

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7045020号  
(P7045020)

(45)発行日 令和4年3月31日(2022. 3. 31)

(24)登録日 令和4年3月23日(2022. 3. 23)

(51)Int. Cl.		F I	
G 0 6 Q 50/10	(2012. 01)	G 0 6 Q 50/10	
G 1 0 L 15/10	(2006. 01)	G 1 0 L 15/10	5 0 0 Z
G 1 0 L 15/22	(2006. 01)	G 1 0 L 15/22	3 0 0 Z

請求項の数 6 (全 17 頁)

(21)出願番号	特願2017-37648(P2017-37648)	(73)特許権者	504157024
(22)出願日	平成29年2月28日(2017. 2. 28)		国立大学法人東北大学
(65)公開番号	特開2018-142280(P2018-142280A)		宮城県仙台市青葉区片平二丁目1番1号
(43)公開日	平成30年9月13日(2018. 9. 13)	(73)特許権者	503420833
審査請求日	令和2年2月4日(2020. 2. 4)		学校法人常翔学園
			大阪府大阪市旭区大宮五丁目16番1号
		(74)代理人	100106909
			弁理士 棚井 澄雄
		(74)代理人	100188558
			弁理士 飯田 雅人
		(74)代理人	100141139
			弁理士 及川 周
		(74)代理人	100189326
			弁理士 齋藤 崇暢

最終頁に続く

(54)【発明の名称】対話支援装置及び対話装置

(57)【特許請求の範囲】

【請求項1】

利用者と対話装置が対話し、利用者の音声と画像を用いて対話の成立又は不成立を判断し、対話不成立の場合には通信回線を介してオペレータが介入する対話装置に用いられる対話支援装置であって、

対話装置と対話する利用者の音声データを取得する音声データ取得部と、

前記利用者が撮像された画像データを取得する画像データ取得部と、

前記音声データが示す音声の特徴量である音声特徴量を取得する音声特徴量取得部と、

前記画像データが示す画像の特徴量である画像特徴量を取得する画像特徴量取得部と、

前記利用者と前記対話装置との対話が成立していない可能性を示す複数の確率値を前記音声特徴量および前記画像特徴量に基づいて取得するとともに、取得した前記複数の確率値を、前記利用者と前記対話装置との対話が成立していない可能性を示す一の確率値に統合し、統合した前記一の確率値が所定の閾値を超えた場合、前記対話への介入が必要であることをオペレータに通知する介入用通知判定部と、

前記介入用通知判定部が、前記対話装置と前記利用者との対話にオペレータを介入させる必要があるか否かを判定するために前記一の確率値と比較する前記閾値を自装置に設定する判定基準値設定部と、

を備え、

前記判定基準値設定部は、前記閾値をより高い値に設定することでオペレータの介入頻度をより低くし、前記閾値をより低い値に設定することでオペレータの介入頻度をより高

くする、

対話支援装置。

【請求項 2】

前記閾値は、オペレータの呼び出し回数の基準値に対する実際の呼び出し回数の比に対応づけられた値である、

請求項 1 に記載の対話支援装置。

【請求項 3】

利用者と対話装置が対話し、利用者の音声と画像を用いて対話の成立又は不成立を判断し、対話不成立の場合には通信回線を介してオペレータが介入する対話装置に用いられる対話支援装置であって、

10

対話装置と対話する利用者の音声データを取得する音声データ取得部と、

前記利用者が撮像された画像データを取得する画像データ取得部と、

前記音声データが示す音声の特徴量である音声特徴量を取得する音声特徴量取得部と、

前記画像データが示す画像の特徴量である画像特徴量を取得する画像特徴量取得部と、

前記対話装置と前記利用者との対話にオペレータを介入させる必要があるか否かを判定するための判定基準値を自装置に設定する判定基準値設定部と、

前記画像特徴量及び前記音声特徴量に基づいて前記利用者と前記対話装置との対話の不成立の程度を示す指標値を算出し、算出した前記指標値が前記判定基準値を超えた場合、前記対話への介入が必要であることをオペレータに通知する介入用通知判定部と、

を備え、

20

前記判定基準値設定部は、オペレータの要員計画に基づいて前記判定基準値を変更し、

前記判定基準値設定部は、前記判定基準値をより高い値又はより低い値に変更することで、オペレータの介入頻度をより低い頻度又はより高い頻度に調整する、

対話支援装置。

【請求項 4】

前記介入用通知判定部は、オペレータの介入が必要であると判定した場合、前記オペレータの音声を前記対話装置に出力させる対話制御部をさらに備える、

請求項 1 から 3 のいずれか一項に記載の対話支援装置。

【請求項 5】

利用者と対話し、利用者の音声と画像を用いて対話の成立又は不成立を判断し、対話不成立の場合には通信回線を介してオペレータが介入する対話装置であって、

30

自装置と対話する利用者の音声データを取得する音声データ取得部と、

前記利用者が撮像された画像データを取得する画像データ取得部と、

取得された前記音声データ及び前記画像データを認識し、前記利用者の発話の内容又は動作に応じた内容の音声を出力する応答部と、

前記音声データが示す音声の特徴量である音声特徴量を取得する音声特徴量取得部と、

前記画像データが示す画像の特徴量である画像特徴量を取得する画像特徴量取得部と、

前記利用者と前記対話装置との対話が成立していない可能性を示す複数の確率値を前記音声特徴量および前記画像特徴量に基づいて取得するとともに、取得した前記複数の確率値を、前記利用者と前記対話装置との対話が成立していない可能性を示す一の確率値に統合し、統合した前記一の確率値が所定の閾値を超えた場合、前記対話への介入が必要であることをオペレータに通知する介入用通知判定部と、

40

前記介入用通知判定部が、前記対話装置と前記利用者との対話にオペレータを介入させる必要があるか否かを判定するために前記一の確率値と比較する前記閾値を自装置に設定する判定基準値設定部と、

を備え、

前記判定基準値設定部は、前記閾値をより高い値に設定することでオペレータの介入頻度をより低くし、前記閾値をより低い値に設定することでオペレータの介入頻度をより高くする、

対話装置。

50

**【請求項 6】**

利用者と対話し、利用者の音声と画像を用いて対話の成立又は不成立を判断し、対話不成立の場合には通信回線を介してオペレータが介入する対話装置であって、

自装置と対話する利用者の音声データを取得する音声データ取得部と、

前記利用者が撮像された画像データを取得する画像データ取得部と、

取得された前記音声データ及び前記画像データを認識し、前記利用者の発話の内容又は動作に応じた内容の音声を出力する応答部と、

前記音声データが示す音声の特徴量である音声特徴量を取得する音声特徴量取得部と、

前記画像データが示す画像の特徴量である画像特徴量を取得する画像特徴量取得部と、

前記対話装置と前記利用者との対話にオペレータを介入させる必要があるか否かを判定するための判定基準値を自装置に設定する判定基準値設定部と、

前記画像特徴量及び前記音声特徴量に基づいて前記利用者と自装置との対話の不成立の程度を示す指標値を算出し、算出した前記指標値が前記判定基準値を超えた場合、前記対話への介入が必要であることをオペレータに通知する介入用通知判定部と、

