(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4735965号
(P4735965)
(45)発行日　平成23年7月27日(2011.7.27)　(24)登録日　平成23年5月13日(2011.5.13)

(51)Int.Cl.　F I
H04M 11/00 (2006.01)　H04M 11/00　301

請求項の数 2　(全 24 頁)

(21)出願番号　特願2005-322141 (P2005-322141)
(22)出願日　平成17年11月7日 (2005.11.7)
(65)公開番号　特開2007-129626 (P2007-129626A)
(43)公開日　平成19年5月24日 (2007.5.24)
審査請求日　平成20年3月31日 (2008.3.31)

（出願人による申告）平成１７年４月１日付け、支出負担行為担当官　総務省大臣官房会計課企画官、研究テーマ「ネットワーク・ヒューマン・インターフェースの総合的な研究開発（ネットワークロボットの技術）」に関する委託研究、産業活力再生特別措置法第３０条の適用を受ける特許出願

前置審査

(73)特許権者　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100090181
弁理士　山田　義人
(72)発明者　小泉　智史
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　塩見　昌裕
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　神田　崇行
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54)【発明の名称】遠隔地間対話システム

(57)【特許請求の範囲】
【請求項１】
ネットワークを介して接続される２つの対話装置を含む遠隔地間で対話を行うためのシステムであって、
各前記対話装置は、
音声を取得する取得手段、
前記取得手段で取得した前記音声を相手側の前記対話装置へ送信する送信手段、
相手側の前記対話装置から送信された音声を受信する受信手段、および
前記受信手段で受信した前記音声を出力する出力手段を含んでいて、
少なくとも空白時間と発話音声の特徴に関する情報を含む複数の間パターンを記憶する間パターン記憶手段、
各前記対話装置における少なくとも音声取得状態および音声出力状態を含む対話状態の履歴を記録する履歴記録手段、
前記履歴記録手段によって記録された前記履歴に基づいて両方の前記対話装置で無発話状態であると判定されるとき、少なくとも空白時間および当該空白前の発話音声の特徴を含む間の状況と前記複数の間パターンとの照合を行う照合手段、
前記照合手段による照合の結果マッチする前記間パターンがあるとき、当該間パターンに対応する所定の音声を、当該所定の音声の発話者の相手側に存在する前記対話装置の前記出力手段から出力する間制御手段、および
前記履歴記録手段によって記録された前記履歴に基づいて両方の前記対話装置で発話が

重複したと判定されるとき、一方の音声を録音して、その後発話が終了したときに当該録音音声を他方の前記対話装置の前記出力手段から出力する遅延再生手段を備える、遠隔地間対話システム。

【請求項２】

前記対話装置の少なくとも一方が身振りを実行可能なロボットであるとき、前記間制御手段は、前記音声の出力とともに、前記間パターンに対応する所定の身振りを当該対話装置に実行させる、請求項１記載の遠隔地間対話システム。

【発明の詳細な説明】

【技術分野】

【０００１】

この発明は遠隔地間対話システムに関し、特にたとえば、遠隔地に離れた対話者同士の音声をネットワークを介して通信する、遠隔地間対話システムに関する。

【背景技術】

【０００２】

遠隔地間で対話を行う場合には、遅延により対話の空白時間が長くなり発話のタイミングが取りづらくなって、両対話者の発話が重複する事態が生じ易い。従来、このような発話の重複を防止して、遠隔地間の円滑な対話を生成しようとする技術は存在しなかった。

【０００３】

なお、たとえば特許文献１には、音声入出力装置の出力する音声と利用者の発話によって入力される音声との重畳を検出する技術の一例が開示されている。

【０００４】

また、特許文献２および特許文献３には、単に対話音声を出力するだけでなく、対話者の前にロボットを設置してゼスチャを行わせる技術の一例が開示されている。特許文献２の技術では、話し手の音声に基づいて当該話し手側のロボットが身振りを実行し、一方、聞き手側で受信した音声に基づいて当該聞き手側のロボットが身振りを実行することで、会話の実感を高めている。また、特許文献３の技術では、話し手側の身振り情報の送信に応じて相手側ロボットで当該身振りが再現される。

【特許文献１】特開平７－２６４１０３号公報

【特許文献２】特開２０００－３４９９２０号公報

【特許文献３】特開２００１－１５６９３０号公報

【発明の開示】

【発明が解決しようとする課題】

【０００５】

特許文献１の技術では、出力音声と入力音声との重畳を検出して、エコーキャンセラの動作が変化されるが、話者の発話タイミングを制御することはできない。また、特許文献２および３の技術では、対話者の前に設置したロボットに身振りをさせることによって、円滑な対話の実現を図っているが、話者の発話タイミングを制御することはできない。このように、従来技術では、発話のタイミングを制御することができなかったので、遅延により対話に異常な空白時間が生じても対応できなかった。したがって、両対話者の発話が重なることを防止することができず、円滑な対話を実現することができなかった。

【０００６】

それゆえに、この発明の主たる目的は、遠隔地間対話に適切な間を与えることができて、円滑な対話を実現できる、遠隔地間対話システムを提供することである。

【課題を解決するための手段】

【０００７】

請求項１の発明は、ネットワークを介して接続される２つの対話装置を含む遠隔地間で対話を行うためのシステムである。各対話装置は、音声を取得する取得手段、取得手段で取得した音声を相手側の対話装置へ送信する送信手段、相手側の対話装置から送信された音声を受信する受信手段、および受信手段で受信した音声を出力する出力手段を含んでいる。当該システムは、少なくとも空白時間と発話音声の特徴に関する情報を含む複数の間

パターンを記憶する間パターン記憶手段、各対話装置における少なくとも音声取得状態および音声出力状態を含む対話状態の履歴を記録する履歴記録手段、履歴記録手段によって記録された履歴に基づいて両方の対話装置で無発話状態であると判定されるとき、少なくとも空白時間および当該空白前の発話音声の特徴を含む間の状況と複数の間パターンとの照合を行う照合手段、照合手段による照合の結果マッチする間パターンがあるとき、当該間パターンに対応する所定の音声を、当該所定の音声の発話者の相手側に存在する対話装置の出力手段から出力する間制御手段、および履歴記録手段によって記録された履歴に基づいて両方の対話装置で発話が重複したと判定されるとき、一方の音声を録音して、その後発話が終了したときに当該録音音声を他方の対話装置の出力手段から出力する遅延再生手段を備える。

【０００８】

請求項１の発明では、遠隔地間対話システムは２つの対話装置を含み、各対話装置が取得した音声を通信して相手側で出力することによって、遠隔地間での対話者同士の対話が行われる。間パターン記憶手段には複数の間パターンが記憶されている。間パターンは、対話における適切な間の取り方を示し、少なくとも空白時間と発話音声の特徴に関する情報を含む。たとえば、発話音声の特徴は、基本周波数（ピッチ）、振幅および音節の平均持続時間等を含んでよい。履歴記録手段は、各対話装置における少なくとも音声取得状態および音声出力状態を含む対話状態の履歴を記録する。対話状態の履歴は、後述される実施例では発話フラグテーブルであり、各時刻の発話の有無状態（ＳＰＥＡＫＩＮＧフラグ、ＳＩＬＥＮＴフラグ）および処理の状態（ＲＥＣＯＲＤＩＮＧフラグ、ＩＮＴＥＲＰＯＬＡＴＩＮＧフラグ）などが記録される。照合手段は、両方の対話装置で無発話状態であると判定されるとき、少なくとも空白時間および当該空白前の発話音声の特徴を含む間の状況と、複数の間パターンとの照合を行う。つまり、対話が無音状態である場合に、現在の間の状況が複数の間パターンのいずれかにマッチしているかが確認される。間制御手段は、照合の結果マッチする間パターンがあるとき、当該間パターンに対応する所定の音声を、当該所定の音声の発話者の相手側に存在する対話装置の出力手段から出力する。ただし、発話が重複してしまった場合には、遅延再生手段によって、一方の音声が録音され、その後発話が終了したときに、当該録音した音声が他方の対話装置から出力される。

なお、上記間パターン記憶手段、履歴記録手段、照合手段、間制御手段および遅延再生手段は、２つの対話装置のいずれか一方に、または、このシステムに含まれる別の装置（実施例では発話タイミング制御サーバ）に設けられてよい。あるいは、これらの手段は、２つの対話装置に分散して設けられてもよい。

【０００９】

請求項１の発明によれば、会話における無音状態が検出されたときに、適切な間を与える間パターンに対応する音声を出力することができる。したがって、発話が重なってしまうのを防止することができ、円滑な対話を成立させることができる。また、万一発話が重複してしまっても、両発話が同時に相手側で出力されるのを回避することができる。

【００１０】

請求項２の発明は、請求項１の発明に従属し、対話装置の少なくとも一方が身振りを実行可能なロボットであるとき、間制御手段は、音声の出力とともに、さらに間パターンに対応する所定の身振りを当該対話装置に実行させる。

【００１１】

請求項２の発明では、対話装置の少なくとも一方は、身振りを実行可能なロボットであってよい。間制御手段は、当該ロボットに、間パターンに対応する音声を出力させるとともに、間パターンに対応する身振りを実行させる。したがって、音声と身振りを使用して、対話に適切な間を与えることができるので、より円滑な遠隔地間対話を成立させることができる。

【発明の効果】

【００１４】

この発明によれば、対話に無音状態が検出されたときに適切な間を取るように言葉を挿

入するようにしたので、対話の空白時間を適切な長さにすることができる。このため、遅延によって空白時間が長くなって対話者に違和感を与えてしまうようなことを回避できる。したがって、対話者は発話のタイミングを計りやすくなるので、両者の発話の重複を防止することができ、円滑な対話を成立させることができる。

【００１５】

この発明の上述の目的，その他の目的，特徴および利点は、図面を参照して行う以下の実施例の詳細な説明から一層明らかとなろう。

【発明を実施するための最良の形態】

【００１６】

図１を参照して、この実施例の遠隔地間対話システム（以下、単に「システム」とも言う。）１０は、遠隔地に離れた対話者同士が対話を行うためのものである。システム１０は少なくとも２つの対話装置１２（１２ａ，１２ｂ）を含む。２つの対話装置１２は、ネットワーク、たとえば公衆インターネット網を介して接続されており、対話装置１２ａ側の対話者Ａおよび対話装置１２ｂ側の対話者Ｂの発話した音声の音声データを互いに通信する。また、この実施例のシステム１０は、発話タイミング制御サーバ（以下、単に「サーバ」とも言う。）１４を含み、サーバ１４はネットワークを介して少なくとも２つの対話装置１２と通信可能に接続される。

【００１７】

この実施例では、一方の対話装置１２としてコンピュータ１２ａが適用され、他方の対話装置１２としてコミュニケーションロボット（以下、単に「ロボット」とも言う。）１２ｂが適用された場合を説明する。

【００１８】

対話装置１２ａはマイク１６およびスピーカ１８を備える。また、対話装置１２ａはたとえばパーソナルコンピュータであり、ＣＰＵ、メインメモリ、通信装置および入力装置等を備えている。メインメモリには、この発明の対話装置１２として機能するために必要なプログラムおよびデータが記憶される。プログラムおよびデータは、メインメモリに予め固定的に記憶されてもよいし、または、情報記憶媒体やネットワークから取得されてよい。ＣＰＵは、当該プログラムに従って、メインメモリのうちのワーキングメモリに一時的なデータを生成または取得しつつ対話のための処理を実行する。

