

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2003-223188
(P2003-223188A)

(43) 公開日 平成15年8月8日(2003.8.8)

(51) Int.Cl. ⁷	識別記号	F I	テマコード*(参考)
G 1 0 L 15/28		G 0 6 F 3/16	3 2 0 H 5 D 0 1 6
G 0 6 F 3/16	3 2 0		3 4 0 A
	3 4 0	G 1 0 L 3/00	5 7 1 A
G 1 0 L 15/00			5 5 1 A
			5 6 1 H
審査請求 未請求 請求項の数18 O L (全 17 頁)			

(21) 出願番号 特願2002-19457(P2002-19457)

(22) 出願日 平成14年1月29日(2002.1.29)

(71) 出願人 000003078

株式会社東芝

東京都港区芝浦一丁目1番1号

(72) 発明者 蟻生 政秀

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(74) 代理人 100083161

弁理士 外川 英明

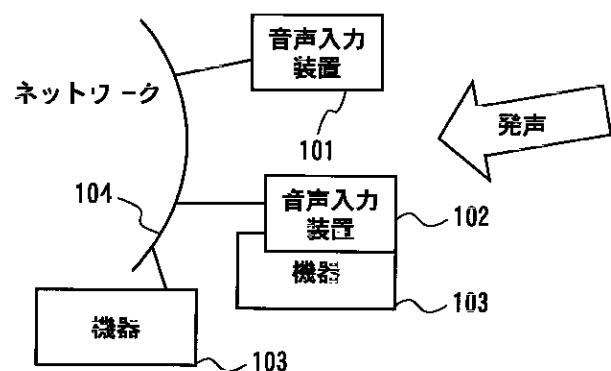
Fターム(参考) 5D015 KK02 LL05 LL07

(54) 【発明の名称】 音声入力システム、音声入力方法及び音声入力プログラム

(57) 【要約】

【課題】 ユーザの発声が複数の音声入力装置に入りうる場合にユーザに負担をかけずに発声に対する処理をきめることができる音声入力システム、音声入力方法及びそのプログラムを提供する。

【解決手段】 複数の音声入力装置がネットワークを通して接続されており、これら音声入力装置同士が情報の授受を行い、音声が入力された場合には決めておいた情報を他の音声入力装置に送り、各音声入力装置は自装置での音声の処理結果と他の音声入力装置の処理結果から、その音声に対する処理の判断を行うようにすることでユーザに負担をかけずに音声入力への処理を行うことを可能とする。



【特許請求の範囲】

【請求項1】複数の音声入力装置がネットワークに接続され、前記ネットワークに接続された各々の音声入力装置において入力される音声情報をそれぞれ検知し、前記ネットワークに接続された各々の音声入力装置で前記検知した前記音声に関する情報を判断情報として前記ネットワークを介して他の音声入力装置に授受し、前記ネットワークに接続された各々の音声入力装置が自音声入力装置と他の音声入力装置における前記判断情報をもとに前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする音声入力システム。

【請求項2】前記ネットワークに接続された複数の音声入力装置は所定のルールに基づいて順位関係を構成し、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記順位情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項3】前記ネットワーク上に接続された複数の音声入力装置は所定のルールに基づいて複数にグループ化され、前記グループに関する情報を記憶する領域を備え、前記グループに関する記憶領域は、ネットワークに接続された音声入力装置に関する記憶領域と関連付けて作業を行う仕組みを備え、取り込まれた音声情報に対して前記グループに関する記憶領域の情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項4】前記ネットワークに接続された複数の音声入力装置は共通の時間情報を持ち、前記ネットワークに接続された各々の音声入力装置で前記音声情報を検知したらその検知時間情報を他の音声入力装置に伝え、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記検知時間情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項5】前記ネットワークに接続された複数の音声入力装置は前記入力される音声情報の音量について共通の尺度を備え、前記ネットワークに接続された各々の音声入力装置で前記音声情報を検知したら前記音量情報を他の音声入力装置に伝え、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記音量情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項6】前記ネットワークに接続された複数の音声入力装置は入力される音声情報の信号対雑音比を推定する手段を備え、前記ネットワークに接続された各々の音声入力装置で前記音声情報を検知したらその信号対雑音