を備え、

前記判定基準値設定部は、オペレータの要員計画に基づいて前記判定基準値を変更し、

前記判定基準値設定部は、前記判定基準値をより高い値又はより低い値に変更することで、オペレータの介入頻度をより低い頻度又はより高い頻度に調整する、

対話装置。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、対話支援装置及び対話装置に関する。

**【背景技術】****【0002】**

近年、利用者の発話に応じた音声を出力可能な対話装置を利用して、利用者に様々なサービスを提供する試みがなされている。しかしながら、対話装置のみによる人との音声対話には限界があり、対話装置が完全に自立して利用者と対話することは難しい。そこで、利用者と対話装置との間の対話に生じている問題を検出し、対話装置と利用者との間にオペレータを介在させる技術が提案されている（例えば特許文献1参照）。

**【先行技術文献】****【特許文献】****【0003】**

【特許文献1】特開2007-190641号公報

**【発明の概要】****【発明が解決しようとする課題】****【0004】**

しかしながら、従来技術は、対話装置と利用者との対話にオペレータを介入させるか否かを、対話システムによって対話が成立していない（対話不成立）と判断された場合に介入させるものである。オペレータは通常、複数の業務や複数台の対話装置等を担当し、また、対話システムの使用状況とは無関係の業務をしていることも多い。利用者と対話装置との対話が成立していないという問題が生じた場合、介入指示がオペレータに通知される。しかし、介入指示が通知された場合であっても、実際の対話状況は、介入が必須である場合もあれば、介入することが望ましいが必ずしも介入が必須ではない場合もある。そのため、オペレータの介入頻度は、対話システムの用途、日時等における繁閑や業務時間におけるオペレータの他業務とのバランス等を考慮して適切に調整されることが望ましい。

**【0005】**

しかし従来技術では対話ロボットの実使用場面における用途やオペレータの要員計画等、システム運営者側の諸事情を考慮して判断するものではなく、そのため、対話装置と利用者との対話にオペレータを、オペレータ等の事情を考慮して状況に応じて適宜介入させ

ることができないという問題があった。また、従来は、オペレータを介入させるか否かの判定を利用者の音声や画像等の特徴ごとに介入判定を行っていたことから、対話不成立を判定する精度も悪く、発話前の状態で対話が困難である場合の判定も難しく、介入頻度を適切に変更し、適切にオペレータを介入させることが困難であった。

【 0 0 0 6 】

上記事情に鑑み、本発明は、対話ロボットと利用者との対話にオペレータを状況に応じて適宜介入させることができる技術を提供することを目的としている。

【 課題を解決するための手段 】

【 0 0 0 7 】

本発明の一態様は、利用者と対話装置が対話し、利用者の音声と画像を用いて対話の成立又は不成立を判断し、対話不成立の場合には通信回線を介してオペレータが介入する対話装置に用いられる対話支援装置であって、対話装置と対話する利用者の音声データを取得する音声データ取得部と、前記利用者が撮像された画像データを取得する画像データ取得部と、前記音声データが示す音声の特徴量である音声特徴量を取得する音声特徴量取得部と、前記画像データが示す画像の特徴量である画像特徴量を取得する画像特徴量取得部と、前記対話装置と前記利用者との対話にオペレータを介入させる必要があるか否かを判定するための判定基準値を自装置に設定する判定基準値設定部と、前記画像特徴量及び前記音声特徴量に基づいて前記利用者と前記対話装置との対話の不成立の程度を示す指標値を算出し、算出した前記指標値が前記判定基準値を超えた場合、前記対話への介入が必要であることをオペレータに通知する介入用通知判定部と、を備え、前記判定基準値設定部は、前記判定基準値をより高い値又はより低い値に変更することで、オペレータの介入頻度をより低い頻度又はより高い頻度に調整する、対話支援装置である。

【 0 0 0 8 】

本発明の一態様は上記の対話支援装置であって、前記判定基準値設定部は、オペレータの要員計画に基づいて前記判定基準値を変更する。

【 0 0 0 9 】

本発明の一態様は上記の対話支援装置であって、前記介入用通知判定部は、オペレータの介入が必要であると判定した場合、前記オペレータの音声を前記対話装置に出力させる対話制御部をさらに備える。

【 0 0 1 0 】

本発明の一態様は上記の対話支援装置であって、前記介入用通知判定部は、前記画像特徴量及び前記音声特徴量の各入力に対して前記利用者が前記対話装置と円滑に対話できているか否かを示す第1の指標値を取得し、前記各入力に対して取得した第1の指標値に基づいて前記対話装置と前記利用者との対話にオペレータを介入させるか否かを判定するための第2の指標値を取得し、取得した前記第2の指標値と前記判定基準値との大小関係に基づいて、前記対話にオペレータを介入させるか否かを判定する。

【 0 0 1 1 】

本発明の一態様は、利用者と対話し、利用者の音声と画像を用いて対話の成立又は不成立を判断し、対話不成立の場合には通信回線を介してオペレータが介入する対話装置であって、自装置と対話する利用者の音声データを取得する音声データ取得部と、前記利用者が撮像された画像データを取得する画像データ取得部と、取得された前記音声データ及び前記画像データを認識し、前記利用者の発話の内容又は動作に応じた音声を出力する応答部と、前記音声データが示す音声の特徴量である音声特徴量を取得する音声特徴量取得部と、前記画像データが示す画像の特徴量である画像特徴量を取得する画像特徴量取得部と、前記対話装置と前記利用者との対話にオペレータを介入させる必要があるか否かを判定するための判定基準値を自装置に設定する判定基準値設定部と、前記画像特徴量及び前記音声特徴量に基づいて前記利用者と自装置との対話の不成立の程度を示す指標値を算出し、算出した前記指標値が前記判定基準値を超えた場合、前記対話への介入が必要であることをオペレータに通知する介入用通知判定部と、を備え、前記判定基準値設定部は、前記判定基準値をより高い値又はより低い値に変更することで、オペレータの介入頻度をより

低い頻度又はより高い頻度に調整する、対話装置である。

【発明の効果】

【0012】

本発明により、対話ロボットと利用者との対話にオペレータを状況に応じて適宜介入させることが可能となる。

【図面の簡単な説明】

【0013】

【図1】第1実施形態の対話システム100の構成の概略を示す図である。

【図2】対話システム100の構成の他の具体例を示す概略図である。

【図3】第1実施形態の対話支援装置1の機能構成の具体例を示すブロック図である。

10

【図4】第1実施形態の対話支援装置1による介入判定処理の流れを示すフローチャートである。

【図5】第1実施形態の対話システムにおける判定基準値と呼び出し回数比との関係の具体例を示す図である。

【図6】第2実施形態の対話支援装置1aの機能構成の具体例を示すブロック図である。

【図7】第2実施形態における制約情報及び条件情報の具体例を示す図である。

【図8】対話支援装置1と一体に構成された対話ロボット2aの機能構成の具体例を示す図である。

【発明を実施するための形態】

【0014】

20

< 第1実施形態 >

図1は、第1実施形態の対話システム100の構成の概略を示す図である。対話システム100は、対話支援装置1、対話ロボット2及びオペレータ端末3を備える。対話支援装置1は、対話ロボット2と利用者との対話が円滑に進むように、現在の対話の状況を認識し、必要な支援を行う装置である。対話の状況の認識には、対話ロボット2の持つ内部状態、利用者の発する音声、利用者の画像およびその他のセンサーデータを利用する。対話支援装置が行う支援としては、現在の対話状況を対話ロボット2に送信して対話の流れを制御することや、対話が破綻したときに外部のオペレータに連絡をして利用者への対応を要請することなどが挙げられる。