【００１９】

マイク１６は対話者の発話した音声を取得するためのものであり、当該音声は音声入出力ボードでデータに変換されて、音声データとしてメインメモリに記憶される。スピーカ１８は、対話相手の音声およびシステム１０の備える音声を出力するためのものである。ＣＰＵは受信した音声データを音声入出力ボードに与えて当該音声をスピーカ１８から出力する。通信装置は、ネットワークを介して他方の対話装置１２やサーバ１４にデータを送受信する。また、入力装置は、キーボードまたはポインティングデバイス等である。

【００２０】

この実施例では、相手側の対話装置１２としてロボット１２ｂが使用されるので、ユーザが入力装置を用いてロボット１２ｂの身振りを指示可能になっている。ロボット１２ｂの身振りは、表示装置の画面に表示されたリストから選択されてよいし、あるいは入力装置の各キーに割り当てられてもよい。対話者Ａは発話しながら入力装置を用いて動作を指示することによって、相手側のロボット１２から自分の音声を出力することができ、しかも当該ロボット１２に所望の身振りを行わせることができる。

【００２１】

なお、対話装置１２ａは、音声入出力可能かつ通信可能なコンピュータであればよく、ＰＣに限られず、ゲーム機、携帯電話、携帯ゲーム機などの他のコンピュータであってよい。

【００２２】

他方の対話装置１２ｂは人間のような身体部位を有するロボットであり、身体部位を動かすことによって所定の身振りを対話者Ｂに提示することができる。このロボット１２ｂ

は、マイク２０およびスピーカ２２を備えている。詳しくは、図２にロボット１２ｂの外観の一例が示され、図３には当該ロボット１２の電気的な構成の一例が示される。

【００２３】

図２を参照して、ロボット１２ｂは台車２４を含み、この台車２４の下面には、このロボット１２ｂを自律移動させる車輪２６が設けられる。この車輪２６は、車輪モータ（図３において参照番号「２８」で示す。）によって駆動され、台車２４すなわちロボット１２ｂを前後左右任意の方向に動かすことができる。なお、図２では示さないが、この台車２４の前面には、衝突センサ（図３において参照番号「３０」で示す。）が取り付けられ、この衝突センサ３０は、台車２４への人や他の障害物との接触を検知する。ロボット１２ｂの移動中に接触を検知すると、直ちに車輪２６の駆動を停止することができる。

【００２４】

台車２４の上には、多角形柱のセンサ取付パネル３２が設けられ、このセンサ取付パネル３２の各面には、超音波距離センサ３４が取り付けられる。この超音波距離センサ３４は、取付パネル３２すなわちロボット１２ｂの周囲の主として人との間の距離を計測するためのものである。

【００２５】

台車２４の上には、さらに、ロボット１２ｂの胴体が、その下部が上述の取付パネル３２に囲まれて、直立するように取り付けられる。この胴体は下部胴体３６と上部胴体３８とから構成され、これら下部胴体３６および上部胴体３８は、連結部４０によって連結される。連結部４０には、図示しないが、昇降機構が内蔵されていて、この昇降機構を用いることによって、上部胴体３８の高さすなわちロボット１２ｂの高さを変化させることができる。昇降機構は、腰モータ（図３において参照番号「４２」で示す。）によって駆動される。

【００２６】

上部胴体３８のほぼ中央には、１つの全方位カメラ４４と、１つのマイク２０とが設けられる。全方位カメラ４４は、ロボット１２ｂの周囲を撮影するもので、後述の眼カメラ４６と区別される。マイク２０は、上述のように、周囲の音、とりわけ人の声を取り込む。

【００２７】

上部胴体３８の両肩には、それぞれ、肩関節４８Ｒおよび４８Ｌによって、上腕５０Ｒおよび５０Ｌが取り付けられる。肩関節４８Ｒおよび４８Ｌは、それぞれ３軸の自由度を有する。すなわち、右肩関節４８Ｒは、Ｘ軸，Ｙ軸およびＺ軸の各軸廻りにおいて上腕５０Ｒの角度を制御できる。Ｙ軸は、上腕５０Ｒの長手方向（または軸）に平行な軸であり、Ｘ軸およびＺ軸は、そのＹ軸に、それぞれ異なる方向から直交する軸である。左肩関節４８Ｌは、Ａ軸，Ｂ軸およびＣ軸の各軸廻りにおいて上腕５０Ｌの角度を制御できる。Ｂ軸は、上腕５０Ｌの長手方向（または軸）に平行な軸であり、Ａ軸およびＣ軸は、そのＢ軸に、それぞれ異なる方向から直交する軸である。

【００２８】

上腕５０Ｒおよび５０Ｌのそれぞれの先端には、肘関節５２Ｒおよび５２Ｌを介して、前腕５４Ｒおよび５４Ｌが取り付けられる。肘関節５２Ｒおよび５２Ｌは、それぞれ、Ｗ軸およびＤ軸の軸廻りにおいて、前腕５４Ｒおよび５４Ｌの角度を制御できる。

【００２９】

なお、上腕５０Ｒおよび５０Ｌならびに前腕５４Ｒおよび５４Ｌの変位を制御するＸ，Ｙ，Ｚ，Ｗ軸およびＡ，Ｂ，Ｃ，Ｄ軸では、「０度」がホームポジションであり、このホームポジションでは、上腕５０Ｒおよび５０Ｌならびに前腕５４Ｒおよび５４Ｌは下方向に向けられる。

【００３０】

また、図２では示さないが、上部胴体３８の肩関節４８Ｒおよび４８Ｌを含む肩の部分や上述の上腕５０Ｒおよび５０Ｌならびに前腕５４Ｒおよび５４Ｌを含む腕の部分には、それぞれ、タッチセンサ（図３において参照番号「５６」で包括的に示す。）が設けられ

ていて、これらのタッチセンサ５６は、人がロボット１２ｂのこれらの部位に接触したかどうかを検知する。

【００３１】

前腕５４Ｒおよび５４Ｌのそれぞれの先端には、手に相当する球体５８Ｒおよび５８Ｌがそれぞれ固定的に取り付けられる。なお、この球体５８Ｒおよび５８Ｌに代えて、この実施例のロボット１２ｂと異なり指の機能が必要な場合には、人の手の形をした「手」を用いることも可能である。

【００３２】

上部胴体３８の中央上方には、首関節６０を介して、頭部６２が取り付けられる。この首関節６０は、３軸の自由度を有し、Ｓ軸，Ｔ軸およびＵ軸の各軸廻りに角度制御可能である。Ｓ軸は首から真上に向かう軸であり、Ｔ軸およびＵ軸は、それぞれ、このＳ軸に対して異なる方向で直交する軸である。頭部６２には、人の口に相当する位置に、上述のスピーカ２２が設けられる。なお、スピーカ２２は、ロボット１２ｂが、それの周囲の人に対して音声または声によってコミュニケーションを図るために用いられてよい。また、スピーカ２２は、ロボット１２の他の部位たとえば胴体に設けられてもよい。

【００３３】

また、頭部６２には、目に相当する位置に眼球部６４Ｒおよび６４Ｌが設けられる。眼球部６４Ｒおよび６４Ｌは、それぞれ眼カメラ４６Ｒおよび４６Ｌを含む。なお、左右の眼球部６４Ｒおよび６４Ｌをまとめて参照符号「６４」で示し、左右の眼カメラ４６Ｒおよび４６Ｌをまとめて参照符号「４６」で示すこともある。眼カメラ４６は、ロボット１２ｂに接近した人の顔や他の部分ないし物体等を撮影してその映像信号を取り込む。

【００３４】

なお、上述の全方位カメラ４４および眼カメラ４６のいずれも、たとえばＣＣＤやＣＭＯＳのように固体撮像素子を用いるカメラであってよい。

【００３５】

たとえば、眼カメラ４６は眼球部６４内に固定され、眼球部６４は眼球支持部（図示せず）を介して頭部６２内の所定位置に取り付けられる。眼球支持部は、２軸の自由度を有し、α軸およびβ軸の各軸廻りに角度制御可能である。α軸およびβ軸は頭部６２に対して設定される軸であり、α軸は頭部６２の上へ向かう方向の軸であり、β軸はα軸に直交しかつ頭部６２の正面側（顔）が向く方向に直交する方向の軸である。この実施例では、頭部６２がホームポジションにあるとき、α軸はＳ軸に平行し、β軸はＵ軸に平行するように設定されている。このような頭部６２において、眼球支持部がα軸およびβ軸の各軸廻りに回転されることによって、眼球部６４ないし眼カメラ４６の先端（正面）側が変位され、カメラ軸すなわち視線方向が移動される。

【００３６】

なお、眼カメラ４６の変位を制御するα軸およびβ軸では、「０度」がホームポジションであり、このホームポジションでは、図２に示すように、眼カメラ４６のカメラ軸は頭部６２の正面側（顔）が向く方向に向けられ、視線は正視状態となる。

【００３７】

図３を参照して、このロボット１２ｂは、全体の制御のためにマイクロコンピュータまたはＣＰＵ６６を含み、このＣＰＵ６６には、バス６８を通して、メモリ７０，モータ制御ボード７２，センサ入力／出力ボード７４および音声入力／出力ボード７６が接続される。

【００３８】

メモリ７０は、図示しないが、ＲＯＭやＨＤＤおよびＲＡＭ等を含み、ＲＯＭまたはＨＤＤには、このロボット１２ｂをこの発明の対話装置１２として機能させるためのプログラムおよびデータが予め格納されている。ＣＰＵ６６は、このプログラムに従って処理を実行する。また、ＲＡＭは、バッファメモリやワーキングメモリとして使用される。

【００３９】

モータ制御ボード７２は、たとえばＤＳＰ(Digital Signal Processor)で構成され、右

腕、左腕、頭および眼等の身体部位を駆動するためのモータを制御する。すなわち、モータ制御ボード７２は、ＣＰＵ６６からの制御データを受け、右肩関節４８ＲのＸ，Ｙおよびｚ軸のそれぞれの角度を制御する３つのモータと右肘関節５２Ｒの軸Ｗの角度を制御する１つのモータを含む計４つのモータ（図３ではまとめて「右腕モータ」として示す。）７８の回転角度を調節する。また、モータ制御ボード７２は、左肩関節４８ＬのＡ，ＢおよびＣ軸のそれぞれの角度を制御する３つのモータと左肘関節５２ＬのＤ軸の角度を制御する１つのモータとを含む計４つのモータ（図３ではまとめて「左腕モータ」として示す。）８０の回転角度を調節する。モータ制御ボード７２は、また、首関節６０のＳ，ＴおよびＵ軸のそれぞれの角度を制御する３つのモータ（図３ではまとめて「頭部モータ」として示す。）８２の回転角度を調節する。モータ制御ボード７２は、また、腰モータ４２および車輪２６を駆動する２つのモータ（図３ではまとめて「車輪モータ」として示す。）２８を制御する。さらに、モータ制御ボード７２は、右眼球部６４Ｒのα軸およびβ軸のそれぞれの角度を制御する２つのモータ（図３ではまとめて「右眼球モータ」として示す。）８４の回転角度を調節し、また、左眼球部６４Ｌのα軸およびβ軸のそれぞれの角度を制御する２つのモータ（図３ではまとめて「左眼球モータ」として示す。）８６の回転角度を調節する。