比情報を他の音声入力システムに伝える手段を備え、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記信号対雑音比情報音量情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項7】前記ネットワークに接続された複数の音声入力装置は過去の状態履歴を保存しておく記憶領域を備え、前記ネットワークに接続された各々の音声入力装置は前記過去の状態履歴を他の音声入力装置に伝え、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記過去の状態履歴を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項8】前記ネットワークに接続された複数の音声入力装置は音響的な情報や言語的な情報を保存する手段及び取り込まれた音声情報と前記音響的な情報や言語的な言語情報を使ってマッチングを行って音声認識をする手段を備え、前記ネットワークに接続された各々の音声入力装置は音声情報を検知したら、この音声情報に基づく音声認識情報を他の音声入力装置に伝え、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記音声認識情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項9】前記ネットワークに接続された複数の音声入力装置は前記音声情報の音源に関して識別できる手段とを備え、前記ネットワークに接続された各々の音声入力装置は音声情報を検知したらその音源に関する情報を他の音声入力装置に伝え、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記音源に関する情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項10】前記ネットワークに接続された複数の音声入力装置は各々の音声入力装置及び他の音声入力装置を制御できる共通の制御命令体系を備え、取り込まれた音声に対してネットワーク上の他の音声入力装置へ前記制御命令を送信及び他の音声入力装置からの前記制御命令の受信を行って、その命令内容を実行することを特徴とする請求項1記載の音声入力システム。

【請求項11】前記ネットワークに接続された複数の音声入力装置は各々の音声入力装置と他の音声入力装置以外のネットワークで接続された制御可能機器の情報を記憶する領域を備え、取り込まれた音声情報に対して、前記他の制御可能機器情報も利用することで、入力音声情報の処理及び制御可能機器との情報の授受を行うことを特徴とする請求項1記載の音声入力システム。

【請求項12】前記ネットワークに接続された各々の音声入力装置には音声入力手段以外のセンサ・デバイス及びその信号処理手段を備え、取り込まれた音声情報に対して、前記他のセンサ・デバイス情報を前記判断情報として前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とする請求項1記載の音声入力システム。

【請求項13】前記ネットワークに接続された前記複数の音声入力装置は各々の音声入力装置及び他の音声入力装置の情報をもとに検知した音声情報の処理を判断する基準を他の音声入力装置の情報を利用して変更する、あるいはユーザの設定によって変更することが可能なことを特徴とする請求項1記載の音声入力システム。

【請求項14】前記ネットワークに接続された複数の音声入力装置は各々のシステムの状態を表示する表出部を備え、取り込まれた音声情報やシステムがユーザに伝えたい情報についてどのように表出するかを制御する機能を備え、処理の決定及び実行と情報の表出を行うことを特徴する請求項1記載の音声入力システム

【請求項15】前記ネットワークに接続された複数の音声入力装置はその一部乃至全ての手段が前記音声入力装置以外の機能における手段と共有化されていることを特徴とする請求項1記載の音声入力システム。

【請求項16】前記ネットワークに接続された複数の音声入力装置の機能の一部が物理的には分かれていても、ネットワーク技術による結びつきで一つとして機能することができることを特徴とする請求項1記載の音声入力システム。

【請求項17】ネットワークに接続された複数の音声入力装置において入力される音声情報をそれぞれ検知するステップと、前記ネットワークに接続された各々の音声入力装置で前記検知した前記音声に関する情報を判断情報として前記ネットワークを介して他の音声入力装置に授受するステップと、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記判断情報をもとに前記音声情報に対する処理の決定及び実行の判断を行うステップとを含むことを特徴とする音声入力方法。

