクトを行い、再入力を促す手順がある。このようなリジェクト機能を設けることで、例えば音声以外の雑音が入力された場合に誤認識が発生してシステムが誤動作するのを防ぐことができる。

10

【 0 0 0 5 】

このような音声認識方法では、得られた認識信頼度が所定の値より低い場合、リジェクトなどの機能により誤認識・誤システム動作を防ぐことができるが、反面、通常の音声入力に対しても認識信頼度が低い場合にはリジェクトを行うので、ユーザによっては、特定の認識可能な語彙に対して誤ってリジェクトされてしまう場合が生じる。

【 0 0 0 6 】

そこで、このような誤認識・誤リジェクトの対策として、音声認識に利用している一般不特定話者向けの音響モデルを現在のユーザの音響モデルに適応するため、ユーザ自身の発声を用いて再学習させる（話者適応・話者学習）ことで認識精度を向上させる方法や、リジェクトが行われた際にユーザの再発声時の認識精度を向上させる方法が提案されている。

20

【 0 0 0 7 】

例えば、話者適応・話者学習の方法としては、少数の音声を用いて音響モデルを学習し、さらに誤認識される単語については話者学習を行う方法（例えば、特許文献1参照）が開示されている。一方、再発声時の認識精度を向上させる方法としては、言い直しだと判定した場合には前回と今回の両認識候補を用いて認識結果を定める手法（例えば、特許文献2参照）や、言い直しの発声に対しては前回の認識結果の上位候補を認識対象語彙とする手法（例えば、特許文献3参照）が開示されている。

【特許文献1】特開2003-162292号公報

30

【特許文献2】特開平11-149294号公報**【特許文献3】特許第3112037号****【発明の開示】****【発明が解決しようとする課題】****【 0 0 0 8 】**

上記のような従来の方法では、少数の学習用発声でユーザの音響モデルを学習可能とする工夫や、再発声時の認識候補や認識対象語彙を変化させることで認識精度を向上させる工夫がされている。

【 0 0 0 9 】

しかしながら、これらの学習による話者適応では、ユーザに適したモデルを学習させる際に少数とはいえ、操作とは直接関係の無い単語を一定量ユーザに発声させるため、ユーザの負担は少なくない。また、再入力時の認識精度向上方法では、再発声時の認識精度は上がるものの、再び前回リジェクトされた発声と同じ発声をユーザが行ったときはやはりリジェクトされてしまい、その度に再発声を行わなくてならない。

40

【 0 0 1 0 】

例えば、特許文献1では、初めに少数ではあるが話者適応用の学習発声をユーザに促し、さらに度々誤認識する単語については誤認識を起こす部分の発声を話者に促し、その入力を基に話者学習を行うが、ユーザに余分な発声を促すためユーザの負担を増やしてしまう。また、特許文献2では、言い直しと検出された場合に前回の認識結果を含め出力する認識候補を調整するが、前回リジェクトされた発声と同じ発声が入力されたときにリジェクトされてしまい、その度に再発声を行わなくてならない。また、特許文献3では、リジ

50

エクトされた次の認識は前回の上位候補のみを認識対象語彙として認識を行うが、特許文献2と同様、前回リジェクトされた発声と同じ発声が入力されたときに正しく認識できない。

【0011】

そこで、本発明はこのような従来の課題を解決するためになされたものであって、ユーザに学習用の特別な発声を要求することなく、ユーザの負担が少なく自然に音声認識の個人適応を行うことができ、かつ誤認識を減らすことができる音声認識装置および音声認識方法を提供することを目的とする。

【課題を解決するための手段】

【0012】

上記目的を達成するため、本発明に係る音声認識装置は、入力された音声を認識し、認識結果に応じて、ユーザとの対話に関するシステムの状態であるシステム状態を遷移させて、対話を行う音声認識装置であって、音声認識辞書を用いて、入力された音声を認識して、認識結果を出力する音声認識手段と、前記音声認識手段の認識結果により前記システム状態を遷移させて応答を行う対話制御手段と、今回の認識結果で前記システム状態が先に進まず停滞している状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定手段と、前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定としてリジェクトの閾値の変更と、音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更との少なくとも1つを行なう変更制御手段とを備えることを特徴とする。

10

【発明の効果】

【0013】

本発明に係る音声認識装置および音声認識方法によれば、ユーザの発声の特徴とシステムの音声認識用パラメータや音声認識辞書の不適合を解消するために、学習用の特別な発声を要求するのではなく、一度の言い直しまたは言い換えにて正しく認識された結果を利用してユーザに適した学習を行うため、ユーザにとって自然で負担の少ない音声認識の個人適応を行うことができる。さらに、音声認識の個人適応を行うので、次からは前回誤認識した発声と同様の発声を行っても正しく認識が可能となるため、誤認識が減ることにより、円滑な音声操作を実現することができる。

20

【発明を実施するための最良の形態】

【0014】

本発明の実施の形態に係る音声認識装置は、入力された音声を認識し、認識結果により対話を行う音声認識装置であって、入力された音声を音声認識辞書を用いて認識して認識結果を出力する音声認識手段と、前記音声認識手段の認識結果によりシステム状態を遷移させて応答を行う対話制御手段と、今回の認識結果で前記システム状態が先に進まず停滞している状態である停滞状態から脱出したか否かを判定するとともに、前記停滞状態から脱出したと判定した場合、今回の認識結果が言い直しおよび言い換えの少なくとも1つであるか否かを判定する停滞脱出判定手段と、前記言い直しまたは言い換えであると判定された場合、対話制御に関する設定および音声認識に関する設定の少なくとも1つを変更する変更制御手段とを備えることを特徴とする。

30

【0015】

これによって、通常の音声操作の中で、ユーザ適応を隨時行っていくため、ユーザ適応のために特別な発声が必要なく、ユーザにとって自然で負担の少ない音声認識の個人適応を行うことができる。さらに、音声認識の個人適応を行うので、次からは前回誤認識した発声と同様の発声を行っても正しく認識が可能となるため、誤認識が減ることにより、円滑な音声操作を実現することができる。

40

【0016】

また、前記システム状態の停滞状態は、前記音声認識結果のリジェクトによる同一システム状態が続く状態であり、前記停滞脱出判定手段は、今回の認識結果が前回の認識結果

50

と同一単語である場合、言い直しであると判定し、今回の認識結果が前回の認識結果と同一単語では無いが、あらかじめ定められた同じシステム動作を実行する認識単語である場合、言い換えであると判定してもよい。

【0017】

また、前記システム状態の停滞状態は、2つのシステム状態の往復が繰り返し続く状態であり、前記停滞脱出判定手段は、今回の認識結果が前々回の認識結果と同一単語である場合、言い直しであると判定し、今回の認識結果が前々回の認識結果と同一単語では無いが、あらかじめ定められた同じシステム動作を実行する認識単語である場合、言い換えであると判定してもよい。

【0018】

これによって、誤ってリジェクトされることによる音声操作の停滞および誤って認識されることによる音声操作の停滞が減ることになり、円滑な音声操作が実現できる。

【0019】

前記変更制御手段は、前記対話制御に関する設定の変更としてリジェクトの閾値の変更を行い、前記音声認識に関する設定の変更として前記音声認識辞書への新規追加または変更を行ってもよい。これによって、リジェクション精度及び認識精度向上が可能となり、ユーザにとって負担が少ない音声認識の個人適応と円滑な各種音声操作を実現することができる。

【0020】

また、前記変更制御手段は、前記リジェクトの閾値を認識対象単語ごとに設定し変更してもよい。これによって、認識対象単語ごとの個人適応が可能となり、よりユーザにとって負担が少ない音声認識の個人適応と円滑な各種音声操作を実現することができる。

【0021】

また、前記変更制御手段は、前記リジェクトの閾値、および、前記音声認識辞書への新規追加または変更を、ユーザごとに設定してもよい。これによって、複数のユーザが利用しても適切な適応が可能となり、よりユーザにとって負担が少ない音声認識の個人適応と円滑な各種音声操作を実現することができる。

【0022】

また、前記音声認識装置は、さらに、前記停滞状態から脱出した際に、今回の認識結果が前回の認識結果と同一単語では無く、かつあらかじめ定められた同じシステム動作を実行する認識単語でない場合、今回の認識結果の省略語を作成する省略語作成手段を備え、前記音声認識手段は、前記省略語を用いて前回の認識結果を再認識し、前記変更制御手段は、前記音声認識手段の再認識結果に応じて前記省略語を前記音声認識辞書へ新規追加してもよい。これによって、省略語をユーザが利用しても適切な適応が可能となり、よりユーザにとって負担が少ない音声認識の個人適応と円滑な各種音声操作を実現することができる。

【0023】

なお、本発明は、このような音声認識装置として実現することができるだけでなく、このような音声認識装置が備える特徴的な手段をステップとする音声認識方法として実現したり、それらのステップをコンピュータに実行させるプログラムとして実現したりすることもできる。そして、そのようなプログラムは、CD-ROM等の記録媒体やインターネット等の伝送媒体を介して配信することができるのは言うまでもない。