【００４０】

なお、この実施例の上述のモータは、車輪モータ２８を除いて、制御を簡単化するためにそれぞれステッピングモータまたはパルスモータであるが、車輪モータ２８と同様に、直流モータであってよい。また、この実施例では、ロボット１２ｂの腕、頭、眼などの身体部位を駆動するアクチュエータとして電力を駆動源とするモータを用いた。しかしながら、このロボット１２ｂとしては、たとえば空気圧（または負圧）、油圧、圧電素子あるいは形状記憶合金などによる他のアクチュエータによって身体部位を駆動するロボットが適用されてもよい。

【００４１】

センサ入力／出力ボード７４も、同様に、ＤＳＰで構成され、各センサやカメラからの信号を取り込んでＣＰＵ６６に与える。すなわち、超音波距離センサ３４の各々からの反射時間に関するデータがこのセンサ入力／出力ボード７４を通して、ＣＰＵ６６に入力される。また、全方位カメラ４４からの映像信号が、必要に応じてこのセンサ入力／出力ボード７４で所定の処理が施された後、ＣＰＵ６６に入力される。眼カメラ４６からの映像信号も、同様にして、ＣＰＵ６６に与えられる。また、タッチセンサ５６からの信号がセンサ入力／出力ボード７４を介してＣＰＵ６６に与えられる。

【００４２】

スピーカ２２には音声入力／出力ボード７６を介して、ＣＰＵ６６から音声データが与えられ、それに応じて、スピーカ２２からはそのデータに従った音声または声が出力される。また、マイク２０からの音声入力が、音声入力／出力ボード７６を介して音声データとしてＣＰＵ６６に取り込まれる。

【００４３】

通信ＬＡＮボード８８も、同様に、ＤＳＰで構成され、ＣＰＵ６６から与えられた送信データを無線通信装置９０に与えて、当該データを無線通信装置９０から送信させる。また、通信ＬＡＮボード８８は無線通信装置９０を介してデータを受信し、受信データをＣＰＵ６６に与える。

【００４４】

図１に戻って、サーバ１４は、両対話者の発話のタイミングを制御するために設けられる。サーバ１４は、ＣＰＵ、メインメモリ、通信装置等を備える。メインメモリにはこのサーバ１４を制御するためのプログラムおよびデータが記憶される。ＣＰＵは当該プログラムに従って処理を実行する。

【００４５】

また、サーバ１４は音声解析履歴データベース（ＤＢ）９２および間パターンＤＢ９４を含む。音声解析履歴ＤＢ９２には、対話装置１２で取得された対話者の音声の解析デー

タの履歴が記憶される。

【００４６】

間パターンＤＢ９４には、後述するように、対話に適切な間を与えるための間パターンデータ（図４参照）が記憶されている。間パターンデータは、予め発話の計測を行って得た発話データからパターン認識によって抽出される。計測を実際の使用者を対象として行うと、間の取り方の個人的特徴を抽出できる。ただし、標準的なまたは一般的な間の取り方も存在すると考えられるので、任意の人を被験者としてその発話を計測して間パターンデータを抽出してよい。

【００４７】

このシステム１０では、各対話装置１２が、一定時間ΔＴごとにマイク１６または２０で音声を検出する。ΔＴはたとえば１フレームまたは所定のフレーム数であってよい。１フレームはたとえば１／３０秒である。対話装置１２は、検出した結果すなわち発話の有無に応じた処理を行う。対話装置１２は、検出時刻における発話状態（音声取得状態）および実行した処理（音声出力状態）など、当該装置１２における対話状態に関する情報をサーバ１４に送信する。サーバ１４は、当該対話状態に関する情報を受信して、当該対話装置１２における状態を逐一記憶する。このような対話状態の履歴は、発話フラグテーブル（図５参照）としてメモリに記憶される。発話フラグテーブルでは、後述するように、検出時刻ごとの対話装置１２における少なくとも音声取得状態および音声出力状態を含む対話状態を示すフラグが記憶されている。

【００４８】

なお、便宜上、ここでは、対話装置１２ａ側からみた動作を説明する。しかし、対話装置１２ａと対話装置１２ｂとの相違は、主に身体動作の提示に関する機能のみであるから、対話装置１２ｂの動作も、対話装置１２ａの場合と同様である。

【００４９】

対話装置１２ａは、マイク１６から音声を検出した場合には、サーバ１４の発話フラグテーブルを参照する。そして、対話装置１２ａは、１つ前の検出時刻における（つまりΔＴ前の）相手側の状態に応じた処理を実行する。具体的には、前検出時刻における相手の状態フラグがＳＰＥＡＫＩＮＧフラグでない場合には、つまり、前検出時刻において相手が自分に音声を送信している状態ではない場合には、対話装置１２ａは、マイク１６で検出した音声データをメモリにローカルファイルとして記録しつつ、当該音声データを相手側の対話装置１２ｂに送信する。これに応じて、対話装置１２ｂは、当該音声データを受信して当該音声をスピーカ２２から出力する。このように、一方の対話者Ａが発話し、かつ、前検出時刻で他方の対話者Ｂの音声が送信されていない場合には、対話者Ａの音声データが直ちに送信され、当該音声が相手側の対話装置１２ｂで出力されて、相手Ｂに聞かせられる。

【００５０】

また、この場合、対話装置１２ａは、ＳＰＥＡＫＩＮＧフラグをサーバ１４に送信し、これに応じて、サーバ１４は発話フラグテーブルに当該対話装置１２ａの当該検出時刻ｔにおける状態として当該ＳＰＥＡＫＩＮＧフラグを記憶する。ＳＰＥＡＫＩＮＧフラグは、一方の対話者が発話している状態、すなわち、発話音声が直接相手側に送信されて再生されている状態を意味する。さらに、対話装置１２ａは、メモリのＳＥＮＤフラグをオンにして、自分の処理状態として、音声を相手に送信中であることを記憶する。

【００５１】

なお、当該音声を記録したローカルファイルには、当該発話が終わったときに音声解析が実行され、当該発話音声の特徴ないし状態はサーバ１４の音声解析履歴ＤＢ９２に記憶される。

【００５２】

一方、対話装置１２ａはマイク１６から音声を検出しなかった場合には、ＳＩＬＥＮＴフラグをサーバ１４に送信し、これに応じて、サーバ１４は発話フラグテーブルに当該ＳＩＬＥＮＴフラグを記憶する。ＳＩＬＥＮＴフラグは、当該検出時刻ｔにおいて対話者が

発話してない状態を意味する。このように、音声が検出されない場合には、発話フラグテーブルにＳＩＬＥＮＴフラグが記録される。

【００５３】

サーバ１４では、両対話者とも発話していない状態（ＳＩＬＥＮＴフラグ）が検出されたとき、間パターンと対話における現在までの発話状況（間の状況）との照合が行われる。間の状況は、少なくとも対話の空白時間（無音時間）、および当該空白前の対話における発話音声の特徴（音声解析結果）を含む。

【００５４】

間パターンＤＢ９４に記憶される間パターンデータの一例が図４に示される。間パターンＤＢ９４には、会話に適切な間を与えることができる複数の間パターンデータが記憶されている。間パターンは、少なくとも空白時間とその前の発話音声の特徴に関する情報を含む。この実施例では、間パターンデータは、会話の空白時間、最終発話者、条件（Ｉ）、間機能言葉、発話者および動作コマンド等の情報を含む。空白時間（ｔ）は、両者無音状態が継続している時間の条件である。最終発話者は、当該空白前の対話での最後の発話者ＡまたはＢの条件である。条件（Ｉ）は、当該空白前の対話の音声の解析結果の条件であり、たとえば基本周波数（ピッチ）、振幅および音節の平均持続時間等の要素を含む。

【００５５】

このような空白時間とその前の発話音声の特徴によって規定される間パターンに見合う間の状況が検出されたとき、当該間パターンに基づく言葉および身振りが対話に挿入される。具体的には、この実施例では、最終発話者、条件（Ｉ）および空白時間条件（ｔ）に合う対話後の空白時間が生じた場合に、当該パターンで指定された発話者側から間機能言葉が発せられる。また、発話者の代わりに当該間機能言葉を発する対話装置１２がロボット１２ｂである場合には、当該出力する言葉に対応する身振りも動作コマンドに従って再現される。

【００５６】

間機能言葉は、無音時間に挿入されることによって無音を間として機能させ会話に適切な間を与えるための言葉である。たとえば、応答、合の手、間投詞などの言葉であってよい。図４では、「うんうん」、「うーん」、「はいはい」および「えーと」などが間機能言葉として示される。また、各間機能言葉には、当該言葉とともに提示される身体動作を実行するための動作コマンドが対応付けられている。図４では、「うんうん」には「うなづく」のコマンド、「うーん」には「首傾げる」のコマンド、「はいはい」には「うなづく」のコマンド、「えーと」には「視線を上方に向ける」コマンドがそれぞれ対応付けられる。

【００５７】

なお、図４において、発話者は、人間である対話者を意味しており、当該間機能言葉を出力する対話装置１２は、この発話者の相手側の場所に存在する対話装置１２である。たとえば、図４の一番上のパターンの場合、最後の発話者が対話者Ａであり、間機能言葉を出力する対話装置は、発話者Ｂの相手である対話者Ａ側の対話装置１２ａとなる。一番上のパターンは「うんうん」という言葉と「うなづく」行動に対応付けられており、最終発話者の相手側である発話者が応答する動作を、最終発話者側に存在するロボット１２ｂが表現することで、間が与えられる。一方、上から２番目のパターンは「うーん」という言葉と「首を傾げる」行動に対応付けられており、最終発話者の相手側に存在するロボット１２ｂが発話および身振りをさらに続けることによって、間が与えられる。

【００５８】

間パターンを用いた照合の結果、現在の間の状況にマッチする間パターンデータが間パターンＤＢ９４に存在する場合には、つまり、間パターンに従って適切な間を取る必要がある状況であると判断される場合には、サーバ１４は、必要な対話装置１２に間を取るための言葉の再生を指示する。この実施例では、当該言葉の音声データと再生指示とが送信される。これに応じて、当該対話装置１２側で当該音声が出力される。さらに、対話装置１２が身体表現可能な対話装置１２ｂである場合には、サーバ１４は当該間を取るための

言葉に対応する身振りの実行を指示する。この実施例では、当該身振りに動作コマンドと再生指示とが送信される。これに応じて、当該対話装置１２ｂでは、対応する身体部位が動かされて当該身振りが実行される。

【００５９】

このように、対話に無発話状態（無音状態）が検出されたときに、現在の間の状況と間パターンとの照合を行うようにした。そして、必要があれば適切な間を取るように言葉や身振りを空白時間に挿入するようにしたので、対話における空白時間を適切な時間に維持することができる。このため、遅延によって対話における空白時間が長くなって対話者に違和感を与えてしまうようなことを回避できる。したがって、対話者は発話のタイミングを計りやすくなり、両対話者の発話が重複する事態が生ずるのを防止することができる。このように、言葉や身振りの挿入によって対話者の発話タイミングを制御することができるので、対話を継続させたり発話を促進したりすることができるし、自然な会話の流れを作り出すことができる。したがって、円滑な対話を成立させることができる。

【００６０】

また、このシステム１０では、万一両対話者の発話が重複した場合には、一方の発話の出力を遅らせることによって、両発話が完全に重なってしまうのを回避する機能を備えるようにしている。