【請求項18】ネットワークに接続された複数の音声入力装置において入力される音声情報をそれぞれ検知し、前記ネットワークに接続された各々の音声入力装置で前記検知した前記音声に関する情報を判断情報として前記ネットワークを介して他の音声入力装置に授受し、前記ネットワークに接続された各々の音声入力装置は自音声入力装置と他の音声入力装置における前記判断情報をもとに前記音声情報に対する処理の決定及び実行の判断を行う機能を実現することを特徴とする音声入力プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声を扱う装置に関するものであり、特にユーザの発声が複数の音声入力に入りうる場合の音声入力システム、音声入力方法及び音声入力プログラムに関する。

【0002】

【従来の技術】これまでは音声によって機器を制御する場合や、音声をある機器に入力する場合にはユーザと音声入力機器は1対1で対応していることを主に想定していた。しかしながら、例えば一つの部屋に複数の音声入力装置がある場合などユーザの発声が複数の音声入力装置に入ってしまうことは十分あり得る。その場合に従来は、ユーザが特に対象機器を指定したり、音声入力しようと思っている機器以外に対しては音声入力を抑制するような操作を行ったりする必要があった。

【0003】

【発明が解決しようとする課題】本発明は ユーザに負担をかけずにユーザの発声を入力したい音声入力装置に入力する音声入力システム、音声入力方法及び音声入力プログラムを提供することを目的とする。

【0004】

【課題を解決するための手段】本発明の音声入力システムは複数の音声入力装置がネットワークに接続され、前記ネットワークに接続された各々の音声入力装置において入力される音声情報をそれぞれ検知し、前記ネットワークに接続された各々の音声入力装置で前記検知した前記音声情報を必要に応じて処理したものである音声に関する情報を判断情報として前記ネットワークを介して他の音声入力装置に授受し、前記ネットワークに接続された各々の音声入力装置が自音声入力装置と他の音声入力装置における前記判断情報をもとに前記音声情報に対する処理の決定及び実行の判断を行うことを特徴とするものである。

【0005】また、本発明の音声入力方法はネットワークに接続された複数の音声入力装置において入力される音声情報をそれぞれ検知するステップと、前記ネットワークに接続された各々の音声入力装置で前記検知した前記音声に関する情報を判断情報として前記ネットワークを介して他の音声入力装置に授受するステップと、前記ネットワークに接続された各々の音声入力装置が自音声入力装置と他の音声入力装置における前記判断情報をもとに前記音声情報に対する処理の決定及び実行の判断を行うステップとを含むことを特徴とするものである。

【0006】また、本発明の音声入力プログラムはネットワークに接続された複数の音声入力装置において入力される音声情報をそれぞれ検知し、前記ネットワークに接続された各々の音声入力装置で前記検知した前記音声に関する情報を判断情報として前記ネットワークを介して他の音声入力装置に授受し、前記ネットワークに接続された各々の音声入力装置が自音声入力装置と他の音声入力装置における前記判断情報をもとに前記音声情報に

対する処理の決定及び実行の判断を行う機能を実現することを特徴とするものである。

【0007】

【発明の実施の形態】以下、図面を参照しながら本発明による音声入力システムについて説明する。はじめに、本発明の全体の概要を図1を用いて説明する。

【0008】本発明の音声入力システムでは、ネットワーク104に複数の単体の音声入力装置101や音声入力装置102を有した機器103、例えばビデオテープレコーダが接続され、これらの単体の音声入力装置101や機器103に搭載された音声入力装置102によりユーザの発声する音声命令や伝言のメッセージあるいは会話等を計測し、入力された音声信号を信号処理手段によって適当な信号に変換する。そしてこの変換された信号から音声入力システムは入力された音声に対する処理を単体の音声入力装置101や機器103に搭載された音声入力装置102で行うことができる。

【0009】また単体の音声入力装置101や機器103に搭載された音声入力装置102はネットワーク104を介して情報の授受が可能となっており、入力された音声に対する処理として、ネットワーク上の他の単体の音声入力装置や機器に搭載された音声入力装置と情報の送受信ができる。