【0024】

以下、本発明の各実施の形態について、それぞれ図面を参照しながら説明する。

【0025】

(実施の形態1)

図1は、本発明の実施の形態1に係る音声認識装置を備えた音声対話型情報検索システムの構成を示すブロック図である。

【0026】

音声対話型情報検索システムは、音声を入力して対話しながら情報を検索するためのシ

10

20

30

40

50

ステムであり、図1に示すように音声認識部101、音声認識辞書102、音声認識パラメータ記憶部103、停滞脱出判定部104、対話制御部105、対話履歴記憶部106、システム仕様記憶部107、データベース検索部108、データベース記憶部109、応答音声・画面出力部110、およびタイマー111を備えている。

【0027】

音声認識部101は、音声認識辞書102および音声認識パラメータ記憶部103を用いて、ユーザより入力された音声の音声認識を行い、認識結果を出力する。音声認識辞書102は、認識対象語彙が登録されている辞書である。音声認識パラメータ記憶部103は、音声認識用パラメータを記憶している。

【0028】

対話制御部105は、予めシステムの開発者によって決められた動作仕様に従って対話を制御し、ユーザからの入力に対し次のシステム状態を決定する。具体的には、対話制御部105は、音声認識部101より入力される音声認識結果、停滞脱出判定部104より入力される停滞脱出か否かの判定結果、対話履歴記憶部106より入力される現在および過去の対話履歴を利用してシステム仕様記憶部107から次のシステム状態を決定する。また、対話制御部105は、必要があれば音声認識辞書102や音声認識パラメータ103の変更、およびデータベース検索をデータベース検索部108に要求する。なお、システム状態とはシステムの開発者によって決められたシステムの動作仕様におけるシステムの一状態を示す。

【0029】

停滞脱出判定部104は、対話制御部105より入力される現在と過去のユーザの認識結果等の情報を用いてシステムの状態遷移が停滞状態から脱出したか否かを判定する。対話履歴記憶部106は、対話制御部105から入力される音声認識結果やシステムの出力（出力画面情報・出力応答情報）結果など各システム状態における様々な情報を保存する。システム仕様記憶部107は、開発者によってあらかじめ決められたシステムの動作仕様を記憶している。

【0030】

データベース検索部108は、対話制御部105からの情報検索要求に対し、データベース記憶部109にあるデータベースから検索を行う。データベース記憶部109は、データベース検索部108の検索対象データベースを格納している。応答音声・画面出力部110は、対話制御部105より入力されるシステム状態に応じた画面や応答音声を出力する。タイマー111は、対話制御部105の要求により現時刻を対話制御部105に出力する。

【0031】

次に、上記のように構成された音声対話型情報検索システムにおいて、番組情報を検索する際の具体的な動作について説明する。図2は音声対話型情報検索システムでの対話全体の動作の流れを示すフローチャートである。

【0032】

対話制御部105は、対話開始のシステム状態を決定し、決定したシステム状態での画面と応答音声を決定し、応答音声・画面出力部110から出力することで、ユーザに対して入力要求を行う（ステップS201）。図3は具体的な出力画面例を示すである。ここでは、例えば図3に示すように番組情報を検索する際のメニュー画面が出力され、エージェントの吹き出しの内容301が応答音声として音声出力される。なお、吹き出し自体も画面表示してもよい。また、この例では図3における認識可能な語彙は四角で囲まれた語彙のみであるとする。例えば、四角「1. 番組名検索」302を選択するのに認識可能な語彙としては「1番」「1」「番組名検索」「1. 番組名検索」であるとする。

【0033】

音声認識部101は、システムからの応答音声・画面による入力要求に対しするユーザからの入力音声の認識処理を行う（ステップS202）。より詳細には、まず、対話制御部105は、音声認識部101に現在のシステム状態で認識可能な語彙の通知と音声認識

10

20

30

40

50

処理実行の要求を行う。より具体的には、図3に示すシステム状態においては、音声認識部101は四角で囲まれた語彙を認識対象語彙として音声認識処理を開始する。次に、音声認識部101は、ユーザの入力音声に対して認識処理を行い、対話制御部105に対し、認識結果を出力する。ここで、出力される認識結果は、ユーザの発声に最も近い認識対象語彙の単語だけではなく、認識に関する後に記述するような詳細な情報も含め出力する。

【0034】

図4および図5は出力される認識結果の具体的な例を示す図であり、図4は1位の認識結果を中心とした音声認識の全体的な情報を示しており、図5は他の候補も含めた認識結果の情報を示している。ここで、項目401は認識結果が出力された日時であり、項目402は入力された音声の区間、即ち音声認識部101が認識処理を行っていた区間のうち音声であると判断した区間である。項目403は認識対象語彙の中で最も近いと判断された単語、即ち認識結果の候補が1位の単語であり、項目404は音声認識辞書とは関係なく音響的に近いカナ文字を認識結果とした文字列であり、一般には音声タイプライタの結果と呼ばれるものである。項目405は入力音声区間の中で認識結果の単語がマッチングした区間である。項目406は認識度合を示す認識スコアであり、スコアが高い方がより認識度合が高いことを示している。項目407は認識信頼度を示し、どの程度認識結果が妥当かを示している。認識信頼度は一般的には、認識候補のスコアの差や音声タイプライタと認識候補の差などを用いて計算する場合が多い。項目408はリジェクト用閾値であり、音声認識パラメータ記憶部102に記憶されている変数である。

10

20

【0035】

対話制御部105は、このリジェクト閾値と認識信頼度との比較を行いシステムとして認識結果を受け入れるか否かの判定を行う。具体的にはリジェクト閾値より認識信頼度が低い場合、対話制御部105は認識結果をリジェクト、即ち入力結果として処理せず、再度同じシステム状態での入力を促す。例えば、図4の例では認識信頼度が「4.5」でリジェクト閾値が「3.5」であるので、対話制御部105はこの認識結果「番組名検索」をシステムへの入力として認め対話制御を行う。なお、このリジェクト閾値は、予めシステム開発者が決定しても良いし、評価実験を行うことにより決定してもよい。具体的には何人かの被験者にこの辞書セットの単語を発声させ、その結果を基に決定してもよい。

30

【0036】

また、図5において、項目501は認識候補の認識スコアのよってソートされた結果の認識候補順位であり、項目502から項目505は各認識候補の情報であり、その内容は図4で説明した認識結果1位の結果の情報と同じである。

【0037】

対話制御部105は、ステップS202で音声認識部101から入力された音声認識結果と対話履歴記憶部106に蓄積されている前回の認識結果を停滞脱出判定部104に出力する。

【0038】

次に、停滞脱出判定部104は、今回の入力が停滞の脱出であるか否かの判定を行い、その結果を対話制御部105に出力する(ステップS203)。対話制御部105は、この結果を対話履歴記憶部106に書き込む。

40

【0039】

ここで、停滞脱出判定部104における停滞脱出判定動作について、音声認識の誤リジェクトによる停滞を例に取り、詳細に説明する。図6は、停滞脱出判定部104における停滞脱出判定動作の流れを示すフローチャートである。

【0040】

まず、停滞脱出判定部104は、今回の音声認識結果および前回の認識結果を取得する(ステップS601)。そして、その音声認識結果に基づいてリジェクトか否かの判定を行う(ステップS602)。この判定の結果、リジェクトと判定した場合(ステップS602でYES)、停滞脱出判定部104は停滞脱出でないという判定結果を出力する。こ

50

れは、リジェクトとは認識結果の信頼度が低いため認識結果として採用されないということであるので、その場合は次のシステム状態へ進まない状態、即ち停滞からの脱出ではないためである。

【0041】

一方、リジェクトでないと判定した場合（ステップS602でNO）は、対話履歴から前回の発声がリジェクトであったか否かの判定を行う（ステップS603）。この判定の結果、前回の発声がリジェクトでないと判定した場合（ステップS603でNO）は、前回の発声においては停滞が発生していないため、停滞脱出判定部104は今回の発声は停滞の脱出ではないという判定結果を出力する。

【0042】

一方、前回の発声をリジェクトと判定した場合（ステップS603でYES）、前回の発声によりシステムは停滞状態であったことを示すため、停滞脱出判定部104は、今回の発声により停滞状態から脱出できたという判定し、言い直しであるか否かの判定を行う（ステップS604）。ここでのいい直しとは、前回の発声と今回の発声が同じであることを意味する。例えば、ユーザが図3のような出力画面において「番組名検索」と発声し、リジェクトされて再度入力を促されたときにもう一度「番組名検索」と発声する場合などである。この言い直し判定は、前回の認識結果と今回の認識結果とを比較することでを行い、言い直しであると判定した場合（ステップS604でYES）、停滞脱出判定部104は言い直しによる停滞の脱出であるという判定結果を出力する。