【００６１】

具体的には、対話装置１２ａで音声が検出された場合において、前検出時刻の相手の状態フラグがＳＰＥＡＫＩＮＧフラグであるときには、つまり、両者の発話が重複している場合には、対話装置１２ａは、マイク１６で検出した音声の録音を開始し、当該音声データを音声ファイルとしてメモリに記憶する。さらに、対話装置１２ａは、ＲＥＣＯＲＤＩＮＧフラグをサーバ１４に送信し、これに応じて、サーバ１４は発話フラグテーブルに当該対話装置１２ａの当該検出時刻ｔにおける状態として当該ＲＥＣＯＲＤＩＮＧフラグを記憶する。ＲＥＤＯＲＤＩＮＧフラグは、音声データを録音中であり、当該音声が相手側に送信されていない状態を意味する。

【００６２】

また、対話装置１２ａは、メモリのＲＥＣＯＲＤフラグをオンにして、自分の処理状態として、音声を録音中であることを記憶する。

【００６３】

また、サーバ１４の発話フラグテーブルでは、録音した音声ファイルの再生を制御するための情報としてＰＬＡＹフラグが記憶される。この実施例では、対話装置１２ａは、録音しているときは、ＰＬＡＹフラグの値に１を加算するようにサーバ１４に指示する。ＰＬＡＹフラグの初期値は０であり、録音が行われているときは毎検出時刻ごとに前の検出時刻の値に１だけ加算され、録音が行われていないときには前の検出時刻の値が維持される。

【００６４】

その後、対話装置１２ａで音声が検出されなくなったときには、録音した音声ファイルがサーバ１４に送信される。これに応じて、サーバ１４は、受信した音声ファイルを、当該録音が行われた検出時刻ｔに対応付けてメモリに記憶する。発話フラグテーブルでは、当該音声ファイルを格納した記憶位置が記憶される。なお、音声ファイルにはサーバ１４に送信される前に音声解析処理が施され、当該解析データがサーバ１４に送信されて音声解析履歴ＤＢ９２に記憶される。

【００６５】

サーバ１４は、両対話装置１２とも音声を出力していないことが検出された場合、つまり、両対話者の状態のいずれにもＳＰＥＡＫＩＮＧフラグが記憶されていないことが検出された場合、いずれかの対話者の音声ファイルが再生されずに記憶されているか否かを判定する。未再生の録音ファイルが残っている場合、つまり、ＰＬＡＹフラグの値が１以上である場合には、録音の開始された時刻の早い方の音声ファイルの再生が実行される。具体的には、サーバ１４は、当該ファイルの再生が終了するまで、音声データと再生指示と

を相手側の対話装置１２に送信する。これに応じて、対話装置１２は、受信した音声データに基づいて、当該音声を出力する。

【００６６】

このようにして、両対話者の発話が重複した場合には、後から発話し始めた側の音声を録音し、その後両方の発話が終了したときに、当該録音音声を相手側で出力することができる。なお、両発話が同時に始まった場合には優先順位に従って音声を遅延再生できる。したがって、重複したときの発話の出力を遅らせることができるので、円滑な遠隔地間対話を成立させることができる。

【００６７】

図５には、サーバ１４に記憶される発話フラグテーブルの一例が示される。発話フラグテーブルでは、検出時刻ｔごとに、ユーザ、状態フラグ、対象、保存音声ファイルの記憶位置、保存コマンドファイルの記憶位置、およびＰＬＡＹフラグ等の情報が記憶される。ユーザ情報は、当該データの主体であり、たとえばＡは当該データが対話装置１２ａの状態であることを意味し、Ｂは当該データが対話装置１２ｂの状態であることを意味する。また、対象は、ユーザの発話対象を示す。

【００６８】

状態フラグは、対話装置１２での音声取得状態および音声出力状態を示し、上述のように、ＳＰＥＡＫＩＮＧフラグ、ＳＩＬＥＮＴフラグ、ＲＥＣＯＲＤＩＮＧフラグが記憶される。なお、図５の時刻ｔ＝Ｔ＋２ΔＴでは、状態フラグはＩＮＴＥＲＰＯＬＡＴＩＮＧフラグである。上述のように、両対話者の状態フラグがＳＩＬＥＮＴフラグであった場合において、間パターンに従って間機能言葉が挿入されたときには、当該時刻の状態フラグとして、このＩＮＴＥＲＰＯＬＡＴＩＮＧフラグが上書きされるようになっている。これによって、当該検出時刻が、対話における空白時間としては計測されなくなる。

【００６９】

保存音声ファイルは、録音された音声ファイルの記憶位置を示している。たとえば、図５では、時刻Ｔ＋４ΔＴおよび時刻Ｔ＋５ΔＴにおいて、ユーザＡの状態フラグとしてＲＥＣＯＲＤＩＮＧフラグが記憶されており、当該時刻の録音に対応する音声ファイルの保存場所が示されている。また、時刻Ｔ＋４ＴでのＰＬＡＹフラグは１であり、録音が開始されたことを意味し、次の時刻Ｔ＋５ＴでのＰＬＡＹフラグは２であり、録音が継続されていることを意味し、その次の時刻Ｔ＋６ＴでのＰＬＡＹフラグは２のままであり、録音が終了されていることを意味する。

【００７０】

なお、保存コマンドファイルは、録音が行われている間に、ユーザによって当該対話装置１２ａで入力された動作コマンドを記録したファイルの記憶位置を示している。このコマンドファイルは音声ファイルと一緒に相手側対話装置１２ｂに送信され、したがって、対話装置１２ｂでは、録音した音声とともに入力指示された身振りが実行される。

【００７１】

図６から図９には、対話装置１２のＣＰＵの入力処理における動作の一例が示される。入力処理を開始すると、図６の最初のステップＳ１では、初期化が行われる。たとえば、ＳＥＮＤフラグがオフされ、ＲＥＣＯＲＤフラグがオフされ、また、時刻（またはフレーム番号）ｔに現在の時刻Ｔ（または初期値Ｔ）が代入される。続くステップＳ３から図９のステップＳ６９までの処理は一定時間ΔＴごとに、たとえば１フレームごとに繰り返し実行される。

【００７２】

ステップＳ３では、マイク１６または２０の入力をチェックし、ステップＳ５で当該入力データに基づいて、音声入力があるか否かを判断する。ステップＳ５で“ＹＥＳ”であれば、つまり、対話者が発話している場合には、ステップＳ７で、サーバ１４の発話フラグテーブルを参照する。たとえば、対話装置１２は発話フラグの要求をサーバ１４に送信する。サーバ１４はこれに応じて発話フラグテーブルのデータを当該対話装置１２に送信する。対話装置１２は発話フラグテーブルデータを受信してメモリに記憶する。なお、開

始後には音声入力の無い状態が続くので、最初の発話の前には発話フラグテーブルには両対話者の状態としてＳＩＬＥＮＴフラグが記憶されている。

【００７３】

続いて、ステップＳ９で、間計測処理を実行する。この間計測処理の動作の一例は図１０に詳細に示される。間計測処理を開始すると、図１０の最初のステップＳ８１では、発話フラグテーブルに基づいて、現時刻ｔのΔＴ前の時刻における自分の状態フラグがＳＩＬＥＮＴフラグであるか否かを判断する。このステップＳ８１では、現在の検出時刻で音声入力があり、かつ、前回の検出時刻で音声入力がなかったか否かを判断している。つまり、この対話装置１２側のユーザが話し始めたタイミングであるか否かを判断している。

【００７４】

ステップＳ８１で“ＹＥＳ”であれば、今回話し始めるまでの２種類の間を発話フラグテーブルに基づいて計測する。具体的には、ステップＳ８３で、自分が前に言葉を話し終えてから話し始めるまでの空白時間を計測する。また、ステップＳ８５で、相手が言葉を話し終えてから自分が話し始めるまでの空白時間を計測する。そして、ステップＳ８７で、計測データをサーバ１４に送信する。これに応じて、サーバ１４は間計測データを記憶する。ステップＳ８７を終了し、または、ステップＳ８１で“ＮＯ”である場合には、処理は図６のステップＳ１１に戻る。

【００７５】

このようにして、間の計測データの履歴をサーバ１４で記憶していくことによって、対話者がどのような間を取りながら対話を行っているかをサーバ１４で記録することができる。この間の履歴データと音声解析履歴データから、間のパターンを抽出することができる。

【００７６】

続いて、図６のステップＳ１１では、発話フラグテーブルに基づいて、ΔＴ前のときの対話相手のフラグがＳＰＥＡＫＩＮＧフラグであるか否かを判断する。このステップでは、相手が話しているのに、この対話装置１２側の対話者も発話をしているのか否かを判断している。ステップＳ１１で“ＹＥＳ”であれば、つまり、両対話者の発話が重複した場合には、ステップＳ１３で、ＲＥＣＯＲＤフラグはオンであるか否かを判断する。

【００７７】

ステップＳ１３で“ＮＯ”であれば、つまり、発話の重複が始まったばかりである場合には、ステップＳ１５で、音声の録音を開始し、取得した音声データを音声ファイル化してメモリに記憶する。たとえば、音声データはＰＣＭ方式データであり、音声ファイルはＷＡＶＥ形式であってよい。なお、音声データを送信前に適宜な方式で圧縮し、再生前に復号するようにしてよい。また、ステップＳ１７で、メモリのＲＥＣＯＲＤフラグをオンにして、録音中であることを記憶する。

【００７８】

なお、相手が先に話し始めている場合には、対話装置１２は相手側から音声を受信してスピーカから出力しているので、この対話装置１２側の対話者は、通常は無理に発話を続けずに、自分の発話を止めて相手の音声を聞くと考えられる。このため、録音される音声は非常に短時間のものになると考えられるので、この実施例では、音声は録音完了後に一括してサーバ１４へ送信するようにしている。しかし、他の実施例では、その都度サーバ１４に音声を送信するようにしてもよい。

【００７９】

一方、ステップＳ１３で“ＹＥＳ”であれば、つまり、既に録音を開始している場合には、ステップＳ１９で、ステップＳ１５で開始された録音を継続し、取得した音声データを音声ファイルに記憶する。

【００８０】

ステップＳ１７またはＳ１９を終了すると、ステップＳ２１で、サーバ１４の発話フラグテーブルにＲＥＣＯＲＤＩＮＧフラグを記録する。具体的には、対話装置１２は、時刻ｔ、発話者（この対話装置１２の識別情報）、対象（相手側対話装置１２の識別情報）等

の情報とともに、録音中であることを示す情報（ＲＥＣＯＲＤＩＮＧフラグ）をサーバ１４に送信する。これに応じて、サーバ１４は、受信した情報に基づいて、発話フラグテーブルに、時刻、発話者、対象およびＲＥＣＯＲＤＩＮＧフラグを記憶する。

【００８１】

なお、システム１０が３つ以上の対話装置１２を含む場合、発話の対象（相手側対話装置１２の識別情報）を入力装置の操作等によって選択できるようにしてもよい。

【００８２】

さらに、ステップＳ２３で、サーバ１４の発話フラグテーブルのＰＬＡＹフラグに、ΔＴ前の値に１を加算した値を記録する。具体的には、対話装置１２は、時刻ｔ、発話者および対象等の情報とともに、ＰＬＡＹフラグの増加指示をサーバ１４に送信する。これに応じて、サーバ１４は、発話フラグテーブルのＰＬＡＹフラグの時刻ｔの１つ前の値を読み出して、この値に１を加算し、当該算出値を時刻ｔのＰＬＡＹフラグの値として記憶する。未再生の音声ファイルが残っていない状態で録音が開始されたときは、ＰＬＡＹフラグに１が記憶され、録音が継続中である限りＰＬＡＹフラグの値は時刻ｔの進行に合わせて１つずつ増加される。ステップＳ２３を終了すると、処理は図９のステップＳ５９に進む。