【0015】

30

対話ロボット2は本発明における対話装置の一例である。対話装置とは、利用者と音声を主とした対話を行うことによって、利用者からの情報収集および利用者への情報提供を行う装置である。音声だけでなく画像やその他のセンサ情報を利用する装置であっても良い。対話ロボット2は、利用者と対話できるロボットであって、利用者の発話音声と画像を入力認識し応答するものをいう。対話ロボット2の形態は人型ロボットの形状に限られず情報端末の形態であってもよい。対話ロボット2は、利用者と音声を主とした対話を行うことによって、利用者からの情報収集及び利用者への情報提供を行う装置であれば他のどのような情報を用いるものであってもよい。例えば、音声だけでは対話の成立、不成立を精度良く判定することができない場合、音声データに加えて対話中の利用者の画像データを用いるものであってもよい。この場合、画像データは対話ロボット2に備わったカメラ等の撮像手段によって取得されてもよいし、別途設けられた撮像手段によって取得されてもよい。一般には、対話ロボット2等の対話装置が撮像手段を備えていることが望ましい。

40

【0016】

対話支援装置1、対話ロボット2及びオペレータ端末3は、通信回線4を介して互いに通信可能である。対話支援装置1は、対話ロボット2と利用者との対話が円滑に進むように、現在の対話の状況を認識し、必要な支援を行う装置である。対話の状況の認識には、対話ロボット2の持つ内部状態、利用者の発する音声、利用者の画像およびその他のセンサーデータを利用する。対話支援装置2が行う支援としては、現在の対話状況を対話ロボット2に送信して対話の流れを制御することや、対話が破綻したときに外部のオペレータ

50

に連絡をして利用者への対応を要請することなどが挙げられる。

【 0 0 1 7 】

対話支援装置 1 は、利用者とロボットとの対話を監視し、必要に応じて利用者の対話ロボット 2 との対話を、オペレータとの対話に切り替える。対話ロボット 2 は、入力された音声を示す内容に応じた音声を出力する対話装置である。対話ロボット 2 は、利用者との対話を行うほか、自装置に対して発せられた利用者の音声を、通信回線 4 を介して対話支援装置 1 に送信する。オペレータ端末 3 は、対話支援装置 1 の制御に基づき対話ロボット 2 に接続され、オペレータの音声を対話ロボット 2 から出力させる機能を有する。オペレータ端末 3 は、対話が不成立でオペレータの介入が必要と判定される場合にオペレータに通知し、オペレータは必要に応じて対話ロボットを介して状況を手入力し介入すべきと判断すれば対話ロボットに代わり利用者との対話する。

10

【 0 0 1 8 】

図 2 は、対話システム 1 0 0 の構成の他の具体例を示す概略図である。対話ロボット 2 は、対話支援装置を含んでも、含まなくともよいが、対話支援装置 1 の機能を含んでいると通信回線の通信量を軽減することができ、好ましい。図 1 が、対話ロボット 2、対話支援装置 1 が各々通信回線を介して接続された一例を示したのに対して、図 2 ( A ) は対話支援装置 1 を介して対話ロボット 2 と通信回線が接続された一例を示す。また、図 2 ( B ) は、対話支援装置 1 が対話ロボット 2 に含まれる一例を示す。

【 0 0 1 9 】

図 3 は、第 1 実施形態の対話支援装置 1 の機能構成の具体例を示すブロック図である。対話支援装置 1 は、バスで接続された CPU ( Central Processing Unit ) やメモリや補助記憶装置などを備え、プログラムを実行する。対話支援装置 1 は、プログラムの実行によって通信部 1 0 1、画像データ取得部 1 0 2、音声データ取得部 1 0 3、画像特徴量取得部 1 0 4、発話区間識別部 1 0 5、音声特徴量取得部 1 0 6、介入用通知判定部 1 0 9 及び対話制御部 1 1 0 を備える装置として機能する。なお、対話支援装置 1 の各機能の全て又は一部は、ASIC ( Application Specific Integrated Circuit ) や PLD ( Programmable Logic Device ) や FPGA ( Field Programmable Gate Array ) 等のハードウェアを用いて実現されてもよい。プログラムは、コンピュータ読み取り可能な記録媒体に記録されてもよい。コンピュータ読み取り可能な記録媒体とは、例えばフレキシブルディスク、光磁気ディスク、ROM、CD-ROM 等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置である。プログラムは、電気通信回線を介して送信されてもよい。

20

30

【 0 0 2 0 】

通信部 1 0 1 は、自装置を通信回線 4 に接続する通信インタフェースを備えて構成される。通信部 1 0 1 は、通信回線 4 を介して対話ロボット 2 及びオペレータ端末 3 と通信する。

【 0 0 2 1 】

画像データ取得部 1 0 2 は、対話ロボット 2 と対話中の利用者が撮像された画像データを取得する ( 図 4 のステップ S 1 0 1 に対応 ) 。例えば、利用者を撮像する撮像部 ( 図示せず ) は対話ロボット 2 に備えられる。この場合、画像データ取得部 1 0 2 は、対話ロボット 2 との通信により画像データを取得する。また、撮像部は、対話中の利用者を撮像可能な位置に設置された撮像装置であってもよい。この場合、撮像装置は通信回線 4 に接続され、画像データ取得部 1 0 2 は、撮像装置との通信により画像データを取得してもよい。

40

【 0 0 2 2 】

音声データ取得部 1 0 3 は、対話ロボット 2 に対して発せられた利用者の音声データを取得する ( 図 4 のステップ S 1 0 4 に対応 ) 。例えば、利用者の音声を入力する音声入力部 ( 図示せず ) は対話ロボット 2 に備えられる。この場合、音声データ取得部 1 0 3 は、対話ロボット 2 との通信により音声データを取得する。また、音声入力部は、対話中の利用者の音声を取得可能な位置に設置された音声入力装置であってもよい。この場合、音声

50

入力装置は通信回線 4 に接続され、音声データ取得部 103 は、音声入力装置との通信により画像データを取得してもよい。

【0023】

画像特徴量取得部 104 は、画像データ取得部 102 によって取得された画像データが示す画像の特徴量（以下「画像特徴量」という。）を取得する（図 4 のステップ S102 に対応）。第 1 実施形態における画像特徴量は、利用者の挙動に関する特徴量である。具体的には、画像特徴量は、利用者の体全体が単位時間あたりに移動した量や、利用者の顔及び視線の向き、及びそれらの時間変化に関する特徴量を含む。画像特徴量取得部 104 は、取得した画像特徴量を示す情報を介入用通知判定部 109 に出力する。