【0043】

一方、言い直しでないと判定した場合（ステップS604でNO）は、停滞脱出判定部104は、言い換えであるか否かの判定を行う（ステップS604）。ここでの言い換えとは、前回の発声と今回の発声が発声語彙は異なるが、発声内容が同じ、即ち発声によるシステム動作が同じ発声を意味する。例えば、ユーザが図3のような出力画面において「番組名検索」と発声し、リジェクトされて再入力を促されたときに「1番」と発声する場合などである。この言い換えの判定は、言い直しの判定と同様に前回の認識結果と今回の認識結果の比較を行うことで判定を行う。より具体的には、前回の認識結果と今回の認識結果との語彙が異なり、且つシステム仕様として認識結果が同じ動作を実行する語彙であれば言い換えであると判定する。システム仕様として認識結果が同じ動作か否かの判定は、システム仕様記憶部107に定義される各システム仕様により判定する。具体的には、システム仕様記憶部107には図7に示されるように、認識結果として受け付ける語彙とその語彙を受け付けたときどの状態に遷移するかが記憶されており、ここで一つの選択可能項目に対応する単語を言い換え対象語として扱う。

【0044】

この判定の結果、言い換えであると判定した場合（ステップS605でYES）、停滞脱出判定部104は、言い換えによる停滞脱出であるという判定結果を出力する。一方、言い換えでないと判定した場合（ステップS605でNO）、停滞脱出判定部104は停滞脱出ではないという判定結果を出力する。

【0045】

以上のように、停滞脱出判定部104は停滞脱出判定の動作を行う。

次に、対話制御部105は、停滞脱出判定処理（ステップS203）までに得られている音声認識結果および停滞脱出判定結果に基づいて、音声認識辞書やリジェクト閾値、音響モデルといった音声認識パラメータの変更を行う（ステップS204）。

【0046】

次に、対話制御部105は、認識結果に基づいて次のシステム状態と、このシステム状態における応答音声および画面の出力について決定し、応答音声・画面出力部110に出力する（ステップS205）。ここで必要であれば、対話制御部105は、データベース検索部108に対しデータベース記憶部109からのデータの検索を要求した結果を応答音声・画面出力部110に出力する。

【0047】

10

20

30

40

50

そして、対話制御部 105 は、システム仕様記憶部 107 に定義されているシステム仕様に従い、対話の終了か否かを判定する。この結果、対話の終了でない場合（ステップ S 206 で N O ）には、再び入力音声の認識処理（ステップ S 202 ）より上記ステップを繰り返し、対話の終了である場合（ステップ S 206 で Y E S ）には、対話を終了する。

【0048】

次に、システムの具体動作例をシステムのシステム出力画面と対話履歴記憶部 106 に保存される対話履歴データの具体例を用いて説明する。

【0049】

図 8 は、動作例で対象とする対話履歴データの具体例を示す図である。項目 801 はシステム状態の変化を一元管理するために振られているステップ番号、項目 802 はシステム状態の種類を示すシステム状態、項目 803 はシステムが応答を出力した日時を示す応答出力開始時刻、項目 804 は音声認識部 101 から得られる音声認識結果の 1 位候補の単語、項目 805 も同様に音声認識結果から得られる認識信頼度、項目 806 は音声認識部 101 からの音声認識結果に基づいて対話制御部 105 が判定したリジェクト判定結果、項目 807 は対話履歴記憶部 106 に保存される前回の認識結果と今回音声認識部 101 が出力した認識結果に基づいて停滞脱出判定部 104 が判定した言い直しによる停滞脱出の判定結果、項目 808 は項目 807 と同様にして停滞脱出判定部 104 が判定した言い換えによる停滞脱出の判定結果、項目 809 は音声認識パラメータ記憶部 103 に保存されており、認識結果からも取得できるリジェクト閾値である。なお、この図には示していないが、各ステップにおける図 4 で示されるような認識結果の詳細情報や図 5 に示されるような表示画面についての情報、具体的には表示されている単語やシステムがどのような応答文を出力したかを示す出力応答文など他の情報も対話履歴記憶部 106 には保存してもよい。

10

20

【0050】

例えば、ユーザが、図 3 に示すメニュー画面で「番組名検索」と発声したとする。この認識結果の認識信頼度（0.47）がリジェクト閾値（0.35）より高いので、対話制御部 105 は、次のシステム状態を決定し、画面遷移と応答文の出力をを行う（図 8 のステップ = 1）。具体的には、システムからは応答音声・画面出力部 110 によって図 9 に示されるような画面と「番組名の頭文字を指定してください」という応答音声が出力される。次に、ユーザは「あ行」と発声し、これも先の発声と同様に、認識確信度（0.36）がリジェクト閾値（0.35）より高いため、正しく受け付けられる（図 8 のステップ = 2）。システムからは応答音声・画面出力部 110 によって図 10 のような画面と「どの番組ですか？」という応答が出力される。次に、ユーザはそのリストには見たい番組が無く「次の画面」と発声するが、この発声に対する認識結果では、認識信頼度（0.33）がリジェクト閾値（0.35）より低いためリジェクトであると判定される（図 8 のステップ = 3）。リジェクトと判定された場合、対話制御部 105 は再度そのシステム状態で（今の場合、対話 = 状態 3）再度入力を促す。なお、この動作はユーザが正しく発声しているのに対し、対話制御部 105 が誤ってリジェクトしてしまったシステムの誤動作であり、リジェクト閾値がユーザにとって正しく設定されていないため生じる動作である。

30

【0051】

40

再度同じシステム状態で、システムより入力を促されたユーザは再び「次の画面」と発声し、その音声認識の結果における認識信頼度（0.38）はリジェクト閾値（0.35）より高いので、対話制御部 105 はその結果を受け付ける（図 8 のステップ = 4）。ここで、このステップでは停滞脱出判定部 104 が「前回の発声はリジェクト」かつ「今回の発声は言い直し」であるので「言い直しによる停滞脱出」と判定し、項目 807 にその情報が記憶される。更にこのステップでは、対話制御部 105 は検出した誤動作と正しい動作を用いて、誤動作したはじめの発声を次からは受け付けるよう個人適応を行う。即ち、音声認識パラメータ、今回の例ではリジェクト閾値を変更し、次のステップからこの値を利用して音声認識を行う。具体的には、現在のリジェクト閾値「0.35」を前回の誤ってリジェクトされた発声における信頼度でも正しく認識できるように「0.30」に変

50

更する。この閾値の変更は、システム開発者が予め設定した、決まった割合で変更を行っても良い。また、現在のリジェクト閾値と誤ってリジェクトされたときの認識信頼度を利用した計算により閾値の変更を行ってもよい。より具体的には、現在のリジェクト閾値と誤ってリジェクトされたときの認識信頼度の差分が一定値以内であれば、リジェクト閾値を誤ってリジェクトされたときの認識信頼度に設定し、差分が一定値以上であれば、現在のリジェクト閾値と誤ってリジェクトされたときの認識信頼度の間の重み付き平均値を利用してリジェクト閾値を設定しても良い。また、リジェクトされた単語と正しく認識された単語の認識信頼度を用いて閾値の変更を行ってもよい。具体的には、現在のリジェクト閾値と誤ってリジェクトされたときの認識信頼度を用いた計算方法と同様の方法で決定する。

10

【0052】

言い直しの結果を受け付けた対話制御部105は、次のシステム状態を決定し、画面遷移と応答文の出力を行う。具体的には、システムからは応答音声・画面出力部110によって、図11に示されるような画面と「どの番組ですか？」という応答音声が出力される。ユーザはこの画面にも見たい番組が無いので、さらに「次の画面」と発声する。この発声の認識結果における認識信頼度はステップ3の時と同じ「0.33」である。この認識信頼度はステップ3ではリジェクトされた値であるが、対話制御部105はこの認識信頼度「0.33」と前ステップで適応させたリジェクト閾値「0.3」とを比較した結果、本ステップではこの発声をリジェクトせず、次のシステム状態を決定し、画面遷移と応答文の出力を行う。具体的には、システムからは応答音声・画面出力部110によって、図12に示されるような画面と「どの番組ですか？」という応答音声が出力される（図8のステップ=5）。ユーザはこの画面の中では見たい番組を見つけ、「iしたい」と番組を選択する発声を行う（図8のステップ=6）。図13は、以上の一連の動作をまとめた図であり、上から順に図8のステップ=1からステップ=6に対応する。

20

【0053】

次に、言い換えを利用したリジェクト閾値の変更動作例について、対話履歴データの具体例を用いて説明する。図14は、動作例で対象とする対話履歴データの具体例を示す図である。なお、対話履歴データの項目は図8と同じであるので、説明は省略する。更に、上記言い直しによるリジェクト閾値の変更動作例との発声の違いはステップ3～ステップ5のみであるので、図14のステップ3からステップ5の動作例についてのみ説明する。

30

【0054】

システムから応答音声・画面出力部110によって図10のような画面と「どの番組ですか？」という応答が出力される。ユーザはそのリストには見たい番組が無いため「次の画面」と発声するが、この発声に対する認識結果では、認識信頼度（0.33）はリジェクト閾値（0.35）より低いためリジェクトであると判定される（図14のステップ=3）。リジェクトと判定された場合、対話制御部105は再度そのシステム状態で（今の場合対話=状態3）再度入力を促す。