【００８３】

一方、ステップＳ１１で“ＮＯ”である場合には、処理は図７のステップＳ２５に進む。つまり、この対話装置１２側の対話者が発話している場合において、１つ前の時刻で相手側が発話の無い状態、または録音中であるときは、この対話装置１２側の対話者の音声を相手に聞かせる。

【００８４】

また、ステップＳ５で“ＮＯ”である場合には、つまり、この対話装置１２側で対話者が発話していない場合には、処理は図８のステップＳ３５に進む。

【００８５】

図７のステップＳ２５では、サーバ１４の発話フラグテーブルにＳＰＥＡＫＩＮＧフラグを記録する。具体的には、対話装置１２は、時刻ｔ、発話者、対象等の情報とともに、発話中であることを示す情報（ＳＰＥＡＫＩＮＧフラグ）をサーバ１４に送信する。これに応じて、サーバ１４は、受信した情報に基づいて、発話フラグテーブルに、時刻、発話者、対象およびＳＰＥＡＫＩＮＧフラグを記憶する。

【００８６】

続くステップＳ２７で、取得した音声データを音声ファイル化して、メモリにローカルファイルとして記憶する。また、ステップＳ２７で、取得した音声データとその再生指示を相手側の対話装置１２に直接（すなわちサーバ１４を介さずに）送信する。相手側の対話装置１２は、音声データと再生指示を受信すると、当該音声データの再生処理を実行して、当該音声をスピーカ１８または２２から出力する。このようにして、この対話装置１２側のみで発話が行われている場合、あるいは相手側が録音中である場合には、この対話装置１２側の音声がローカルファイルに記録されつつ相手側に直接送信され、相手側の対話装置１２で直ちに当該音声が再生されて出力される。

【００８７】

また、ステップＳ３１では、メモリのＳＥＮＤフラグをオンにして、送信中であることを記憶する。さらに、ステップＳ３３では、サーバ１４の発話フラグテーブルのＰＬＡＹフラグに、ΔＴ前の値をそのまま記録する。具体的には、対話装置１２は、時刻ｔ、発話者および対象等の情報とともに、ＰＬＡＹフラグの維持指示をサーバ１４に送信する。これに応じて、サーバ１４は、発話フラグテーブルのＰＬＡＹフラグの時刻ｔの１つ前の値を読み出して、この値を時刻ｔのＰＬＡＹフラグの値として記憶する。このように、録音中でない場合には、ＰＬＡＹフラグの値として前回の値が維持される。ステップＳ３３を終了すると、処理は図９のステップＳ５９に進む。

【００８８】

この対話装置１２で音声入力が行われていない場合には、図８のステップＳ３５で、Ｒ

ＥＣＯＲＤフラグがオンであるか否かを判断する。ステップＳ３５で“ＹＥＳ”であれば、つまり、１つ前の時刻まで録音が行われていた場合には、ステップＳ３７で、音声ファイルへの音声の録音を終了する。また、ステップＳ３９で、音声の録音中に入力装置を用いて入力された動作コマンドのコマンドファイルへの記録を終了する。さらに、ステップＳ４１で、メモリのＲＥＣＯＲＤフラグをオフにする。

【００８９】

そして、ステップＳ４３で、録音した音声ファイルに対する音声解析処理を実行する。この音声解析処理の動作の一例が図１１に詳細に示される。なお、図８のステップＳ５３で実行される音声解析処理の動作も同じである。

【００９０】

音声解析処理を開始すると、図１１のステップＳ９１で、メモリに録音された音声ファイルを読み込む。なお、図８のステップＳ５３で実行される場合には、このステップＳ９１では、ローカルファイルの音声データを読み込む。

【００９１】

次に、ステップＳ９３で、読み込んだ音源の基本周波数（ピッチ）および振幅を算出する。また、ステップＳ９５では、音声データを音節に分割する処理を試みる。そして、ステップＳ９７で、分割した音節が存在するか否かを判断する。ステップＳ９７で“ＹＥＳ”であれば、続くステップＳ９９で、当該音節の持続時間を算出する。さらに、当該音節の持続時間の平均を算出する。ステップＳ９９を終了すると、ステップＳ９７に戻って、分割した音節が残っている場合には、当該音節についてステップＳ９９の処理を繰返す。ステップＳ９７で“ＮＯ”であれば、ステップＳ１０１で、音声解析データをサーバ１４に送信する。したがって、音声解析データは、基本周波数、振幅、および音節の平均持続時間等の情報を含む。この音声解析データは、たとえば、時刻、発話者、対象等の情報に対応付けられてサーバ１４に送信される。これに応じて、サーバ１４は、受信した音声解析データを音声解析履歴ＤＢ９２に記憶する。このようにして、発話音声の特徴が抽出されて、その履歴が記録される。ステップＳ１０１を終了すると、この音声解析処理を終了して、図８のステップＳ４５（ステップＳ４３の場合）、またはステップＳ５５（ステップＳ５３の場合）へ戻る。

【００９２】

ステップＳ４３を終了すると、ステップＳ４５で、録音した音声ファイルとコマンドファイルとをサーバ１４に送信する。音声ファイルとコマンドファイルとは、たとえば時刻、発話者、対象等の情報に対応付けられてサーバ１４に送信される。これに応じて、サーバ１４は、受信した音声ファイルとコマンドファイルとをメモリの所定領域に保存する。発話フラグテーブルでは、録音された時刻の保存音声ファイル情報として、音声ファイルの記憶位置が登録されるとともに、同時刻の保存コマンドファイル情報として、コマンドファイルの記憶位置が登録される（図５参照）。ステップＳ４５を終了すると処理はステップＳ５５に進む。

【００９３】

一方、ステップＳ３５で“ＮＯ”であれば、ステップＳ４７でＳＥＮＤフラグがオンであるか否かを判断する。ステップＳ４７で“ＹＥＳ”であれば、つまり、１つ前の時刻まで音声を相手側の対話装置１２に送信していた場合には、ステップＳ４９で、音声のローカルファイルへの記録を終了し、ステップＳ５１で、メモリのＳＥＮＤフラグをオフにする。そして、ステップＳ５３で、ローカルファイルの音声データに対して、上述のような図１１の音声解析処理を実行する。ステップＳ５３を終了すると、または、ステップＳ４７で“ＮＯ”である場合には、処理はステップＳ５５へ進む。

【００９４】

ステップＳ５５では、サーバ１４の発話フラグテーブルにＳＩＬＥＮＴフラグを記録する。具体的には、対話装置１２は、時刻ｔ、発話者、対象等の情報とともに、音声入力が無いことを示す情報（ＳＩＬＥＮＴフラグ）をサーバ１４に送信する。これに応じて、サーバ１４は、受信した情報に基づいて、発話フラグテーブルに、時刻、発話者、対象およ

びＳＩＬＥＮＴフラグを記憶する。

【００９５】

また、ステップＳ５７で、図７のステップＳ３３と同様にして、サーバの発話フラグテーブルのＰＬＡＹフラグに、ΔＴ前の値をそのまま記録する。ステップＳ５７を終了すると、処理は図９のステップＳ５９へ進む。

【００９６】

図９のステップＳ５９からＳ６７では、相手側の対話装置１２がロボット１２ｂである場合の処理である。したがって、相手側がロボット１２ｂでない場合には、これらの処理は行われなくてよい。

【００９７】

図９のステップＳ５９では、動作コマンドの入力をチェックする。具体的には、入力装置からの入力データを取得して、ロボット１２ｂの身振りのための動作コマンドが選択されたか否かを判定する。たとえば、動作コマンドはディスプレイに選択可能なリストとして表示されてよい。なお、この対話装置１２がロボット１２ｂである場合には、入力装置とディスプレイを設ける必要がある。

【００９８】

そして、ステップＳ６１で、選択された動作コマンドがあるかどうかを判断し、“ＹＥＳ”であれば、ステップＳ６３で、メモリのＲＥＣＯＲＤフラグがオンであるか否かを判断する。ステップＳ６３で“ＹＥＳ”であれば、つまり、音声録音中の場合には、ステップＳ６５で、動作コマンドをメモリのコマンドファイルに記憶する。このように、音声を録音している場合には、動作コマンドの入力も同時に記録して、録音終了後に上述のステップＳ４５でサーバ１４に送信するようにしているので、両対話者の発話が重複した場合には、発話と身振りに対して同時に遅延を与えてから相手側で再生することができる。

【００９９】

一方、ステップＳ６３で“ＮＯ”であれば、ステップＳ６７で、動作コマンドと再生指示とを相手側の対話装置１２ｂに直接送信する。相手側の対話装置１２ｂは、動作コマンドと再生指示を受信すると、当該動作コマンドに対応するプログラムおよびデータに従って動作し、その身振りを実行する。

【０１００】

ステップＳ６５またはＳ６７を終了したとき、またはステップＳ６１で“ＮＯ”の場合には、ステップＳ６９で、所定時間ΔＴ（たとえば１フレーム）を加算することで時刻（あるいはフレーム番号）ｔを更新する。そして、図６のステップＳ３に戻って、次の時刻ｔにおける処理を繰返す。このようにして、対話装置１２では、この対話装置１２側の対話者の発話の状態および相手側の発話の状態に応じた処理が実行される。

【０１０１】

図１２にはサーバ１４の継続促進処理における動作の一例が示される。また、図１３には、サーバ１４の遅延再生処理における動作の一例が示される。

【０１０２】

なお、サーバ１４の他の処理、たとえば受信処理、発話フラグテーブルの作成処理および送信処理などのフロー図は省略する。サーバ１４は上述のような各処理を並列的に実行している。サーバ１４は、上述のように、対話装置１２からデータを受信したときは、当該データをメモリに記憶し、必要に応じて当該データに対応する所定の処理を実行する。たとえば、サーバ１４は、対話装置１２から発話や処理の状態に関するデータを受信したときは発話フラグテーブルを作成する。音声ファイルおよび動作コマンドファイル等を受信したときは、これらのファイルを記憶するとともに、発話フラグテーブルに記憶位置を書き込む。音声解析データを受信したときは、当該データを音声解析履歴ＤＢ９２に記憶する。また、対話装置１２から発話フラグテーブルの要求があったときは、当該対話装置１２に発話フラグテーブルを送信する。

【０１０３】

図１２に示す継続促進処理では、サーバ１４のＣＰＵは、ステップＳ１１１で初期化を

実行し、たとえば変数ｔに初期値Ｔを設定する。この初期値Ｔは発話フラグテーブルの時刻ｔの最初の値Ｔであり、つまり、対話装置１２における時刻ｔの初期値Ｔである。したがって、継続促進処理は発話フラグテーブルの作成後に実行される。続くステップＳ１１３からＳ１３５の処理をサーバ１４のＣＰＵは一定時間ΔＴごとに、たとえば１フレームごとに繰り返し実行する。