【0024】

発話区間識別部 105 は、音声データの周波数解析等により、音声データの示す音声区間から利用者の発話が含まれる区間（以下「発話区間」という。）を識別する（図 4 のステップ S105 に対応）。発話区間識別部 105 は、利用者の発話区間を示す情報を音声特徴量取得部 106 に出力する。

【0025】

音声特徴量取得部 106 は、音声データ取得部 103 によって取得された音声データが示す音声の特徴量（以下「音声特徴量」という。）を取得する（図 4 のステップ S107 に対応）。第 1 実施形態における音声特徴量は、利用者の発話タイミングに関する特徴量及び音声の周波数に関する特徴量である。具体的には、音声特徴量は、対話ロボット 2 が利用者に対して発話を促してから、実際に利用者の発話が始まるまでに要した時間や有聲休止（「あー」「えーと」など、同じ母音が引き延ばされることを特徴とする無意味発話）等に関する特徴量を含む。音声特徴量取得部 106 は、発話区間識別部 105 によって識別された発話区間の情報に基づいて利用者の発話タイミングに関する音声特徴量を取得する。また、音声特徴量取得部 106 は、音声データに基づいて周波数解析を行うことにより音声の周波数に関する特徴量を取得する。音声特徴量取得部 106 は、利用者の発話タイミングに関する特徴量及び音声の周波数に関する特徴量を音声特徴量として介入用通知判定部 109 に出力する。

【0026】

記憶部 107 は、磁気ハードディスク装置や半導体記憶装置などの記憶装置を用いて構成される。記憶部 107 は判定基準値情報を記憶する。判定基準値情報は、対話ロボット 2 と利用者との対話にオペレータを介入させるか否かの判定するための判定基準値を示す情報である。判定基準値情報は、判定基準値設定部 108 によって記憶部 107 に記憶される。

【0027】

判定基準値設定部 108 は、対話ロボット 2 と利用者との対話にオペレータを介入させるか否かを判定するために必要となる判定基準値を自装置に設定する機能を有する。具体的には、判定基準値の設定とは、判定基準値情報を記憶部 107 に記憶させることを意味する。記憶部 107 に記憶させる判定基準値情報はどのような方法で取得されてもよい。判定基準値及びその適用範囲等の判定基準情報の入力者としては、例えば、システム運用者又はオペレータ、あるいは AI（Artificial Intelligence：人工知能）等が挙げられる。例えば、判定基準値設定部 108 は、マウスやキーボード等の入力装置を介して判定基準値情報を取得してもよいし、通信部 101 を介した通信によって他の装置から判定基準値情報を取得してもよい。また、例えば、判定基準値設定部 108 は、種々の情報を用いて判定基準値情報を生成してもよいし、判定基準値となりうる複数の判定基準値情報の中から用いられるべき判定基準値情報を選択してもよい。

【0028】

また、対話ロボット 2 が複数存在する場合、判定基準値設定部 108 は、複数の対話ロボット 2 ごとの判定基準値を設定してもよいし、複数の対話ロボット 2 を分類するグループごとに判定基準値を設定してもよい。

【0029】

10

20

30

40

50

介入用通知判定部109は、対話ロボット2と利用者との対話にオペレータを介入させるか否かを判定する(図4のステップS103、S106、S108~S112に対応)。以下、この判定を介入判定という。具体的には、介入用通知判定部109は、画像特徴量取得部104によって取得された画像特徴量と、音声特徴量取得部106によって取得された音声特徴量と、判定基準値設定部108によって設定された判定基準値と、に基づいて介入判定を行う。介入用通知判定部109は、介入判定の判定結果を対話制御部110に出力する。

【0030】

対話制御部110は、介入用通知判定部109によって行われた介入判定の判定結果に基づいて対話ロボット2の音声出力機能を制御する。

10

【0031】

図4は、第1実施形態の対話支援装置1による介入判定処理の流れを示すフローチャートである。まず、画像データ取得部102が、対話ロボット2と対話中の利用者が撮像された画像データを取得する(ステップS101)。画像データ取得部102は、取得した画像データを画像特徴量取得部104に出力する。画像特徴量取得部104は、画像データ取得部102から画像データを取得する。画像特徴量取得部104は、取得した画像データの画像特徴量を取得する(ステップS102)。画像特徴量取得部104は、取得した画像特徴量を、利用者が対話ロボット2と円滑に対話できているか否かを識別する第1の識別器(以下「第1識別器」という。)に入力する。これにより、画像特徴量取得部104は、第1識別器の出力として、例えば、利用者が対話ロボット2と円滑に対話できていない、すなわち対話が成立していない対話不成立の確率 $p_1$ を取得する(ステップS103)。

20

【0032】

一方、画像データの取得と並行して、音声データ取得部103が、対話ロボット2と対話中の利用者の音声データを取得する(ステップS104)。音声データ取得部103は、取得した音声データを発話区間識別部105及び音声特徴量取得部106に出力する。発話区間識別部105は、音声データ取得部103から音声データを取得する。発話区間識別部105は、取得した音声データに基づいて利用者の発話区間を識別する(ステップS105)。発話区間識別部105は、識別した発話区間を示す情報を音声特徴量取得部106に出力する。

30

【0033】

音声特徴量取得部106は、発話区間識別部105から、利用者の発話区間を示す情報を取得する。音声特徴量取得部106は、利用者の発話区間に関する特徴量を、利用者が対話ロボット2と円滑に対話できているか否かを識別する第2の識別器(以下「第2識別器」という。)に入力する。これにより、音声特徴量取得部106は、第2識別器の出力として、例えば、利用者が対話ロボット2と円滑に対話できていない、すなわち対話が成立していない対話不成立の確率 $p_2$ を取得する(ステップS106)。ここでいう発話区間に関する特徴量は、例えば利用者のフィルタ音声又は言い淀みに関する特徴量である。

【0034】

一方、音声特徴量取得部106は、音声データ取得部103から音声データを取得する。音声特徴量取得部106は、取得した音声データの音声特徴量を取得する(ステップS107)。例えば、音声特徴量は、音声波形の解析によって取得可能な音声の特徴量(声の大きさや高さ、速さ等)である。その意味では、上記の発話区間に関する特徴量も音声特徴量の一つに含まれても良い。音声特徴量取得部106は、取得した音声特徴量を、利用者が対話ロボット2と円滑に対話できているか否かを識別する第3の識別器(以下「第3識別器」という。)に入力する。これにより、音声特徴量取得部106は、第3識別器の出力として、例えば、利用者が対話ロボット2と円滑に対話できていない、すなわち対話が成立していない対話不成立の確率 $p_3$ を取得する(ステップS108)。