【0055】

再度同じシステム状態で、システムより入力を促されたユーザは「次の画面」と同じシステム動作を行うコマンドである「5番」と発声する。この音声認識の結果における認識信頼度（0.38）はリジェクト閾値（0.35）より高いので、対話制御部105はその結果を受け付ける（図14のステップ=4）。ここで、このステップでは停滞脱出判定部104が「前回の発声はリジェクト」かつ「今回の発声は言い換え」であるので「言い換えによる停滞脱出」と判定し、項目1408にその情報が記憶される。さらに、このステップでは、対話制御部105は検出した誤動作と正しい動作を用いて、誤動作したはじめの発声を次からは受け付けるよう個人適応を行う。即ち、音声認識パラメータ、今回の例ではリジェクト閾値を変更し、次のステップからこの値を利用して音声認識を行う。具体的には現在のリジェクト閾値「0.35」を前回の誤ってリジェクトされた発声における信頼度でも正しく認識できるように「0.3」に変更する。以降の動作は上記言い直しによるリジェクト閾値の変更動作例と同じであるので省略する。

40

50

【 0 0 5 6 】

なお、上記具体例の中では「言い直しまたは言い換えによる停滞脱出」を1回検出した段階でリジェクト閾値を変更したが、音声認識パラメータの変更を行う基準としての停滞脱出検出の回数は可変に設定できるようにしてもよい。例えば3回に設定すると、「言い直しまたは言い換えによる停滞脱出」が3回検出されたらリジェクト閾値の変更を行うことになる。この場合、例えば3回分の認識結果における認識信頼度を用いてリジェクト閾値を変更してもよい。より具体的には、3回分の認識結果における信頼度の最低値や平均値、重み付け平均値により決定する。

【 0 0 5 7 】

また、上記具体例ではリジェクト閾値を1つしか持たない例について述べたが、単語ごとにリジェクト閾値を持ち、「言い直しまたは言い換えによる停滞脱出」を単語ごとに検出し、閾値を変更してもよい。具体的には、例えば図15のようなデータを音声認識パラメータ記憶部103に保存する。ここで、項目1501は停滞脱出をしたことによりリジェクト閾値が変更された単語であり、項目1502はその単語のリジェクト閾値である。なお、このリストに無い単語はデフォルト値、例えば上記具体例では「0.35」を利用する。

10

【 0 0 5 8 】

図16は本実施の形態を利用した場合と利用しない場合の対話シーケンスの例を示す図である。この図16に示す例では、本実施の形態を利用した場合の方がユーザの発声が1回少なくて済む。この例では、ユーザは2ページ目で番組の選択を決定しているが、より多くのページを見ていく場合のように多くのステップを有する対話では本実施の形態の有効性は顕著に現れることになることは容易に理解できる。また、一度検索が終わり、再び同じ番組をはじめから選択する場合も本実施の形態を用いれば前回リジェクトされた発声方法でも初めから認識されることになる。

20

【 0 0 5 9 】

このように本実施の形態によると、一連の対話シーケンスの中で、誤動作と正しい動作を検出することで音声認識パラメータを適切に変更することが可能となる。この結果、次に前回誤動作をした発声を行ってもシステムは正しい動作が可能となるため、何度も繰り返し言い直しをする必要が無く、スムーズでユーザに負担の掛からない対話が実現できる。また、本実施の形態による音声認識パラメータの変更は、変更のために特別な発声を促すわけでは無いので、ユーザの負担も少ない。

30

【 0 0 6 0 】

なお、本実施の形態は、図17に示すように上記構成に加えてEPG(Electronic Program Guide)を受信するEPG受信部201を備え、EPGを対象として音声認識を行つて情報を検索する音声対話型情報検索システムにおいても適用することが可能である。この場合、EPG受信部201で受信されたEPGは、データベース記憶部109に記憶される。対話制御部105は、データベース記憶部109に記憶されているEPGを用いて音声認識辞書102を作成する。そして、音声認識部101は、EPGを用いて作成された音声認識辞書102を用いて、ユーザより入力された音声の音声認識を行う。また、データベース検索部108は、データベース記憶部109に記憶されているEPG等を対象として検索を行うことになる。

40

【 0 0 6 1 】**(実施の形態2)**

上記実施の形態1によれば、誤動作と正しい動作を検出することで音声認識パラメータの個人適応が可能となり、ユーザに負担の少ない個人適応が実現できるが、同様の適応を音声認識辞書の追加という形でも行える。本実施の形態では、誤動作と正しい動作の検出しを行い、音声認識辞書の変更または新たに登録を行う方法について述べる。

【 0 0 6 2 】

本実施の形態は、上記実施の形態1とは図1における対話制御部105における停滞脱出判定結果に基づいて個人適応する対象が異なるものであり、他は実施の形態1と同様で

50

ある。従って、基本的には図1から図12を参照することとする。以下、本実施の形態における対話制御部105の動作と、前実施の形態では述べていない音声認識辞書の変更処理について説明する。

【0063】

本実施の形態における辞書変更・登録による個人適応の動作例について、対話履歴記憶部106に記憶されている対話履歴データの具体例を用いて説明する。

【0064】

図18は、対話履歴記憶部106に記憶されている対話履歴データの具体例を示す図である。図18に示される対話履歴データの例は実施の形態1での対話シーケンスにおける対話履歴データの例(図8)と同様の履歴であり、図8には示されていなかった項目「認識結果2」が示されている点、および図8に示されていた項目「応答出力開始時刻」が省略されている点を除いては図8と同じものである。なお、既に述べたが認識結果2は音声認識辞書を使わず、音響的に近いかな文字列を音声認識結果として出力されたものであり、認識結果の一例を示した図4における音声認識結果2と同一のものである。

10

【0065】

以下、図18の項目「ステップ」を用い、順に具体的な動作を説明する。

ステップ3では、ユーザの発声「次の画面」に対し、音声認識部101は認識結果2「スイノダメン」、認識信頼度「0.33」、リジェクト閾値「0.35」を出力する。対話制御部105は、認識信頼度がリジェクト閾値より低いため、リジェクトと判定し、再度そのシステム状態での再度入力を促す。ステップ4では、ユーザの再発声「次の画面」に対し、音声認識部101は認識結果2「ツリノガメン」、認識信頼度「0.38」、リジェクト閾値「0.35」を出力し、停滞脱出判定部104は「言い直しによる停滞脱出」との判定を出力する。対話制御部105は、これらの結果を受けて、誤動作したステップ3における発声が次回からは正しく認識されるように、個人適応を行う。即ち、ステップ3でリジェクトされた発声に対する音声認識結果2の「スイノダメン」をステップ4で正しく認識されたコマンド「次の画面」に対応させて音声認識辞書102に新規に登録を行う。

20

【0066】

図19は音声認識辞書の具体例を示す図である。項目1801は単語ごとにユニークに付与される単語番号、項目1802はシステム仕様で同じ意味として扱われる番号を同一番号として付与された意味番号、項目1803は単語の表記、項目1804は単語の読みである。ここで、上記例においては、図19の単語番号130が新規登録されることになる。

30

【0067】

ステップ5では、ユーザが「次の画面」と発声する。音声認識部101からはステップ3の時と同様に音声認識結果2として「スイノダメン」という結果が出力されるが、このときの音声認識時には音声認識辞書102に「スイノダメン」が「次の画面」と対応された状態で登録されているため、高い確信度(今の場合0.45)が結果として出力される。このように、ステップ5における発声は個人適応されたことに伴い、ステップ3と同様の発声であるにもかかわらず正しく認識が行われる。

40

【0068】

なお、上記具体例の中では言い直しの「言い直しによる停滞脱出」を1回検出した段階で認識辞書の変更を行ったが、認識辞書変更を行う基準としての停滞脱出検出の回数は可変に設定できるようにしてもよい。例えば3回に設定すると、「言い直しによる停滞脱出」が3回検出されたら認識辞書の変更を行うことになる。ここで、3回分の認識結果における認識結果を全て登録しても良いが、組み合わせて作成した文字列を登録してもよい。具体的には「ツギノガメン」に対して「スイノダメン」「ツイノダメン」「スギノダメン」に対して、全てが共通している「ダ」の部分だけを変更した「ツギノダメン」を登録してもよい。さらに、変更されたかな文字を記憶し、このユーザは「ガ」を「ダ」とよく間違えると判定した場合、他の単語についても「ガ」を「ダ」に変更してもよい。具体的に

50

は「前の画面」に対し「マエノダメン」という読みを付与し、音声認識辞書に追加登録しても良い。

【0069】

また、本実施の形態では言い直しの停滞の判定により音声認識辞書の追加・変更の例についてのみ述べたが、実施の形態1と同様にすれば言い換えの場合も音声認識辞書の追加・変更を行うことができる。