【０１０４】

ステップＳ１１３では、メモリの発話フラグテーブルを参照する。たとえば現時刻ｔのデータを読み出す。そして、ステップＳ１１５で、対話者同士でＳＩＬＥＮＴフラグであるか否かを判断する。たとえば、現時刻ｔにおいてユーザと発話対象が互いに対になっている両対話者が存在しており、かつ、当該両対話者の状態フラグがＳＩＬＥＮＴフラグであることを判定する。たとえば図５では、時刻Ｔ＋ΔＴのときがこの状態に相当する。

【０１０５】

ステップＳ１１５で“ＹＥＳ”であれば、つまり、対話において無音状態になっている場合には、ステップＳ１１７で、空白時間を算出する。たとえば、現時刻ｔ以前の発話フラグテーブルのデータを読み出して、現時刻ｔから遡って両対話者のどちらかの状態フラグがＳＩＬＥＮＴフラグでなくなるまでに掛かった時間（またはフレーム数）を算出する。

【０１０６】

続いて、ステップＳ１１９で、音声解析履歴ＤＢ９２から対話者らの最新のデータを抽出する。具体的には、現時刻ｔに最も近い時刻の発話者の音声解析データから、基本周波数、振幅および音節の平均持続時間等を読み出す。このように、ステップＳ１１７とＳ１１９で、少なくとも空白時間と当該空白前の発話音声の特徴を含む間の状況が検出される。

【０１０７】

そして、ステップＳ１２１で、現在の間の状況と間パターンとの照合を実行して、ステップＳ１２３で、現在の対話の間の状況にマッチする間パターンがあるか否かを判断する。上述の図４のように、間パターンデータ内には、空白時間（ｔ）および条件（Ｉ）設定されているので、このような間パターンに合う空白時間および発話音声の特徴（基本周波数、振幅、音節の平均持続時間など）を有する間の状況（すなわち、最終発話者の発話後の無音状態）が生じているか否かを判定する。マッチする間パターンがある場合には、当該間パターンに対応する間機能言葉を選択する。また、間パターンデータに設定されている最終発話者と発話者との関係（相手か自分か）に基づいて、間機能言葉を発話させる対話装置１２を特定する。

【０１０８】

ステップＳ１２３で“ＹＥＳ”であれば、つまり、現在の対話における間の状況が、間パターンに基づく間を挿入すべき状況になっていると判定される場合には、ステップＳ１２５で、選択した間機能言葉の音声ファイルをメモリの作業領域に読み出して、ピッチ、抑揚パターンを調整して、当該調整した間機能言葉の音声ファイルを生成する。これによって、発話者の発話の特徴（たとえば、高揚した口調、淡々とした発話など）に合わせた間機能言葉を出力することが可能になる。したがって、会話に合成音声が挿入されても対話者に違和感をさほど覚えさせないようにすることができるし、また、それまでの会話の調子や流れを継続させることができる。

【０１０９】

また、ステップＳ１２７で、選択された間機能言葉に適した動作コマンドを選択する。この実施例では、間パターンデータにおいて、間機能言葉に対応する動作コマンドが登録されているので、当該動作コマンドを選択する。

【０１１０】

そして、ステップＳ１２９で、音声ファイルと動作コマンドファイルを、発話させる対話装置１２に送信する。ファイル送信後、ステップＳ１３１で、音声と動作の再生指示を同じ対話装置１２に送信する。これによって、対話における無音領域に言葉や身振りを挿

入することができる。なお、その対話装置１２は、音声ファイルの再生を実行し、当該音声を出力する。また、対話装置１２がロボット１２ｂである場合には、さらに動作コマンドの再生を実行し、当該動作コマンドに対応する身振りを行う。

【０１１１】

さらに、ステップＳ１３３で、発話フラグテーブルにおいて、現時刻ｔの状態フラグにＩＮＴＥＲＰＯＬＡＴＩＮＧフラグを上書きする（図５の時刻Ｔ＋２ΔＴを参照）。これによって、以降のステップＳ１１７では、当該時刻ｔが無音であるとは見なされないようにすることができる。

【０１１２】

一方、ステップＳ１２３で“ＮＯ”である場合には、つまり、未だ、間パターンに従った間を与える必要がない場合には、処理はそのままステップＳ１３５に進む。ステップＳ１３５では、所定時間ΔＴ（たとえば１フレーム）を加算することで時刻（あるいはフレーム番号）ｔを更新する。なお、このサーバ１４におけるΔＴは対話装置１２におけるΔＴと同一である。そして、ステップＳ１１３に戻って、次の時刻ｔにおける処理を繰返す。このようにして、対話において無音が検出された場合には、必要に応じて言葉や身振りを挿入することによって、無音時間を適切な間に変えることができる。

【０１１３】

図１３に示す遅延再生処理では、サーバ１４のＣＰＵは、ステップＳ１５１で初期化を実行する。たとえば、ＰＬＡＹＩＮＧフラグをオフにする。ＰＬＡＹＩＮＧフラグは録音された音声ファイルおよび動作コマンドファイルを再生中であるか否かを示す。また、図１２の継続促進処理と同様に、変数ｔに初期値Ｔを設定する。この初期値Ｔは発話フラグテーブルの時刻ｔの最初の値Ｔであり、つまり、対話装置１２における時刻ｔの初期値Ｔである。したがって、この遅延再生処理も発話フラグテーブルの作成後に実行される。続くステップＳ１５３からＳ１７９の処理をサーバ１４のＣＰＵは一定時間ΔＴごとに、たとえば１フレームごとに繰り返し実行する。

【０１１４】

ステップＳ１５３では、メモリの発話フラグテーブルを参照する。ステップＳ１５５で、メモリのＰＬＡＹＩＮＧフラグがオンであるか否かを判断する。ステップＳ１１５で“ＮＯ”であれば、つまり、再生中ではない場合には、ステップＳ１５７で、現時刻ｔにおける両対話者のどちらかの状態フラグとしてＳＰＥＡＫＩＮＧフラグがあるか否かを判断する。ステップＳ１５７で“ＹＥＳ”の場合、一方が発話をしており、その音声が他方の対話装置１２から出力されているはずである。したがって、遅延再生は行わず処理はステップＳ１７９に進む。

【０１１５】

一方、ステップＳ１５７で“ＮＯ”であれば、つまり、両対話装置１２で音声が出力されていない場合には、ステップＳ１５９で、両対話者のどちらかのＰＬＡＹフラグが１以上であるか否かを判断する。ステップＳ１５９で“ＮＯ”であれば、録音されたが未再生である音声ファイルが存在しないので、処理はそのままステップＳ１７９に進む。

【０１１６】

しかし、ステップＳ１５９で“ＹＥＳ”であれば、つまり、録音されたが未再生の音声ファイルが残っている場合には、ステップＳ１６１で、ＰＬＡＹフラグが１である時刻ｔが早いユーザを発話フラグテーブルから参照する。つまり、録音を開始した時刻が早いユーザを特定する。なお、録音の開始が両対話者で同時刻である場合には、予め設定しておいた優先順位（たとえばＢ＞Ａ）に基づいて、ユーザを特定する。

【０１１７】

続いて、ステップＳ１６４で、再生のための設定を実行し、変数Ｆに１を設定し、変数Ｕに特定したユーザを設定する。変数Ｆは音声再生のためのフレームカウンタである。また、ステップＳ１６５で、メモリのＰＬＡＹＩＮＧフラグをオンにして、再生中であることを記憶する。そして、ステップＳ１６７で、変数ＵのＰＬＡＹフラグが変数Ｆの値である音声および動作を再生する。具体的には、当該音声ファイルを読み出して、当該ユーザ

の相手側の対話装置１２に音声ファイルと再生指示とを送信する。なお、当該動作コマンドファイルも保存されている場合には、当該動作コマンドファイルも読み出して、音声ファイルと一緒に相手側の対話装置１２に送信する。これに応じて、当該対話装置１２は、音声ファイルおよび動作コマンドファイルを記憶するとともに、その再生を実行する。これによって、音声がスピーカ１８または２２から出力され、動作コマンドもあった場合には、当該身振りも実行される。このようにして、録音されていた音声および記憶されていた動作の再生が開始される。

【０１１８】

ステップＳ１６７を終了すると、処理はステップＳ１７９へ進む。ステップＳ１７９では、時刻ｔに所定時間ΔＴが加算されて時刻ｔが更新される。ステップＳ１７９を終了すると、処理はステップＳ１５３へ戻って、次の時刻ｔにおける処理を繰返す。

【０１１９】

再生が開始されると、ステップＳ１５５で“ＹＥＳ”と判断され、続くステップＳ１６９で、時刻ｔにおける変数ＵのＰＬＡＹフラグが変数Ｆの値に等しいか否かを判断する。上述のように、録音が終了した場合には、ＰＬＡＹフラグの値は前時刻の値を維持するので、このステップＳ１６９では、再生中の音声ファイルの再生を完了したか否かを判定している。

【０１２０】

ステップＳ１６９で“ＮＯ”であれば、つまり、音声ファイルの再生が未だ完了していない場合には、ステップＳ１７１で、変数Ｆをインクリメントする。その後、ステップＳ１７３で、変数ＵのＰＬＡＹフラグが変数Ｆの値である音声および動作を再生する。これによって、上述のステップＳ１６７と同様にデータが送信され、次のフレームの音声および動作が対話装置１２で再生される。ステップＳ１７３を終了すると、処理はステップＳ１７９へ進む。

【０１２１】

一方、ステップＳ１６９で“ＹＥＳ”であれば、つまり、音声ファイルの再生を完了した場合には、ステップＳ１７５で、メモリのＰＬＡＹＩＮＧフラグをオフにする。また、ステップＳ１７７で、変数ＵのＰＬＡＹフラグの値を全て変数Ｆの値だけ減算する。なお、減算の結果、値が負になったとき、当該ＰＬＡＹフラグの値は０に設定される。これによって、再生された変数Ｕおよび時刻ｔのＰＬＡＹフラグの値がすべて０になる。また、当該変数Ｕのユーザの未再生の音声ファイルが存在する場合には、当該ユーザの最も古く録音された音声ファイルのうち最も早い時刻のＰＬＡＹフラグの値が１になる。したがって、次回は、当該未再生の音声を再生することが可能になる。ステップＳ１７７を終了すると、処理はステップＳ１７９へ進む。

【０１２２】

このようにして、両対話者の発話の重複によって録音された音声および記録された動作コマンドを、後から再生することができる。

【０１２３】

図１４には、対話装置１２のＣＰＵの出力処理の動作の一例が示される。この出力処理は上述の図６から図９の入力処理と並列的に実行される。また、この出力処理は一定時間ごと、たとえば１フレームごとに繰り返し実行される。

【０１２４】

ステップＳ１９１では、音声を受信したか否かが判断され、“ＹＥＳ”であれば、ステップＳ１９３で、受信した音声ファイルないし音声データをメモリに記憶する。

【０１２５】

続いて、ステップＳ１９５では、動作コマンドを受信したか否かが判断され、“ＹＥＳ”であれば、ステップＳ１９７で、受信した動作コマンドファイルをメモリに記憶する。

【０１２６】

続いて、ステップＳ１９９では、再生指示を受信したか否かが判断され、“ＹＥＳ”であれば、ステップＳ２０１で、音声を再生する。具体的には、対話装置１２のＣＰＵは、

受信した音声ファイルを再生を開始し、当該音声データを音声入出力ボードに与えてスピーカから当該音声を出力する。また、当該対話装置１２が身体動作機能を有する対話装置１２ｂである場合には、ステップＳ２０３で、動作を再生する。具体的には、当該動作コマンドに従って対応する身振りを実行する。動作コマンドに対応する身振りを実行するためのプログラムおよび制御データは、対話装置１２ｂのメモリ７０に予め記憶されている。ＣＰＵ６６は動作コマンドに対応するプログラムに従って制御データをモータ制御ボード７２に与えて、対応するモータを制御する。これによって対応する身体部位が動かされて所定の身振りが表現される。