40

【0035】

なお、上述した各識別器は、サンプルデータの特徴量を機械学習することによって生成

50



される。各識別器は、サポートベクターマシンやニューラルネットワークなどの機械学習手法を用いて生成することができる。画像特徴量取得部104及び音声特徴量取得部106は、予め生成された識別器を有してもよいし、サンプルデータの特徴量を機械学習することにより各識別器を生成する機能を有しても良い。また、画像特徴量取得部104及び音声特徴量取得部106は、生成された各識別器を、新たに得られたサンプルデータの特徴量に基づいて更新する機能を有しても良い。

#### 【0036】

続いて、介入用通知判定部109が、確率 $p_1$ 、 $p_2$ 及び $p_3$ に基づいて、利用者が対話ロボット2と円滑に対話できているか否かを最終的に判定する。具体的には、介入用通知判定部109は、機械学習に基づく確率統合の手法（例えば以下の参考文献1を参照。）を用いて確率 $p_1$ 、 $p_2$ 及び $p_3$ を統合し、利用者が対話ロボット2と円滑に対話できていない確率 $p$ を算出する（ステップS109）。

参考文献1：千葉他「対話中のユーザ状態逐次推定のための多段階識別手法に関する検討」、情報処理学会研究報告 Vol.2013 No.21 16

#### 【0037】

介入用通知判定部109は、算出した確率 $p$ （利用者が対話ロボット2と円滑に対話できていない確率）を判定基準値（閾値）と比較する（ステップS110）。 $p$ が以上である場合（ステップS110 - YES）、介入用通知判定部109は、利用者と対話ロボット2との対話にオペレータの介入用通知が必要と判定する（ステップS111）。一方、 $p$ が未満である場合（ステップS110 - NO）、介入用通知判定部109は、オペレータへの介入用通知は不要と判定する（ステップS112）。通知されたオペレータは通常ただちに介入するが、状況により判断することも可能である。

#### 【0038】

このような介入判定処理によれば、対話ロボット2の問いかけから利用者の発話が発生するまでの時間、利用者による発話の休止、利用者の顔の向きや視線の方向、顔の動き等の単位時間あたりの量の観測することにより、利用者が対話ロボット2と円滑に対話できているか否かを判定することができる。例えば、対話が困難である場合、発話までの時間が長くなる（概ね5秒以上）傾向がある。また、この場合、発話の休止頻度が高い、視線が中心を外れる、首を傾げる動作が見られるなどの傾向がある。これらの特徴はそれぞれが独立して観測されるものではなく、互いに何らかの相関を有すると考えられる。そのため、利用者と対話ロボット2とが円滑に対話できているか否かを、これらの各特徴のそれぞれに判定基準値を設けて判定することは必ずしも適切でない。

#### 【0039】

例えば、これらの特徴ごとに独立して介入判定を行った場合、必要以上に頻繁にオペレータが呼び出されたり、呼び出すべきタイミングが適切でなかったりといった問題が生じる可能性がある。実施形態の対話支援装置1は、音声特徴量と画像特徴量と発話区間（間合い）のそれぞれに基づいて対話に問題が生じている確率を算出し、これらの特徴ごとに算出された確率を、更に機械学習に基づく手法で統合する構成を備えることにより、問題が生じている確率をより精度よく算出することが可能となる。音声、画像等の個別要素のみで対話不成立を判定する方法に比べて、個別要素を統合して対話不成立の確率を算出するため、判定精度が向上し、オペレータの呼び出しの要否がより精度良く判定されることにつながる。

#### 【0040】

このように構成された第1実施形態の対話支援装置1は、対話ロボットと利用者との対話にオペレータを効率良く介入させることが可能となる。具体的には、対話支援装置1は、判定基準値情報を取得又は生成して自装置に設定する判定基準値設定部108を備える。この判定基準値設定部108を備えることにより、対話支援装置1は、システム運営者側の諸事情を考慮した上で、対話ロボットと利用者との対話にオペレータを導入させるか否かを判定することができる。オペレータの介入が必要と判断した場合、対話制御部110オペレータの音声に対話装置2に出力させる。

10

20

30

40

50

## 【 0 0 4 1 】

図5は、第1実施形態の対話システムにおける判定基準値と呼び出し回数比との関係の具体例を示す図である。図5の横軸は判定基準値としての閾値を表し、縦軸は呼び出し回数比を表す。呼び出し回数比は、オペレータの呼び出し回数の基準値に対する比を表す。ここでは、 $P$ が0.6のときの呼び出し回数比を基準(1.0)としている。下記の式(3)において、適合率 $P$ が0.6のとき再現率 $R$ も約0.6となるため、ここでは $P = 0.6$ を基準とした。

## 【 0 0 4 2 】

図5に示すグラフは次のようにして求められた。まず、対話システムを利用して実際に対話を行った際に、オペレータの介入を必要とするかどうかについて、さまざまな判断基準値で判定を行った。この時に、実際に介入が必要であった回数を $N$ 、対話システムが、介入が必要と判定した回数のうち実際に介入が必要だった回数を $N_{tp}$ 、対話システムが、介入が必要と判定した回数のうち実際は介入が不要であった回数を $N_{fp}$ とすると、この対話システムの適合率 $P$ (対話システムによる介入判定の的中率)は以下の式(1)で求められる。また、再現率(実際に介入が必要な状況を検出した確率)は、以下の式(2)で求められる。

## 【 0 0 4 3 】

$$P = N_{tp} / (N_{tp} + N_{fp}) \quad \text{式(1)}$$

## 【 0 0 4 4 】

$$R = N_{tp} / N \quad \text{式(2)}$$

## 【 0 0 4 5 】

実際の検出結果を分析した結果、 $P$ と $R$ の間には相関があり、その相関は大凡次の式(3)のように表すことができる。

## 【 0 0 4 6 】

$$R(P) = \min(1.0, 1.1 - 0.9P) \quad \text{式(3)}$$

## 【 0 0 4 7 】

そのため、 $N$ をある値(例えば100回)としたときに、対話システムを使って実際に検出される回数は次の式(4)のように表すことができる。

## 【 0 0 4 8 】

$$N_{tp} + N_{fp} = N \times R / P \quad \text{式(4)}$$

## 【 0 0 4 9 】

ここで、 $P$ を介入判定の指標値として利用すれば、介入が必要な状況100回当たりの呼び出し回数比は次の式(5)で求めることができる。閾値は、判定基準値である。

## 【 0 0 5 0 】

$$100 \times \min(1 / P, 1.1 / P - 0.9) \quad \text{式(5)}$$

## 【 0 0 5 1 】

例えば、第1実施形態の対話システムがイベント会場の案内に用いられる場合を想定する。この場合、利用者が対話システムを利用するために提供される対話ロボット等の利用者端末は、イベント会場のエントランス等の場所に設置されることが多い。このような場所に十分なスペースが確保されていない場合、時間帯によっては対話ロボット周辺に多くの人が滞留してしまい、会場全体の混雑を招く可能性がある。このような状況が想定される用途に対して、従来技術では、オペレータを介入させるか否かの判定基準値を柔軟に変更することができないため、オペレータに対して現場の混雑の状況に応じた効率の良い介入を行わせることができず、対話システム自身が会場の人の流れを阻害する要因となってしまう可能性があった。これに対して、第1実施形態の対話システムによれば、例えば、イベント運営者は、会場の混雑状況に応じて判定基準値を柔軟に変更することができる。これにより、混雑時にはオペレータの介入頻度を向上させ、対話ロボットが利用者と円滑に対話することができなくなる状況が発生することを未然に防止し、会場が混雑することを抑制することが可能になる。