【0070】

このように本実施の形態によると、一連の対話シーケンスの中で、誤動作と正しい動作を検出することで音声認識パラメータだけでなく、音声認識辞書についても適切に変更することが可能となる。この結果、次に前回誤動作をした発声を行ってもシステムは正しい動作が可能となるため、何度も繰り返し言い直しをする必要が無く、スムーズでユーザに負担の掛からない対話が実現できる。また、本実施の形態による音声認識辞書の変更は、認識率を上げるために特別な発声を促すわけではなく自然な対話から認識率を上げるため、ユーザの負担も少ない。

【0071】

なお、本実施の形態における音声認識辞書への追加・変更と上記実施の形態1における音声認識パラメータの変更とを組み合わせて実施することも可能である。

【0072】

(実施の形態3)

上記実施の形態1および実施の形態2によれば、一連の対話シーケンスの中で、誤動作と正しい動作を検出することで音声認識パラメータおよび認識辞書をユーザに適したものに変更しているが、上記実施の形態1および実施の形態2においては、「言い換え」を前回リジェクトされた単語が、今回正しく認識された単語と同一のシステム動作を行う単語であるかをシステム仕様記憶部にある図7のようなデータを用い判定している。しかし、「言い換え」には様々な形があり、事前にシステム仕様に登録できない場合がある。特にEPGを用いた番組検索システムにおいては、日々更新される番組名を認識対象とする必要があり、予め言い換えについてシステム開発者が登録しておくことができない。本実施の形態は、このような場合に対処するものである。

【0073】

図20は、本発明の実施の形態3に係る音声認識装置を備えた音声対話型情報検索システムの構成を示すブロック図である。

【0074】

本実施の形態3と上記実施の形態1および実施の形態2との相違点は、省略語作成部301とユーザ発声記憶部304が追加されたことによる停滞脱出判定部302の動作が異なる点であり、他の動作は上記実施の形態1および実施の形態2と同一である。従って、本実施の形態においては、動作が異なる停滞脱出判定部302の言い換え判定の動作についてのみ説明する。

【0075】

停滞脱出判定部302は、上記実施の形態1および実施の形態2と同様に図6のフローチャートに従って、言い直しおよび言い換えによる停滞脱出の判定を行うが、図6のステップS605における処理、即ち今回の発声が言い換えか否かの判定を行う処理が異なる。図21は本実施の形態における言い換え判定動作の流れを示すフローチャートである。

【0076】

まず、今回の発声が前回の発声と同一のシステム動作を行う認識単語であるか否かを判定する(ステップS2001)。この判定の結果、前回の発声と同一のシステム動作を行う認識単語である場合(ステップS2001でYES)、これまでの実施の形態同様の動作であり、言い換えによる停滞脱出と判定する(図6のステップS605でYESの判定)。一方、前回の発声と同一のシステム動作を行う認識単語でない場合(ステップS2001でNO)、今回の認識対象語彙から省略語が作成される(ステップS2002)。省略語の作成は、今回の認識対象語彙を用いて省略語作成部301において行われる。

10

20

30

40

50

【0077】

省略語作成部301は、今回の認識対象語彙を受け取り、予め定義されているルールに基づいて省略語を作成する。省略語作成方法としては、形態素解析ツールなどを用いて今回の認識対象語彙を形態素に分解し、その分解結果を基に作成する。例えば、一つの形態素を省略語としても良いし、複数の形態素をつなげて省略語としても良い。より具体的には例えば、「発掘あるある広辞苑」という単語に対して「発掘」「あるある」「広辞苑」「あるある広辞苑」といった省略語を作成したり、「冬のレクイエム」という単語に対して「冬レク」といった省略語を作成したりする。省略語作成部301で作成された省略語は、停滞脱出判定部302を介して対話制御部303に保持される。

【0078】

次に、音声認識部101は、対話制御部303に保持されている省略語作成部301で作成された省略語を用いて、ユーザ発声記憶部304に記憶されている前回リジェクトされた発声について、再度認識を行う（ステップS2003）。

【0079】

そして、停滞脱出判定部302は、再認識結果の信頼度とリジェクト閾値とを比較する（ステップS2004）。ここで、再認識結果の信頼度がリジェクト閾値より高い場合（ステップS2004でYES）、対話制御部303は、認識候補1位の省略語を今回認識された単語と同じ動作を行う単語としてシステム仕様記憶部107および音声認識辞書102に登録（ステップS2005）し、言い換えによる停滞脱出と判定する（図6のステップS605でYESの判定）。一方、再認識結果の信頼度がリジェクト閾値より低い場合（ステップS2004でNO）、停滞脱出判定部302は言い換えによる停滞脱出では無いと判定する（図6のステップS605でNOの判定）。

【0080】

以上の動作より、システム仕様で音声認識辞書102に登録されていない省略語をユーザが発声してリジェクトされても、次の発声で正しい表現での発声を行い認識されれば、前回発声した省略語は新たに登録されるため、次回から認識が可能となる。これにより、省略語を発声してしまうユーザに対して何度もリジェクトすることが無く、スムーズでユーザに負担の掛からない対話が実現できる。さらに、本実施の形態による省略語の作成には特別な発声を促すわけでは無いので、ユーザの負担も少ない。

【0081】

（実施の形態4）

上記実施の形態1から実施の形態3によれば、一連の対話シーケンスの中で、誤動作と正しい動作を検出することで音声認識パラメータの変更および認識辞書の変更を行い、個人適応を可能としたが、複数のユーザが利用することを想定していないため、複数のユーザが利用した場合、正しく個人適応できない。本実施の形態は、このような場合に対処するものである。

【0082】

図22は、本発明の実施の形態4に係る音声認識装置を備えた音声対話型情報検索システムの構成を示すブロック図である。

【0083】

本実施の形態と上記実施の形態3との相違点は、ユーザ入力部401およびユーザ情報記憶部402が追加されたことによる対話制御部403における個人適応処理が異なる点であり、他は実施の形態1から実施の形態3までと同一である。従って、本実施の形態においては、複数ユーザが利用する際の対話制御部403の動作について説明する。

【0084】

対話制御部403は、ユーザ入力部401からユーザ名が入力されると、ユーザ情報記憶部402より、入力されたユーザ名に適応された音声認識パラメータや認識対象辞書が登録されているか否かの確認を行う。もし、入力されたユーザ名に適応された音声認識パラメータや認識対象語彙が無い場合、音声認識パラメータや音声認識辞書は初期値を利用しシステムを動作させる。もし、ユーザ適応されていないユーザが、システムを利

10

20

30

40

50

用中に停滞脱出判定部 302 により誤動作と正しい動作のシーケンスが検出され、実施の形態 1 から実施の形態 3 で説明したような音声認識パラメータや音声認識対象語彙の変更が必要となると、対話制御部 403 はユーザ情報記憶部 402 に新規ユーザのユーザ名と音声認識パラメータや辞書を変更した単語についての各種情報を記憶する。

【0085】

一方、ユーザ入力部 401 より入力されたユーザ名に適応された音声認識パラメータや認識対象辞書がユーザ情報記憶部 402 に登録されている場合、対話制御部 403 はユーザ情報記憶部 402 から以前登録されたユーザ名の個人適応後の音声認識パラメータや新規認識辞書登録単語を抽出し、音声認識パラメータ記憶部 103 や音声認識辞書 102 にその情報を登録する。

10

【0086】

図 23 はユーザ情報記憶部 402 に記憶される各種情報の具体例を示す図である。なお、図 23 の例では単語ごとにリジェクト閾値を持つ場合の具体例を示す。項目 2201 はユーザ名であり、項目 2202 は停滞単語、すなわち音声認識パラメータや辞書を変更した単語であり、項目 2203 は停滞脱出回数、すなわち何度停滞脱出が検出されたかを示し、項目 2204 は変更した音声認識パラメータであるリジェクト閾値、項目 2205 は音声認識辞書に追加した新規登録読みである。

【0087】

図 23 に示されるデータがユーザ情報記憶部 402 に記憶されている場合、ユーザ入力部 401 からユーザ名 A が入力されると、対話制御部 403 はユーザ名 A の個人適応情報として「次の画面」の単語に対し、リジェクト閾値「3.4」を、新規読み登録として「ツリノダメン」を、「前の画面」に対しリジェクト閾値「3.5」を、新規読み登録として「マエノダメン」をそれぞれ音声認識パラメータ記憶部 103 および音声認識辞書 102 に登録する。

20

【0088】

以上の動作より、一連の対話シーケンスの中で、誤動作と正しい動作を検出することで音声認識パラメータおよび音声認識辞書の個人適応が可能となるだけではなく、複数のユーザがシステムを利用した際にも正しく個人適応が可能となり、ユーザに負担の少ない個人適応と円滑な対話が実現できる。

【0089】

30

なお、本実施の形態ではユーザ入力部の入力を基にユーザの判別を行い、複数のユーザに対応した個人適応を行ったが、話者識別や話者判別の技術は現在一般的に存在するので、それらの技術を用いてユーザの判別を行ってもよい。

【0090】

(実施の形態 5)

上記実施の形態 1 から実施の形態 4 ではシステムの停滞状態として、誤ったリジェクトによる停滞状態を対象としたが、誤認識により誤ったシステム状態へ遷移した場合に発生する停滞状態について述べていない。そこで、本実施の形態は、このような停滞状態に対処するものである。