【０１２７】

なお、上述の実施例では、両対話者の発話が重複したとき、後から発話された方の音声を録音して、その後どちらも発話しなくなってから、当該録音音声を相手側で出力するようにしていた。しかし、他の実施例では、両対話者の発話が重複したときには、後から発話された方の音声をキャンセルするようにしてもよい。

【０１２８】

また、上述の各実施例では、間機能言葉の音声データをサーバ１４が記憶しておいて、サーバ１４から対話装置１２に送信するようにしていた。しかし、他の実施例では、間機能言葉の音声データを各対話装置１２に予め記憶させておいて、サーバ１４から再生すべき間機能言葉を指定する情報を送信するようにしてもよい。

【０１２９】

また、上述の各実施例では、システム１０は、身体動作機能を有しない対話装置１２ａと身体動作機能を有する対話装置１２ｂとを含んでいた。しかし、他の実施例では、身体動作機能を有しない対話装置１２ａのみが使用されてよく、この場合には、動作コマンド関連の処理が不要である。逆に、身体動作機能を有する対話装置１２ｂのみが使用されてもよい。

【０１３０】

また、上述の各実施例では、システム１０は対話装置１２とは別に各対話装置１２の音声取得状態および音声出力状態を示す情報（すなわち発話フラグテーブル）を管理するサーバ１４を備えた。しかし、他の実施例では、サーバ１４を別途に設けずに、サーバ１４の機能（発話フラグテーブルの管理、継続促進処理、遅延再生処理など）を一方の対話装置１２に備えさせるようにしてよいし、あるいは２つの対話装置１２に分散して備えさせるようにしてもよい。

【図面の簡単な説明】

【０１３１】

【図１】この発明の一実施例の遠隔地間対話システムの構成を示す図解図である。
【図２】身体動作機能を有する対話装置の外観の一例を示す図解図である。
【図３】図２の対話装置の電気的な構成の一例を示すブロック図である。
【図４】間パターンＤＢに記憶される間パターンデータの一例を示す図解図である。
【図５】サーバに記憶される発話フラグテーブルの一例を示す図解図である。
【図６】対話装置の入力処理の動作の一例の一部を示すフロー図である。
【図７】図６の続きの一部を示すフロー図である。
【図８】図６の続きの一部を示すフロー図である。
【図９】図６、図７および図８の続きを示すフロー図である。
【図１０】図６の間計測処理の動作の一例を示すフロー図である。
【図１１】図８の音声解析処理の動作の一例を示すフロー図である。
【図１２】サーバの継続促進処理の動作の一例を示すフロー図である。
【図１３】サーバの遅延再生処理の動作の一例を示すフロー図である。
【図１４】対話装置の出力処理の動作の一例を示すフロー図である。

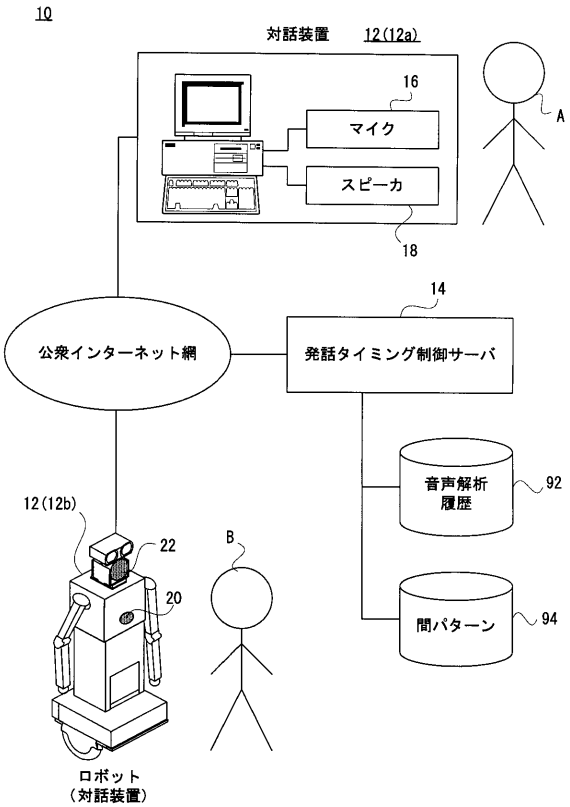
【符号の説明】

【０１３２】

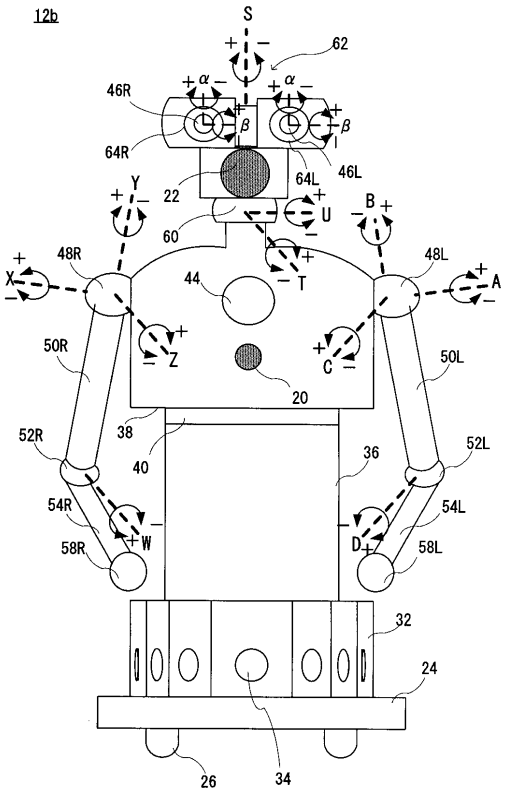
１０　…遠隔地間対話システム

１２，１２ａ，１２ｂ　…対話装置
１４　…発話タイミング制御サーバ
１６，２０　…マイク
１８，２２　…スピーカ
９２　…音声解析履歴データベース
９４　…間パターンデータベース

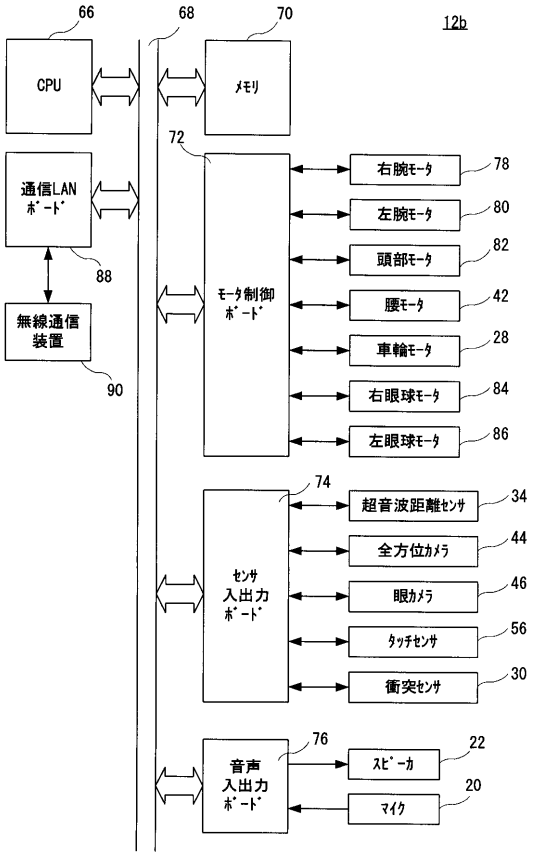
【図１】



【図２】



【図３】



【図４】

間パターンデータ

空白時間(t) [sec]	最後の発話者	条件 (I)	間機能言葉	発話者	動作コマンド
1<t≦1.5	A	ω>150, 1000<A≦1500, 30<m≦50	うんうん	B	うなづく
0.5<t≦2	B	ω>10, 50<m≦100	うーん	B	首傾げる
t≦2	A	ω>300, m≦30	はいはい	B	うなづく
t>2	B	ω>80, A≦1500	えーと	A	視線を上方に向ける
⋮	⋮	⋮	⋮	⋮	⋮

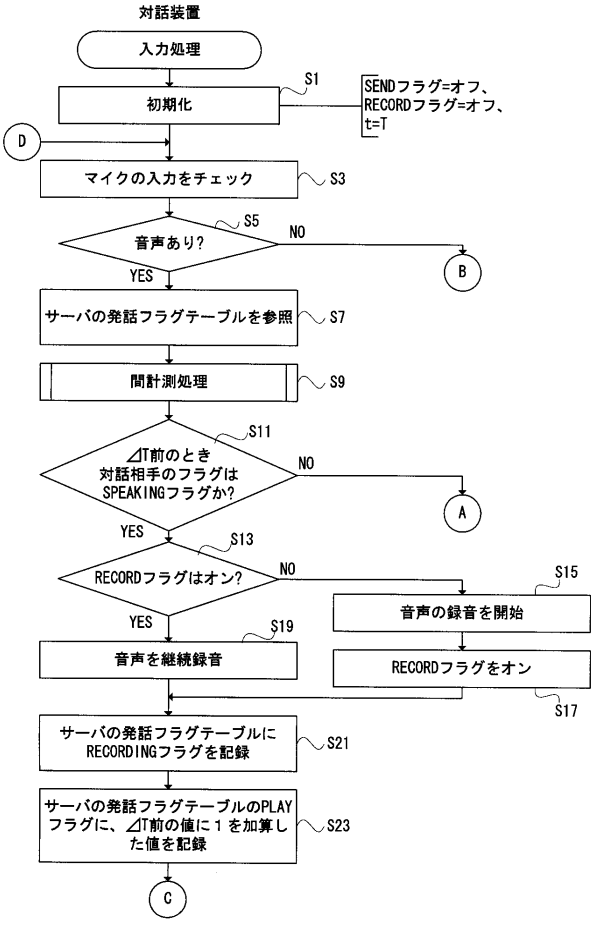
ω：基本周波数[Hz],
A：振幅,
m：音節の平均持続時間[msec]

【図５】

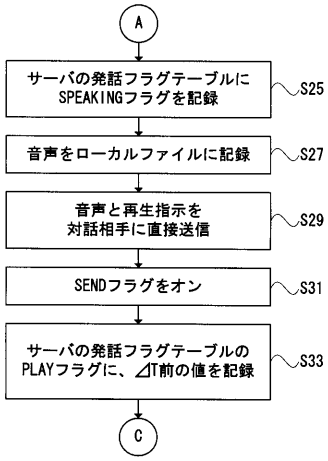
発話フラグテーブル

時刻t	ユーザ	フラグ	対象	保存 音声ファイル	保存 コマンド ファイル	PLAY フラグ
T	A	SPEAKING	B			0
T	B	SILENT	A			0
T+⊿T	A	SILENT	B			0
T+⊿T	B	SILENT	A			0
T+2⊿T	A	INTERPOLATING	B			0
T+2⊿T	B	INTERPOLATING	A			0
T+3⊿T	A	SILENT	B			0
T+3⊿T	B	SPEAKING	A			0
T+4⊿T	A	RECORDING	B	D:¥voice¥v.wav	D:¥command¥g.dat	1
T+4⊿T	B	SPEAKING	A			0
T+5⊿T	A	RECORDING	B	D:¥voice¥v.wav	D:¥command¥g.dat	2
T+5⊿T	B	SPEAKING	A			0
T+6⊿T	A	SILENT	B			2
T+6⊿T	B	SPEAKING	A			0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