【0010】このとき、ネットワークへの情報の送信については情報が一つ一つの各音声入力装置に移っていくようなリレー方式でも、一つの音声入力装置から同時に複数の音声入力装置に送るようなブロードキャスト方式でも構わないが、音声という実時間処理が重要な用途であるので以降はブロードキャスト方式を念頭に置いて説明する。

【0011】ユーザの発声がネットワーク接続された複数の音声入力装置に入力された場合に、各音声入力装置での処理をどうするかという点が本発明によって解決する所である。また、ユーザの発声が単一の音声入力装置にしか入力されなかった場合でも本発明の処理で包含することができる。

【0012】また本発明の実施例としては、ユーザの発声という人間の発声を主に例に挙げて説明しているが、本発明は人間の音声に限定されたものではない。目的に応じて機械の動作音や動物の声でも、音声であれば構わないものとする。

【0013】次に本発明の実施の形態の音声入力システムを構成する音声入力装置について図2を用いて説明する。音声入力装置(20-1~20-3)はそれぞれネットワーク21に接続されており、音声入力装置20-1はビデオテープレコーダ(以下ビデオとする)26に搭載され、音声入力装置20-3はエアーコンディショナー(以下エアコンとする)27に搭載され、また、音声入力装置20-2は単体で接続されている。音声入力装置20-1に入力された音声によりビデオ26の操作を行い、音声入力装置20-3に入力さ

れた音声によりエアコン27の操作を行う。なお、後述するように自分の音声入力装置への音声でなくとも、各機器は音声入力への処理を行うことができる。

【0014】各音声入力装置(20-1~20-3)はそれぞれマイクロホン201、信号処理部202、中央処理部203、記憶部204、ネットワーク接続部205、情報表出部206から構成される。

【0015】ユーザが発声する音声入力はマイクロホン201に入力され、このマイクロホン201でユーザの発声を計測する。これは一般にあるマイクロホンで実用可能である。このマイクロホンは、単一のマイクロホンや複数のマイクロホン(マイクロホンアレイ)、指向性・無指向性マイクロホンなど、マイクロホンとして使えるものから構成できるものとする。

【0016】マイクロホンから取り込まれた音声信号は信号処理部202で後段の処理に必要な形式に処理される。この処理は例えば音声信号のMPEGによる圧縮や、音声認識で用いられるケプストラム特徴に変換する処理などが考えられる。なお、この信号処理部202はその他にも音声入力装置の用途に応じて適当な処理を実行できるように構成できるものとする。

【0017】また、この信号処理部202では次に説明する中央処理部203からの命令を受けて情報表出部206に伝える形式に変換する機能も含まれている。さらに、この情報表出部206では中央処理部203からのユーザに伝えるメッセージ内容から音声合成を行って合成音の信号に変換している。

【0018】なお、その他にも、ディスプレイ表示のための表示内容に変換したりする処理や情報表出部206におけるデバイスや、音声入力装置の用途に応じて処理を実行できるように構成することも可能である。

【0019】ただし、このマイクロホンからの音声信号の処理と情報表出部206へ送る情報についての処理は同一処理機構で行うか否かは問わないものとする。すなわち上記の処理を行う機構を総称して信号処理部202とする。

【0020】また、信号処理部202の入力としてマイクロホン以外のセンサ・デバイスも考えられる。例えばカメラからの動画像や触覚センサ、スイッチ等が挙げられる。その他のセンサ・デバイスからの入力も、音声入力装置の用途に応じて処理できるような信号処理部を構成できるものとする。これについては後述する。

【0021】中央処理部203では音声入力装置全体の処理を制御する。この中央処理部203が音声入力装置の状態を管理し、必要に応じて各処理機構に命令を送る。信号処理部202からの情報やネットワーク接続部205からの情報、そして記憶部204の情報を元に制御内容を決めることができる。また、他の音声入力装置に対して制御情報を送出する。本発明の音声入力システムとして音声はどう処理するかについては後述する。