## 【 0 0 5 2 】

10

20

30

40

50

例えば図5の例の場合、通常時の判定基準値（閾値）を0.6（最大値1.0）に設定して運用しているところ、当該時間帯のみ、システム運用者又はオペレータが対話不成立の確率に対する閾値を0.4に引き下げることで、オペレータの呼び出し頻度を容易に約2倍に上げることができる。この場合、対話不成立が生じていないときであってもオペレータが呼び出される確率が高くはなるが、オペレータの対応頻度を多くすることによって、積極的に案内することで対話不成立を未然に防ぎ、被案内人が対話トラブルにより対話システム前に長く居続けないようにすることで混雑を防ぐことができる。

【0053】

また、イベント会場のオペレータは、開催されるイベントによっては会場内の安全チェック等の別業務を兼任する場合もある。このような場合において、従来技術では、オペレータを介入させるか否かの判定基準値を柔軟に変更することができないため、オペレータの介入頻度が高い状況では、オペレータが安全チェック等の別業務を十分に行うことができない可能性があった。これに対して、第1実施形態の対話システムによれば、例えば、イベント運営者は、オペレータに課されるその日の業務内容に応じて、オペレータの介入頻度を下げよう判定基準値を設定することができる。これにより、オペレータは、安全チェック等の別業務の遂行に注力することが可能になる。

【0054】

例えば図5の例の場合、通常時の判定基準値（閾値）を0.6（最大値1.0）に設定して運用しているところ、当該イベント当日のみ、システム運用者又はオペレータが対話不成立の確率に対する閾値を0.8に引き上げることで、オペレータの呼び出し頻度を約2分の1に容易に下げることができる。この場合、対話不成立が生じてもオペレータが呼び出されなくなる確率が高くはなるが、オペレータがより重要な安全チェック等の業務に注力できるようにすることができる。

【0055】

また、例えば、一人のオペレータが、複数のイベント会場の利用者の対応を行わなければならない場合も想定される。この場合、利用者との対話ロボットとの対話の円滑性がイベント会場ごとに異なる場合もある。例えば、一部のイベントが高齢者向けのイベントである場合には、他のイベントよりもオペレータの介入頻度が高まることが想定される。また、高齢の利用者に対する対応は、若年の利用者に対する対応よりも時間がかかることも想定される。このような場合、従来技術では、オペレータを介入させるか否かの判定基準値を柔軟に変更することができないため、複数のイベント会場においてオペレータ介入の必要性が同じ頻度で発生することになる。その結果、オペレータが、高齢の利用者に対する対応の負荷によって、若年の利用者に対する対応を十分に行えなくなる状況が発生しうる。さらに、高齢の利用者に対する対応と、若年の利用者に対する対応とが同時に発生した場合、両者に対する対応が不十分なものになってしまう可能性がある。これに対して、第1実施形態の対話システムによれば、例えば、イベント運営者は、複数のイベント会場の対話ロボットに対して、イベントや利用者の傾向等に応じた判定基準値を設定することができる。これにより、オペレータは、高齢の利用者に対する対応することが可能になる。

【0056】

例えば図5の例の場合、2つの対話ロボットA及びBについて、通常時の判定基準値（閾値）を0.6（最大値1.0）に設定して運用しているところ、当該イベント当日のみ、システム運用者又はオペレータが対話ロボットA及びBの対話不成立の確率に対する閾値を0.8に引き上げることで、対話ロボットA及びBによるオペレータの呼び出し頻度を約2分の1に容易に下げることができる。この場合、対話不成立が生じてもオペレータが呼び出されなくなる確率が高くはなるが、高齢者以外は通常、対話トラブルに陥っていても比較的自己対処能力が高いので、オペレータが他の対話ロボットCを通じた高齢者のガイドにより注力できるようにすることができる。

【0057】

< 第2実施形態 >

図6は、第2実施形態の対話支援装置1aの機能構成の具体例を示すブロック図である

10

20

30

40

50

。対話支援装置 1 a は、記憶部 1 0 7 に代えて記憶部 1 0 7 a を備える点、判定基準値設定部 1 0 8 に代えて判定基準値設定部 1 0 8 a を備える点で、第 1 実施形態の対話支援装置 1 a と異なる。対話支援装置 1 a のその他の機能部は、第 1 実施形態の対話支援装置 1 と同様であるため、図 3 と同じ符号を付すことによって図 6 での説明を省略する。

【 0 0 5 8 】

記憶部 1 0 7 a は、判定基準値情報に加えて、制約情報及び条件情報をさらに記憶する。制約情報は、対話システムに関する運用上の制約を示す情報である。条件情報は、対話支援装置 1 a が介入判定における判定基準値を設定する際の条件を示す情報である。

【 0 0 5 9 】

判定基準値設定部 1 0 8 a は、制約情報及び条件情報に基づいて介入判定の判定基準値を決定する。判定基準値設定部 1 0 8 a は、決定した判定基準値を示す情報を判定基準値情報として記憶部 1 0 7 a に記憶させることで、自装置に介入判定の判定基準値を設定する。なお、判定基準設定部 1 0 8 a は、介入用通知判定部 1 0 9 が介入判定を行う都度、判定基準値を決定してもよいし、判定基準値の決定を所定期間ごとに行っても良い。例えば 1 日に 1 度閾値を見直すような場合、判定基準設定部 1 0 8 a は一旦決定した判定基準値を、その日に参照される判定基準値として記憶部 1 0 7 a に記憶させてもよい。

【 0 0 6 0 】

図 7 は、第 2 実施形態における制約情報及び条件情報の具体例を示す図である。図 7 ( A ) に示す制約情報テーブル T 1 は、制約情報の一例として、オペレータの要員計画情報を保持する。オペレータの要員計画情報は、システム運営者がオペレータとして確保する要員の計画を示す情報である。この場合、例えば制約情報テーブル T 1 は、日付及び要員の組み合わせごとに制約情報レコードを有する。制約情報レコードは、日付、要員及び勤務予定の各値を有し、" 日付 " の値が示す日における各 " 要員 " の " 勤務予定 " を表す。例えば、制約情報テーブル T 1 は、" 2 0 1 7 年 3 月 1 0 日 " において要員 A 及び B がオペレータとして勤務し ( 値 " " )、要員 C はオペレータとして勤務しない ( 値 " × " ) ことを表す。