【0091】

40

誤認識により誤ったシステム状態へ遷移した場合に発生するシステムの停滞状態の具体例としては、「時間検索」とユーザが発声したのに対し、システムがこれを「ジャンル検索」と認識し、ユーザが思っていたシステム状態と別のシステム状態へ遷移するような場合がある。このとき、ユーザはこの誤認識によるシステムの誤った状態遷移を基に戻るために「戻る」といった元の状態に戻るためのコマンドを発声する。システムの状態が戻るとユーザは再度「時間検索」を発声する。この一連の動作は 2 つのシステム状態の往復が繰り返し続く状態であり、一つの停滞状態といえる。

【0092】

本実施の形態では、上記実施の形態 4 と比べシステム構成としての変更はなく、異なるのは停滞脱出判定部 302 における停滞判定の動作処理（図 6 のフローチャート）であり

50

、他は実施の形態 4 と同様である。

【0093】

本実施の形態における停滞脱出判定部 302 の動作処理について説明する。図 24 は本実施の形態における停滞脱出判定部 302 の動作の流れを示すフローチャートである。なお、下記の説明における過去の認識結果は対話履歴記憶部 106 に記憶されているデータを参照して利用し、言い直しや言い換えの判定は上記実施の形態 1 から実施の形態 4 に述べた方法と同じ方法で行う。

【0094】

まず、停滞脱出判定部 302 は、今回の音声認識結果を取得する（ステップ S2301）。次に、この音声認識結果がリジェクトか否かの判定を行う（ステップ S2302）。この判定の結果、リジェクトと判定した場合（ステップ S2302 で YES）、停滞脱出ではないと判定し、処理を終了する。一方、リジェクトではないと判定した場合（ステップ S2302 で NO）、前回の発声が状態を戻す発声（上記例では「戻る」）であったか否かを判定する（ステップ S2303）。この判定の結果、前回の発声が状態を戻す発声でない場合（ステップ S2303 で NO）、停滞脱出では無いと判定し、処理を終了する。一方、前回の発声が状態を戻す発声である場合（ステップ S2303 で YES）、今回の発声が前々回の発声の言い直しか否かの判定を行う（ステップ S2304）。この判定の結果、言い直しである場合（ステップ S2304 で YES）、いい直しによる停滞脱出と判定し、処理を終了する。一方、言い直しでない場合（ステップ S2304 で NO）、今回の発声が前々回の発声の言い換えか否かの判定を行う（ステップ S2305）。この判定の結果、言い換えである場合（ステップ S2305 で YES）、言い換えによる停滞脱出と判定し、処理を終了する。一方、言い換えでない場合（ステップ S2305 で NO）、停滞脱出ではないと判定し、処理を終了する。なお、このようにして検出された言い直しや言い換えによる停滞脱出は、上記実施の形態 1 から実施の形態 4 で述べた、誤ったリジェクトによる停滞からの脱出と区別して対話履歴保存部 106 に保存する。

【0095】

このようにして誤認識による停滞脱出の判定を行い、音声認識用パラメータや認識時書の変更を行う。具体的には例えば、上記実施の形態 2 で述べたような認識辞書の変更を行う。より具体的には、前々回の誤認識された発声に対しての音声認識結果のうち、音声認識辞書を使わず、音響的に近いかな文字列を音声認識結果として出力された結果（例えば図 4 の認識結果 2）を、今回得られた正しく認識された結果の単語に対応付けて音声認識辞書に追加する。

【0096】

以上の動作より、誤ったリジェクトによるシステム状態の停滞のみでなく、誤認識によるシステム状態の停滞を利用した個人適応を行うので、次に前回誤動作をした発声を行っても誤認識による停滞が発生しなくなるため、スムーズでユーザに負担の掛からない対話が実現できる。また、本実施の形態による音声認識パラメータや音声認識辞書の変更は、専用の特別な発声を促すわけではなく自然な対話から認識率を上げるために、ユーザの負担も少ないのである。

【0097】

なお、上記各実施の形態において、音声認識部は音声認識手段に、停滞脱出判定部は停滞脱出判定手段に、対話制御部は対話制御手段および変更制御手段に、省略語作成部は省略語作成手段に対応する。

【産業上の利用可能性】

【0098】

本発明に係る音声認識装置および音声認識方法は、音声対話型インタフェースを持つ多くのシステムに対して利用可能であり、例えば家庭内の情報検索システムやカーナビゲーションシステム、携帯端末からの情報検索などにおいて有用であり、その利用可能性是非常に大きい。

【図面の簡単な説明】

10

20

30

40

50

【0099】

【図1】本発明の実施の形態1に係る音声認識装置を備えた音声対話型情報検索システムの構成を示すブロック図である。

【図2】本発明における対話全体の処理の流れを示すフローチャートである。

【図3】本発明における音声対話型情報検索システムの出力画面例を示す図である。

【図4】本発明の実施の形態1における音声認識部から出力されて格納される認識結果例を示す図である。

【図5】本発明の実施の形態1における音声認識部から出力されて格納される認識結果例を示す図である。

【図6】本発明の実施の形態1における停滞脱出判定部における処理の流れを示すフロー 10 チャートである。

【図7】本発明の実施の形態1におけるシステム仕様記憶部におけるシステム動作仕様例を示す図である。

【図8】本発明の実施の形態1における対話履歴記憶部に記憶される対話履歴データ例を示す図である。

【図9】本発明における音声対話型情報検索システムの出力画面例を示す図である。

【図10】本発明における音声対話型情報検索システムの出力画面例を示す図である。

【図11】本発明における音声対話型情報検索システムの出力画面例を示す図である。

【図12】本発明における音声対話型情報検索システムの出力画面例を示す図である。

【図13】本発明の実施の形態1の対話例におけるシステム動作概略を示す図である。 20

【図14】本発明の実施の形態1における対話履歴記憶部に記憶される対話履歴データ例を示す図である。

【図15】本発明の実施の形態1における単語ごとにリジェクト閾値を設定しているデータ例を示す図である。

【図16】本発明の実施の形態1における対話シーケンスにおいて、本技術を利用した場合としない場合の比較を示す図である。

【図17】本発明の実施の形態1に係る音声対話型情報検索システムの他の構成を示すブロック図である。

【図18】本発明の実施の形態2における対話履歴記憶部に記憶される対話履歴データ例を示す図である。 30

【図19】本発明の実施の形態2における音声認識辞書に記憶される認識対象語彙の例を示す図である。

【図20】本発明の実施の形態3に係る音声認識装置を備えた音声対話型情報検索システムの構成を示すブロック図である。

【図21】本発明の実施3の形態による言い換え判定動作の処理の流れを示すフローチャートである。

【図22】本発明の実施4に係る音声認識装置を備えた音声対話型情報検索システムの構成を示すブロック図である。

【図23】本発明の実施4の形態におけるユーザ情報記憶部に記憶されるユーザ情報データ例を示す図である。 40

【図24】本発明の実施5の形態における停滞脱出判定部における処理を示すフローチャートである。

【符号の説明】

【0100】

101 音声認識部

102 音声認識辞書部

103 音声認識パラメータ記憶部

104、302 停滞脱出判定部

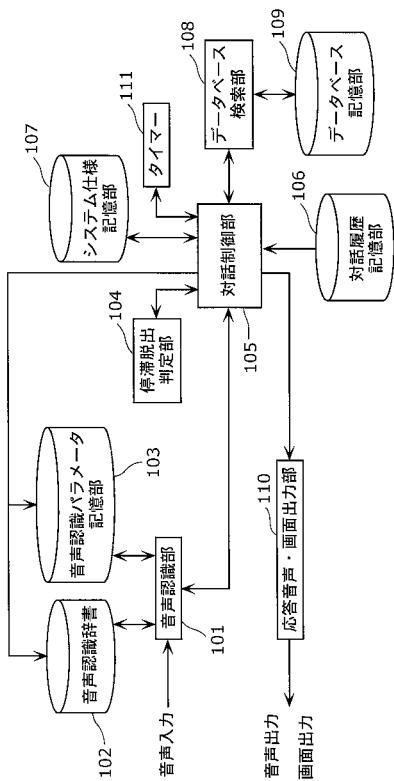
105、303、403 対話制御部

106 対話履歴記憶部

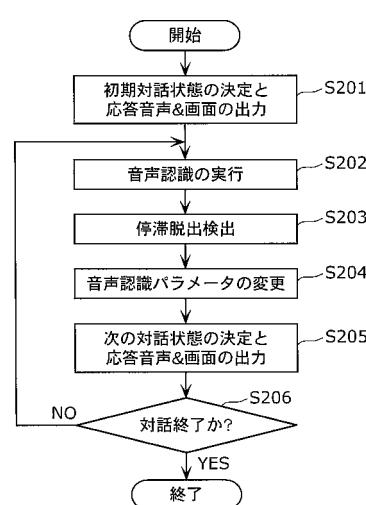
- 107 システム仕様記憶部
 108 データベース検索部
 109 データベース記憶部
 110 応答音声・画面出力部
 111 タイマー
 201 EPG受信部
 301 省略語作成部
 304 ユーザ発声記憶部
 401 ユーザ入力部
 402 ユーザ情報記憶部