【0022】記憶部204では中央処理部203で行う処理のプログラムやその作業領域、信号処理部202からの情報やネットワーク接続部205からの情報を保持しておく機構である。なお、この記憶部204は信号処理部202における情報記憶用やネットワーク接続部からの情報記憶用といったように回路的には別のものであっても構わないとする。

【0023】すなわち、音声入力装置における情報保持機構を総称して記憶部204と呼ぶことにする。この記憶部204は半導体メモリや磁気ディスクなどの機構で実現可能であり、データを保持できる任意の機構で構成可能なものであるが、この実施の形態では半導体メモリが使用されている。

【0024】記憶部204の使われ方や記憶される情報については中央処理部203の処理の説明と共に後述する。

【0025】ネットワーク接続部205はネットワーク21を通して音声入力装置間の情報の授受を行うための機構であり、LANでのネットワーク接続やブルートゥースといった無線技術といった機器間通信技術によって実現できるものとし、ここではLANでのネットワーク接続を用いている。

【0026】また、以上のような音声入力装置の機構のそれぞれ、もしくは全てが、他の機能を持つシステムのものや機構を共有しても構わないとする。例えばビデオ・システムのようなオーディオ・ビジュアル機器に音声入力装置が含まれている場合に、共通の信号処理回路を使ってお互いの機能を実現したり、同じ中央処理回路を用いて音声入力装置やビデオ・システムの機能の制御を行ったりすることが考えられる。

【0027】他にも共通の機構で音声入力装置と他のシステムの機能を実現する例が考えられるが詳細は省略する。

【0028】さらに、回路的な機構として音声入力装置やその他のシステムが別々にあるのではなく、共通の回路でありながら、プログラムのなプロセスとして別のシステムとして制御できる場合も上記に含まれているものとする。

【0029】次に中央処理部203が信号処理部202からの音声信号やネットワーク接続部205からの情報、記憶部204で保持されている情報をもとにして音声をもとにして処理するかについて図3を用いて説明する。図3では図2のビデオ26に搭載された音声入力装置20-1（以下音声入力装置Aとする）とエアコン27に搭載された音声入力装置20-3（以下音声入力装置Bとする）に対して音声が入力される例を示している。さらに、現在ユーザが音声入力装置Bに対して対話処理を行い、音声入力装置Aは待機中の状態を示している。

【0030】まず、ユーザが音声入力装置A及び音声入力装置Bに対して発声すると（step301）、各音声入力装置の信号処理部202ではマイクロホン201で取り込まれ

たユーザからの発声を検知し、信号処理される（step302）。

【0031】ここで、音声入力装置Bは既にユーザと対話処理を行っているので、音声入力装置B自身が対話処理中であって他のシステムの状態が対話状態でないとなれば、音声入力装置Bがユーザの発声した内容に対する処理を行う選択をする。（step303）

次に、音声入力装置Bの中央処理部202は音声入力装置の機能にあわせて取り込まれた音声の処理を行い、音声の内容にしたがって機器を操作し、対話終了後再び待機状態になる（step304）。

【0032】逆に音声入力装置Aでは、音声入力装置Bがユーザとの対話状態であるので、信号処理された後（step302）、それ以上の処理を行わない（step305）とし、待機状態になる。

【0033】こうすることで、ユーザの発声が複数の音声入力装置で検知されてしまうような場合でも、ユーザが現在発声対象としている音声入力装置に対してのみ、楽にアクセスできるようにすることを可能とする。また上記ではユーザが複数の音声入力装置に対して発声する例を挙げたが、ユーザは意図的に複数の音声入力装置に検知されるように音声を発声する必要はなく、このことは以降の実施例でも同様である。

【0034】また、他の音声入力装置が対話状態でなければ処理を行うといった条件付けは、上記以外の条件についてユーザが任意に、もしくは音声入力装置が設定として定めることができるものとする。