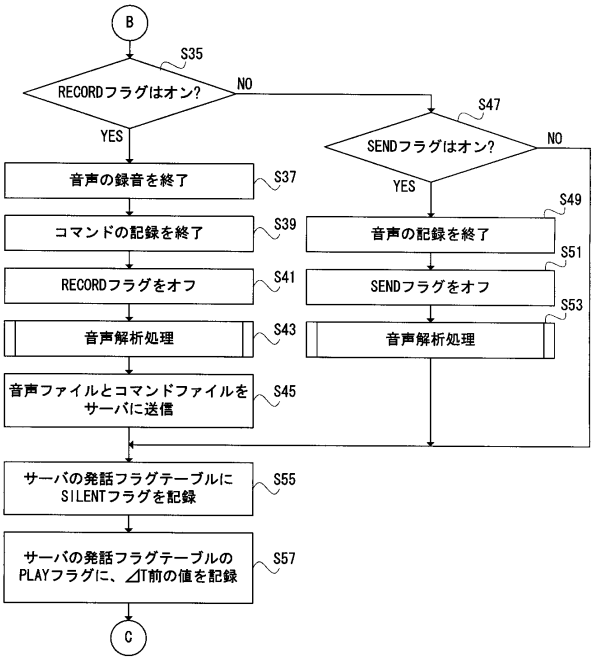
【図６】



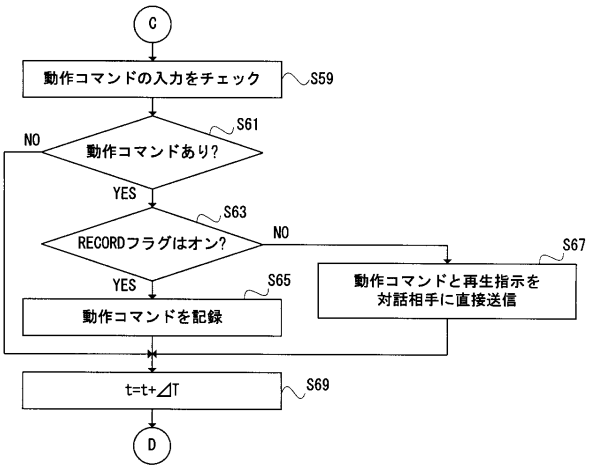
【図７】


A
サーバの発話フラグテーブルに
SPEAKINGフラグを記録
S25
音声をローカルファイルに記録
S27
音声と再生指示を
対話相手に直接送信
S29
SENDフラグをオン
S31
サーバの発話フラグテーブルの
PLAYフラグに、⊿T前の値を記録
S33
C

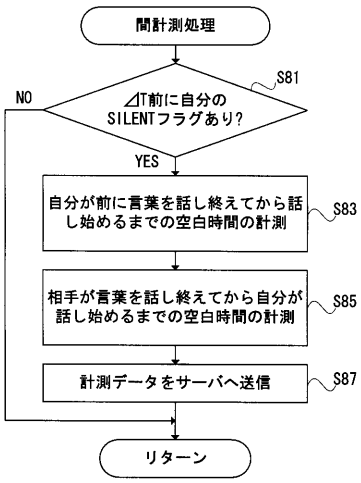
【図８】


B
S35
RECORDフラグはオン?
NO
YES
S47
SENDフラグはオン?
NO
YES
音声の録音を終了
S37
コマンドの記録を終了
S39
RECORDフラグをオフ
S41
音声解析処理
S43
音声ファイルとコマンドファイルを
サーバに送信
S45
S49
音声の記録を終了
S51
SENDフラグをオフ
S53
音声解析処理
サーバの発話フラグテーブルに
SILENTフラグを記録
S55
サーバの発話フラグテーブルの
PLAYフラグに、⊿T前の値を記録
S57
C

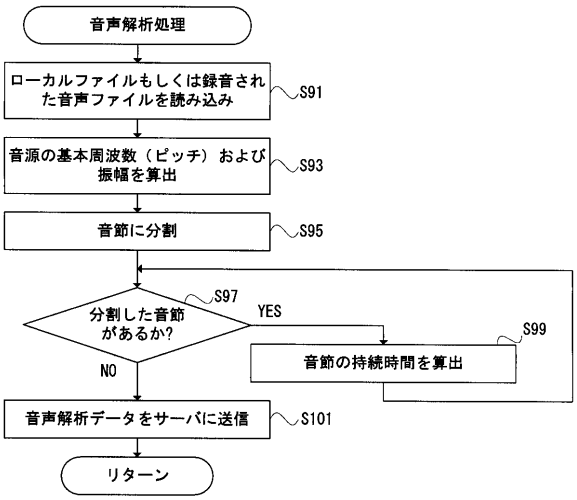
【図９】


C
動作コマンドの入力をチェック
S59
S61
NO
動作コマンドあり?
YES
S63
RECORDフラグはオン?
NO
YES
S65
動作コマンドを記録
S67
動作コマンドと再生指示を
対話相手に直接送信
t=t+⊿T
S69
D

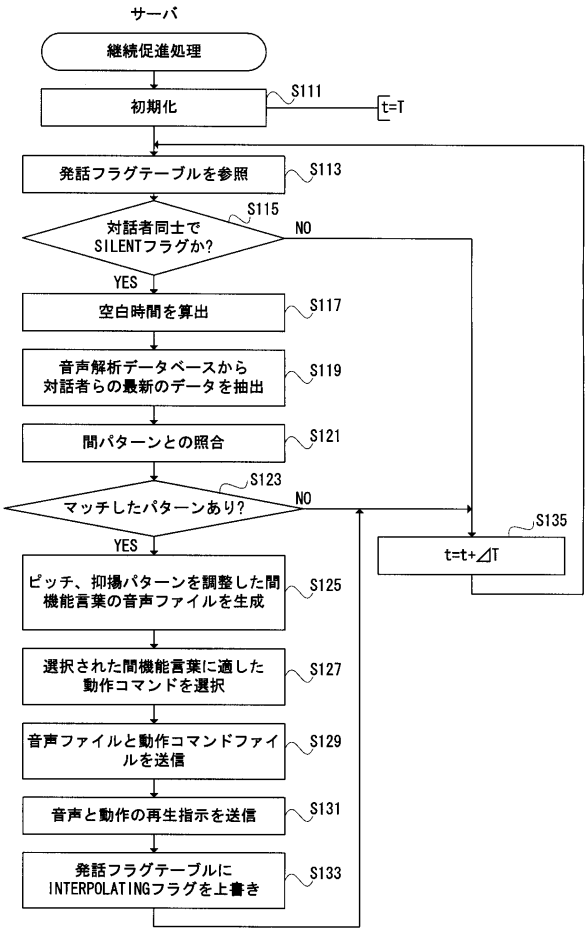
【図１０】


間計測処理
S81
NO
⊿T前に自分の
SILENTフラグあり?
YES
自分が前に言葉を話し終えてから話
し始めるまでの空白時間の計測
S83
相手が言葉を話し終えてから自分が
話し始めるまでの空白時間の計測
S85
計測データをサーバへ送信
S87
リターン

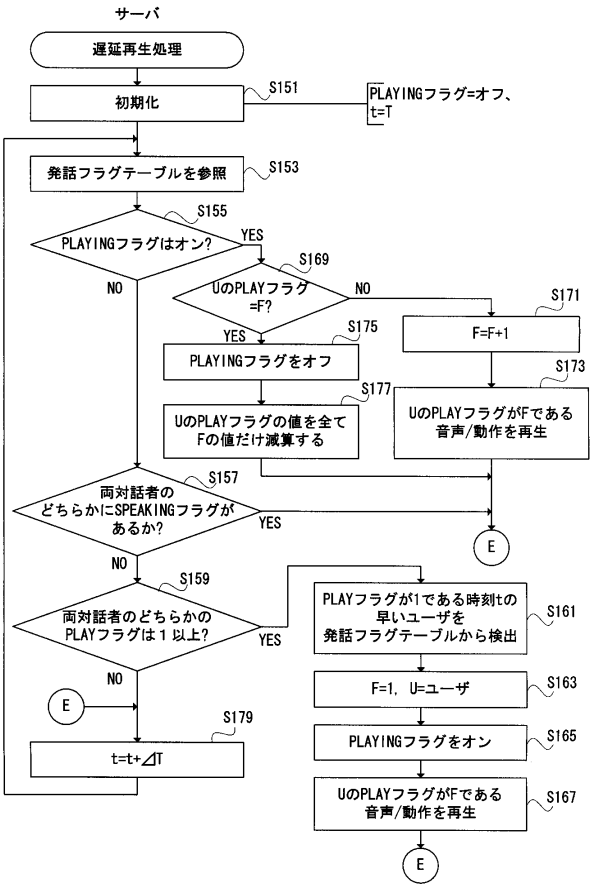
【図１１】



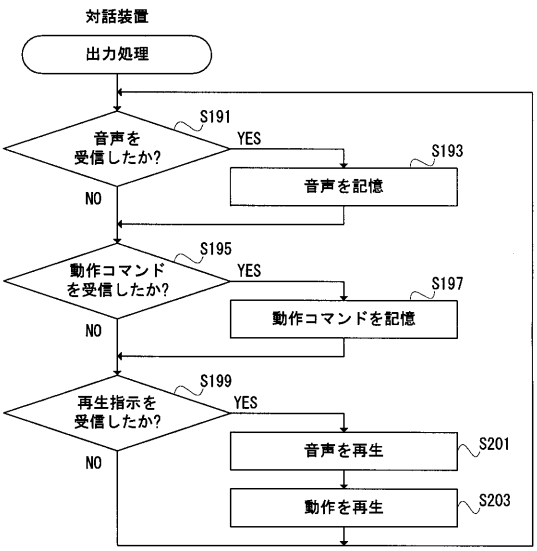
【図１２】



【図１３】



【図１４】



フロントページの続き

(72)発明者　宮下　敬宏
京都府相楽郡精華町光台二丁目２番地２　株式会社国際電気通信基礎技術研究所内
(72)発明者　石黒　浩
京都府相楽郡精華町光台二丁目２番地２　株式会社国際電気通信基礎技術研究所内

審査官　松元　伸次

(56)参考文献　特開２００４－０６４２８１（ＪＰ，Ａ）
特開２００５－０１２８１９（ＪＰ，Ａ）
特開２００３－３０４３０７（ＪＰ，Ａ）
特開２００１－１３４２８９（ＪＰ，Ａ）
特開２０００－３４９９２０（ＪＰ，Ａ）
特開昭６３－２８７１２６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｍ　１／００、１／２４－１／２５３、
１／５８－１／６２、１／６６－３／００、
３／１６－３／２０、３／３８－３／５８、
７／００－７／１６、１１／００－１１／１０、９９／００