10

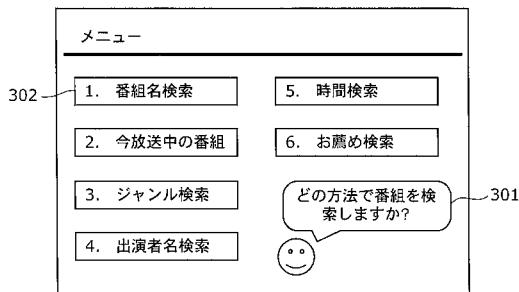
【図1】



【図2】



【図3】



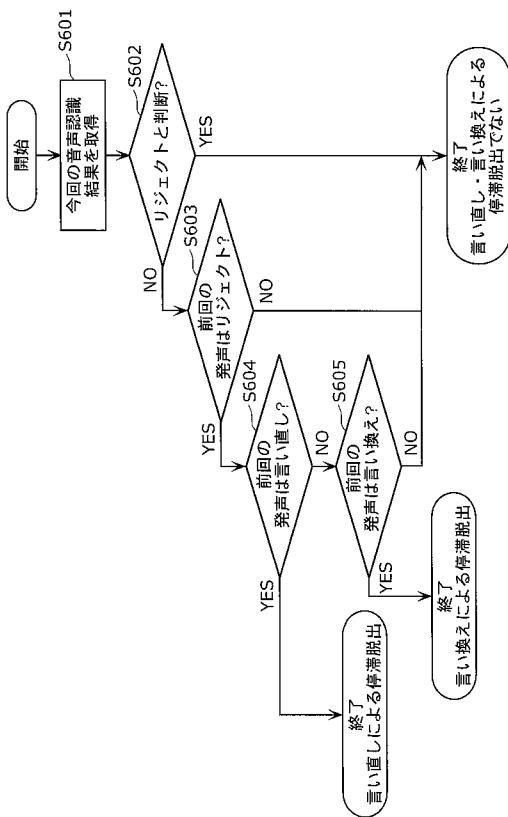
【図4】

401	認識日時	2005/03/08 10:23:48
402	音声区間	10-500
403	認識結果1	番組名検索
404	認識結果2	バンギュミミイケンサ
405	認識区間	10-380
406	認識スコア	1032
407	認識信頼度	4.5
408	リジェクト閾値	3.5

【図5】

候補順位	認識結果	認識区間	認識スコア	認識信頼度
1位	番組名検索	100-380	1032	4.5
2位	出演者名検索	50-420	650	3.8
3位	お薦め検索	150-390	326	2.3

【図6】



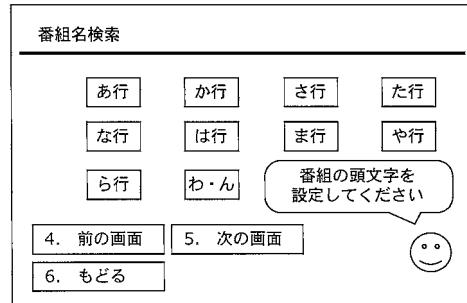
【図7】

システム状態	選択可能項目	対応認識単語	遷移先システム状態
状態1	[1. 番組名検索]	[1. 番組名検索いちばんぐみめいかんさく] [1. 番組名検索いちばんばんぐみめいかんさく] [番組名検索ばんぐみめいかんさく] [1. いちばん]	状態2
	[2. 今放送中の番組]	[2. 今放送中の番組:にいまほうそうちゅうのばんぐみ] [2. 今放送中の番組:にばんいまほうそうちゅうのばんぐみ] [今放送中の番組:にばんいまほうそうちゅうのばんぐみ] [2. ばん]	状態10
	[3. ジャンル検索]	:	状態20

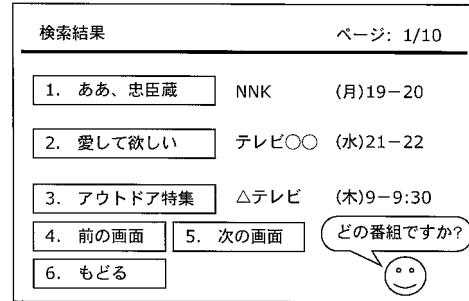
【図 8】

ステップ 状態	応答出力 開始時刻	認識結果1 番組名検索	認識頻度1 信頼度	リジエクト 判定	言い換え 停滯脱出	リジエクト 履歴
1 状態1	2004/10/08 10:23:00	番組名検索	0.47	NO	NO	NO
2 状態2	2004/10/08 10:28:00	あ行	0.36	NO	NO	NO
3 状態3	2004/10/08 10:28:30	次の画面	0.33	YES	NO	NO
4 状態3	2004/10/08 10:29:40	次の画面	0.38	NO	YES	NO
5 状態4	2004/10/08 10:30:45	次の画面	0.33	NO	NO	NO
6 状態5	2004/10/08 10:31:20	じしたい	0.50	NO	NO	NO

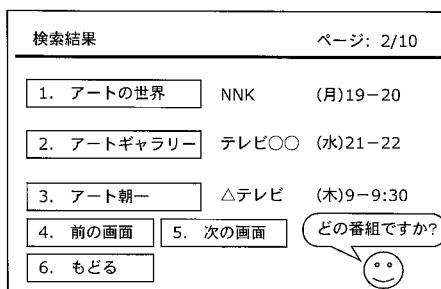
【図 9】



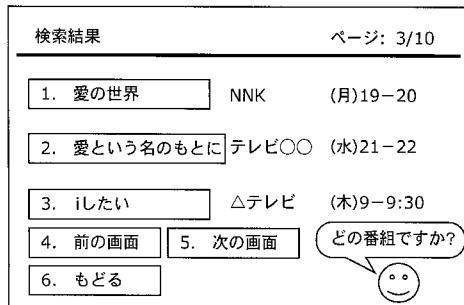
【図 10】



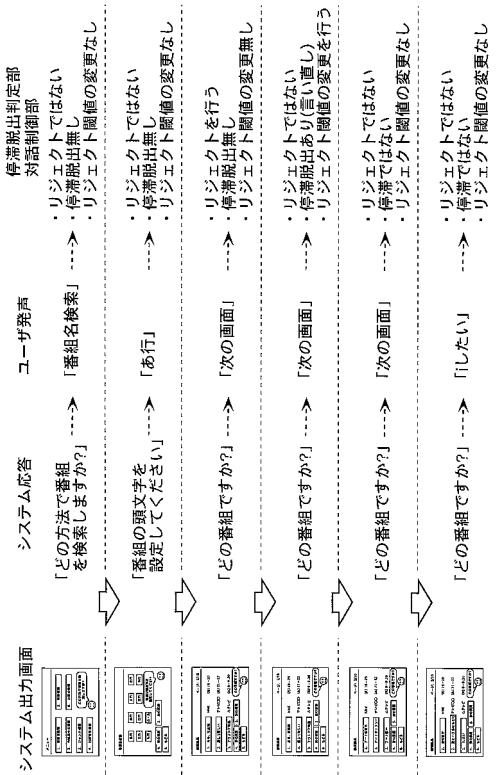
【図 11】



【図 12】



【図 13】



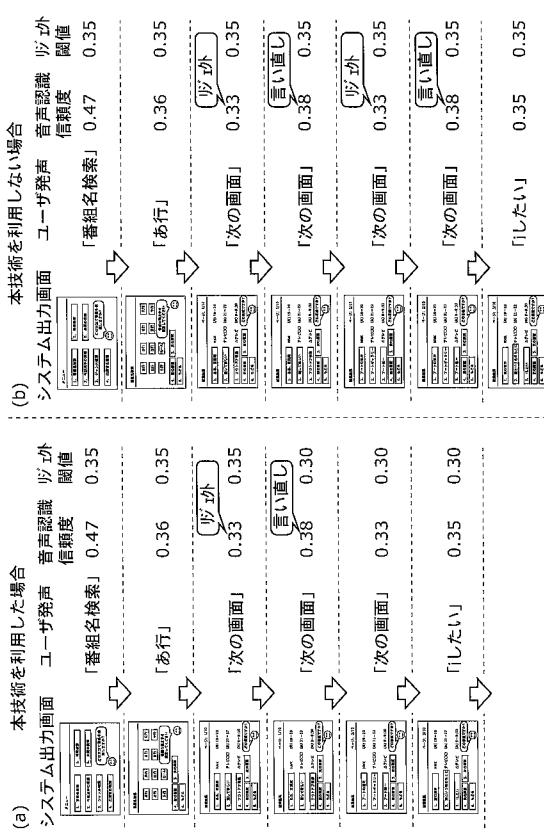
【図 1 4】

ステップ	システム状態	応答出力開始時刻	認識結果1	認識結果2	信頼度	リジェクト判定	言い直し停滯脱出閾値	リジェクト
1	状態1	2004/10/08 10:23:00	番組名検索	0.47	NO	NO	NO	0.35
2	状態2	2004/10/08 10:28:00	あ行	0.36	NO	NO	NO	0.35
3	状態3	2004/10/08 10:28:30	「次の画面」	0.33	YES	NO	NO	0.35
4	状態3	2004/10/08 10:29:40	「次の画面」	0.38	NO	NO	YES	0.35
5	状態4	2004/10/08 10:30:45	「次の画面」	0.33	NO	NO	NO	0.30
6	状態5	2004/10/08 10:31:20	「したい」	0.35	NO	NO	NO	0.30