【 0 0 6 1 】

また、図 7 ( B ) に示す条件情報テーブル T 2 は、条件情報の一例として、オペレータの数及び負荷と判定基準値とが対応づけられた情報を保持する。この場合、例えば条件情報テーブル T 2 は、要員総数と回線利用率との組み合わせごとに条件情報レコードを有する。条件情報レコードは、要員総数、回線利用率及び判定基準値の各値を有する。" 要員総数 " の値は、オペレータとして確保される要員の総数を表す。" 回線利用率 " の値は、オペレータと対話ロボット 2 との間の通話回線の利用率を表す。すなわち、回線利用率は、一定時間あたりに占める通話時間 ( 通話状態にある時間 ) の比率である。この場合、通話時間には、音声は途切れていても利用者が注意を傾けている時間等も含まれる。回線利用率が高い状況は、オペレータの介入が頻繁に発生している状況であり、オペレータの負荷が高い状況であると考えられる。そのため、ここでは、オペレータの負荷を表す指標値として回線利用率を用いる。" 判定基準値 " の値は、要員総数及び回線利用率によって表される状況において、オペレータの介入が適切な頻度で行われるように調整された判定基準値の閾値を表す。例えば、条件情報テーブル T 2 は、オペレータの総数が " 2 " 人であり、回線利用率が " 1 0 ~ 5 0 " % の範囲内である場合に設定されるべき判定基準値の閾値が 0 . 6 であることを表している。なお、図 7 ( B ) の条件情報テーブル T 2 における判定基準値は、対話ロボット 2 と利用者との対話が円滑に行われていない確率を表している。

【 0 0 6 2 】

このように構成された第 2 実施形態の対話支援装置 1 a では、判定基準値設定部 1 0 8 a が、制約情報及び条件情報に基づいて判定基準値を設定する。このような判定基準値の設定を行うことにより、対話支援装置 1 a は、システム運用上の制約の範囲内で、より効率良くオペレータを介入させることが可能になる。特にオペレータの要因計画に基づいて判定基準値を変更することで、オペレータの介入をより適切に行わせることが可能になる

。

### 【 0 0 6 3 】

< 変形例 >

以下、上記実施形態の対話支援装置 1 及び 1 a に共通の変形例について説明する。以下では、簡単のため対話支援装置 1 についての変形例として記載するが、以下に記載する変形例は対話支援装置 1 a にも適用可能である。

### 【 0 0 6 4 】

対話支援装置 1 は、音声のみを切り替えるだけでなく、利用者と対話中のオペレータの映像を対話ロボット 2 の表示部に表示させるように構成されてもよい。

### 【 0 0 6 5 】

対話支援装置 1 は、対話ロボット 2 と一体に構成されてもよいし、別体として構成されてもよい。図 8 は、対話支援装置 1 と一体に構成された対話ロボット 2 a の機能構成の具体例を示す図である。図 8 に示す各機能部のうち、対話支援装置 1 と同様の機能部には図 3 と同じ符号を付すことにより、対話支援装置 1 と同様の機能部についての説明を省略する。この場合、対話ロボット 2 a は、対話支援装置 1 が備える各機能部に加え、音声入力部 2 0 1、撮像部 2 0 2、音声対話データベース 2 0 3、音声制御部 2 0 4 及び音声出力部 2 0 5 を備える。音声入力部 2 0 1 は、自装置に利用者の音声を入力するとともに、入力された音声を音声データ取得部 1 0 3 に出力する。撮像部 2 0 2 は、利用者の画像を撮像するとともに、撮像した画像を画像データ取得部 1 0 2 に出力する。音声対話データベース 2 0 3 は、入力された利用者の音声や画像を認識し、利用者の発話の内容又は動作に応じて応答すべき内容を決定するために必要な情報を記憶するデータベースである。音声制御部 2 0 4 は、取得された利用者の音声及び画像と、音声対話データベース 2 0 3 に記憶された情報とに基づいて、利用者に対して応答すべき内容を決定する。音声出力部 2 0 5 は、音声制御部 2 0 4 によって決定された内容を音声として出力する。

### 【 0 0 6 6 】

従来は、オペレータを介入させるか否かの判定を利用者の音声や画像等の特徴ごとに介入判定を行っていたことから、介入頻度を適切に変更することが困難であった。これに対して、上述した実施形態の対話支援装置によれば、複数の指標値をまとめた一つの指標値を調整すればよいことから、介入判定の閾値を適切かつ容易に変更することができる。

### 【 0 0 6 7 】

上述した実施形態における対話支援装置又は対話ロボット（対話装置の一例）をコンピュータで実現するようにしてもよい。その場合、この機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによって実現してもよい。なお、ここでいう「コンピュータシステム」とは、OS や周辺機器等のハードウェアを含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM 等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含んでもよい。また上記プログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであってもよく、FPGA（Field Programmable Gate Array）等のプログラマブルロジックデバイスを用いて実現されるものであってもよい。

### 【 0 0 6 8 】

以上、この発明の実施形態について図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

10

20

30

40

50

【産業上の利用可能性】

【0069】

本発明は、入力された音声が表示内容に応じた音声を出力する対話装置を有するシステムに適用可能である。

【符号の説明】

【0070】

100 対話システム、 1, 1a 対話支援装置、 101 通信部、 102 画像データ取得部、 103 音声データ取得部、 104 画像特徴量取得部、 105 発話区間識別部、 106 音声特徴量取得部、 107 記憶部、 判定基準値設定部 108、 109 介入用通知判定部、 110 対話制御部、 2, 2a 対話ロボット、 201 音声入力部、 202 音声対話データベース、 203 音声制御部、 204 音声出力部、 3 オペレータ端末、 4 通信回線