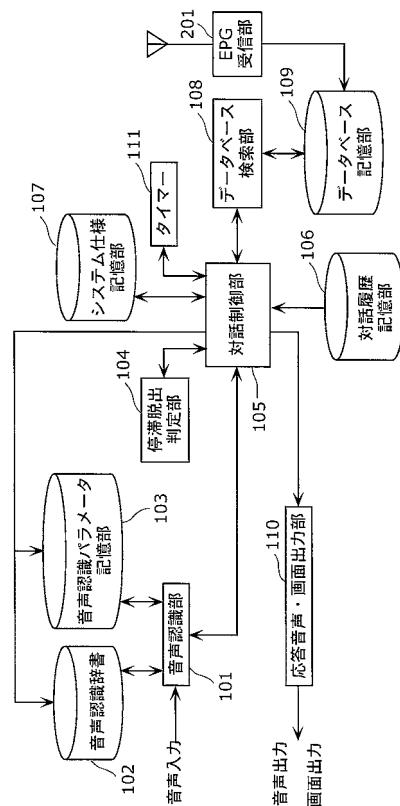
【図 1 5】

停滯単語	リジェクト閾値
次の画面	0.30
前の画面	0.33

【図 1 6】



【図 1 7】



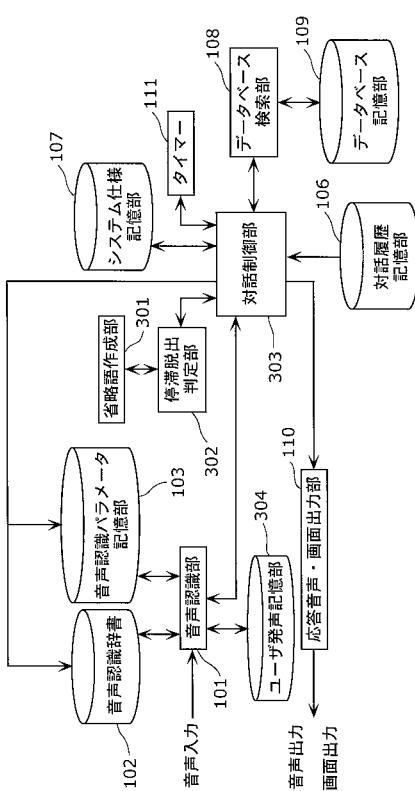
【図18】

ステップ	システム状態	認識結果1	認識結果2	認識信頼度	リジェクト判定	言い直し停滞脱出	言い換えリジェクト
1	状態1	番組名検索 バンギュミイケンサ	0.47	NO	NO	NO	0.35
2	状態2	あ行	アギュウ	0.36	NO	NO	0.35
3	状態3	次の画面	スイノダメン	0.33	YES	NO	0.35
4	状態3	次の画面	ツリノガメン	0.38	NO	YES	0.35
5	状態4	次の画面	スイノダメン	0.45	NO	NO	0.35
6	状態5	したい	アイシタイ	0.50	NO	NO	0.35

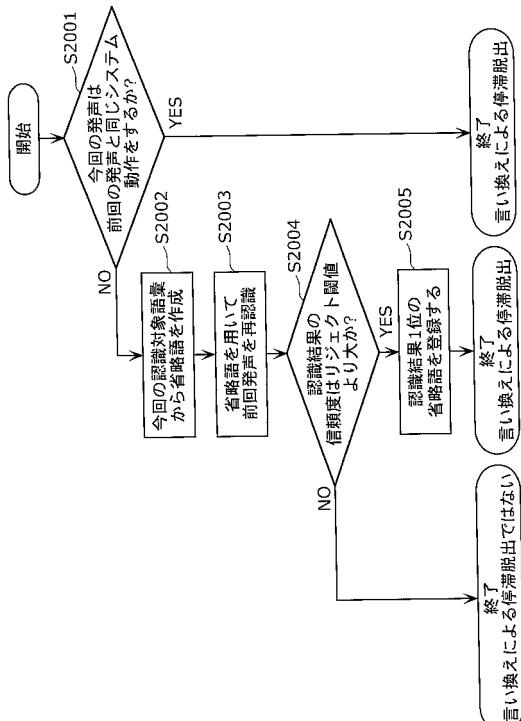
【図19】

単語番号	意味番号	表記	読み
1	1	1.	イチバン
2	1	1.	イチ
3	1	1. 番組名検索	イチバンバングミメイケンサク
4	1	1. 番組名検索	イチバングミメイケンサク
5	1	番組名検索	バングミメイケンサク
.
26	5	5.	ゴバン
27	5	5.	ゴ
28	5	5. 次の画面	ゴバンツギノガメン
29	5	5. 番組名検索	ゴツギノガメン
30	5	次の画面	ツギノガメン
130	5	次の画面	スイノダメン

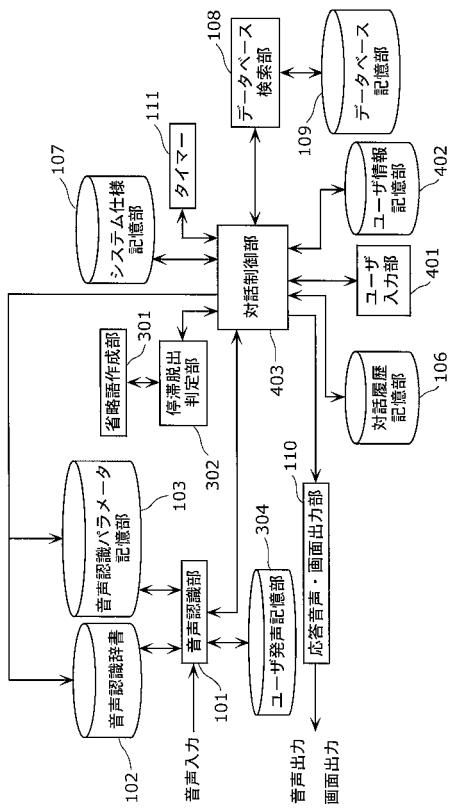
【図20】



【図21】



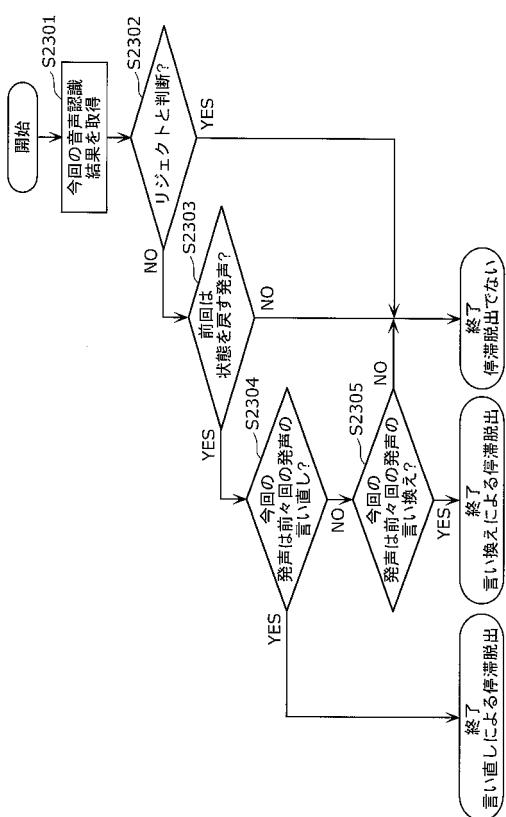
【図22】



【図23】

ユーザ	停滯単語	停滯脱出回数	リジェクト閾値	新規登録読み
A	次の画面	2	0.34	ツリノダメン
	前の画面	2	0.35	マエノダメン
B	戻る	3	0.25	モホル
C	2番	1	0.35	—
	12番	1	0.35	—

【図24】



フロントページの続き

(72)発明者 繩木 貴史

大阪府門真市大字門真1006番地 松下電器産業株式会社内

審査官 毛利 太郎

(56)参考文献 特開2001-125591(JP,A)

特開2004-109563(JP,A)

特開2003-337595(JP,A)

特開平11-149294(JP,A)

特開平11-085180(JP,A)

特開2006-154724(JP,A)

特開昭62-032500(JP,A)

特開平10-097282(JP,A)

特開平09-097095(JP,A)

国際公開第2004/044887(WO,A1)

特開2001-005489(JP,A)

特開2006-018028(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28