10

【図1】

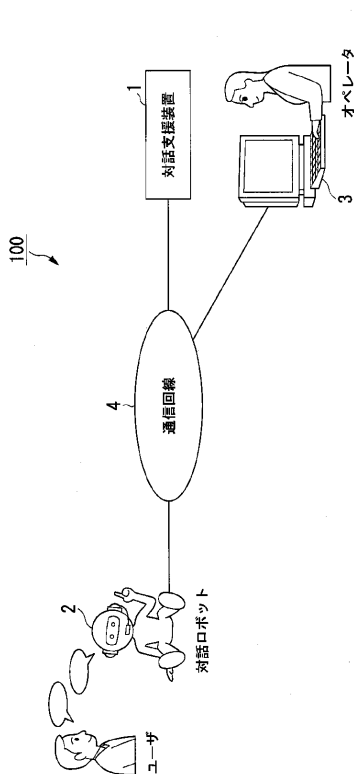


図1

【図2】

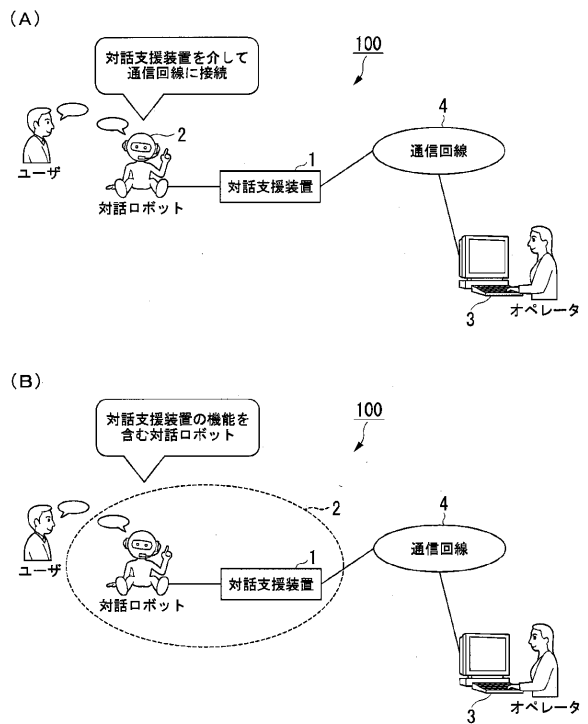


図2

【 図 3 】

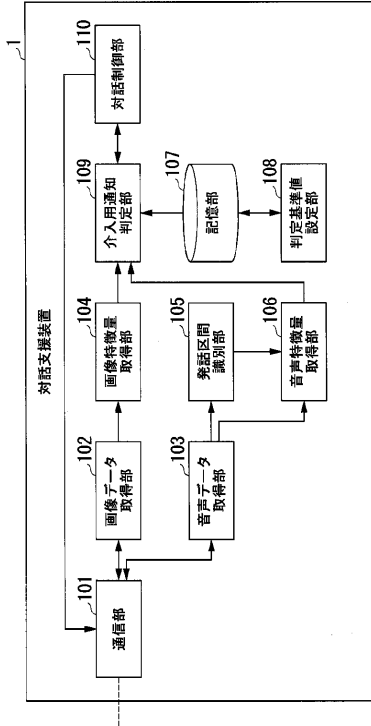


図 3

【 図 4 】

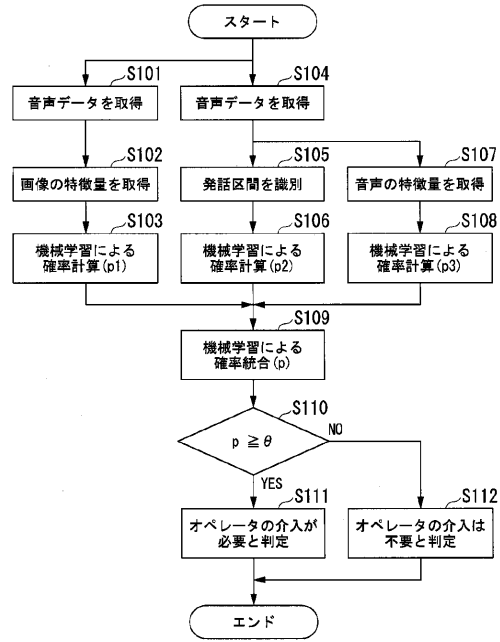


図 4

【 図 5 】

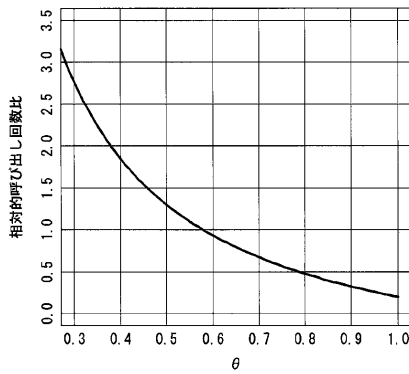


図 5

【 図 6 】

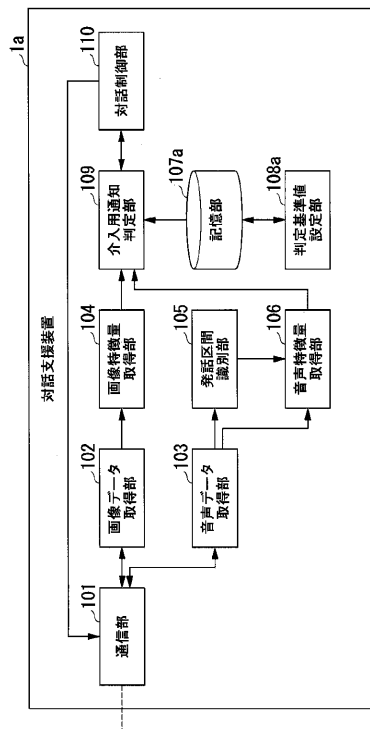


図 6

【図7】

(A)

日付	要員	勤務予定
2017年3月10日	A	○
	B	○
	C	×
2017年3月11日	...	...
...	...	...

(B)

要員総数	回線利用率	判定基準値
1	10-50	0.8
	50-100	0.6
2	10-50	0.6
	50-100	0.4
3	10-50	0.4
	50-100	0.2
...	...	...

図7

【図8】

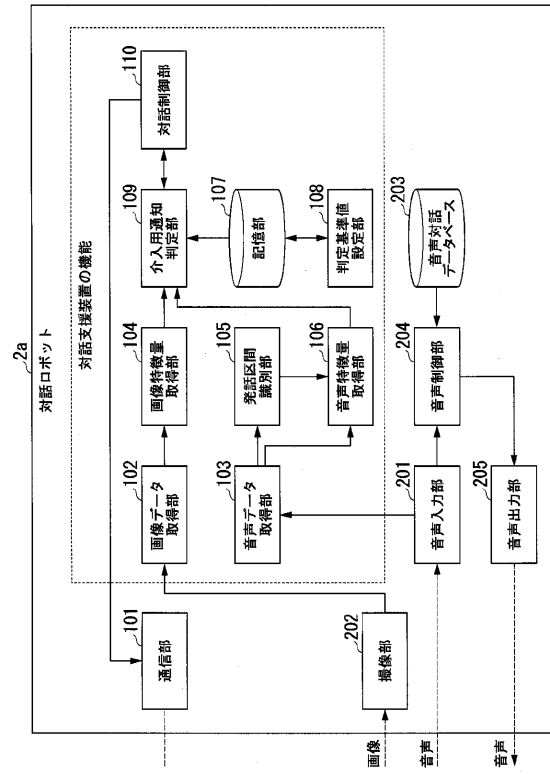


図8



---

フロントページの続き

(72)発明者 伊藤 彰則

宮城県仙台市青葉区片平二丁目1番1号 国立大学法人東北大学内

(72)発明者 廣井 富

大阪府大阪市旭区大宮五丁目16番1号 大阪工業大学内

審査官 岸 健司

(56)参考文献 特開2007-190641(JP,A)

特開2015-186018(JP,A)

特開2005-046320(JP,A)

特開2007-334841(JP,A)

千葉祐弥, 外1名, 対話中のユーザ状態逐次推定のための多段階識別手法に関する検討, 情報処理学会研究報告, 一般社団法人情報処理学会, 2013年04月15日, 音声言語情報処理 No.95, p.1

(58)調査した分野(Int.Cl., DB名)

G06Q 10/00 - 99/00

G10L 15/10

G10L 15/22