【0035】また、ここでの対話は人間とシステムの一対一による音声のやり取りに限定したものではなく、人間からシステムへの一方的な音声発声やシステム側から視覚的な応答を返す場合、あるいはシステムから任意の人間に応答する場合を含んでも構わないものとし、以降の説明で用いられる対話についても同様である。

【0036】また、音声入力装置にはあるルールに基づいた順序関係があり、その順序関係に基づいて取り込まれた音声情報に対する処理を決めることもできる。ルールの具体例としては、音声入力装置の処理能力・ユーザによる設定・使用頻度・音声入力装置の機能に基づく設定値・マイクロホン以外からのセンサの情報や、これらの組み合わせ等が挙げられる。

【0037】次に上記の音声入力装置の機能による順位付けの例を図4を用いて説明する。

【0038】音声入力装置が搭載している機器としてウェアラブル・コンピューター（以下音声入力装置Cとする）と音声入力装置が搭載している機器としてビデオ・システム（以下音声入力装置Dとする）があり、前者の方が特定ユーザ向けなので順位が高く、ビデオ・システムは不特定のユーザが使い得るので順位が低いものとする。

【0039】このときユーザは音声入力装置C及び音声

入力装置Dに対して発声し(step401)、それぞれの音声入力装置は信号処理部202においてマイクロホン201で取り込まれたユーザからの発声を検知した場合に、自音声入力装置の順位を送信しあう(step402)。

【0040】次に、他の音声入力装置の順位と比較し、順位の高い音声入力装置Cがそのユーザの発声を処理する(step403)。

【0041】順位の低い音声入力装置Dは処理は行わず(step404)、待機中のまになる。

【0042】上記の例では順位情報を送信しているが、送信情報に順位以外の情報があっても構わないし、発声を検知してからでなく前もって情報のやり取りをしておく、あるいはプリセットの順位情報をもとに自音声入力装置で処理するかの判断を行っても構わないとする。

【0043】上記のような実施例によって、例えば音声入力装置を搭載する機器として火災報知器や緊急警報器のような非常用機器は他のどんな機器よりも順位が高く、例えば「助けて」という発声に対していかに通常機器で音声命令として登録していてもまずは非常用機器に対する音声入力優先されるということも可能となる。

【0044】また、音声入力装置内に時間を処理する機構を設けて、それによって処理の判断の参考にもできる。図5で例を挙げて説明する。

【0045】図5ではビデオに搭載された音声入力装置(以下音声入力装置Eとする)とエアコンに搭載された音声入力装置(以下音声入力装置Fとする)に対して音声が入力される例を示しており、音声入力装置Eは音声入力装置Fよりユーザに近い位置に設置している。

【0046】このときユーザは音声入力装置E及び音声入力装置Fに対して発声し(step501)、それぞれの音声入力装置は信号処理部202においてマイクロホン201で取り込まれたユーザからの発声を検知した場合に、自音声入力装置の発声検知時間を送信しあう(step502)。

【0047】次に、音声を検知した他の音声入力装置からの検知時間と自音声入力装置の検知時間を比較し、自音声入力装置が最も早かった場合は音声を処理し(step503)、そうでなければ当該音声を処理しないという判断をする(step504)ことで、ユーザが指定しなくともユーザに最も近い音声入力装置が音声の処理を行えるようになる。

【0048】また、音声検知時間がもっとも長かった音声入力装置がユーザの発声を最初から最後まで検知できたとみなして、その音声入力装置が当該音声の処理を行うといったように音声検出の早さ以外の時間情報を判断基準とすることもできる。

【0049】また、ユーザの発声の音量をマイクロホンから取り込まれた音声から計測し、処理の判断の参考にもできる。音量情報を利用した本発明の例として図6を用いて説明する。

【0050】ここでは上述した音声入力装置Eと音声入

力装置Fがある場合に、ユーザは音声入力装置E及び音声入力装置Fに対して発声し(step601)、それぞれの音声入力装置は信号処理部202においてマイクロホン201で取り込まれたユーザからの発声を検知した場合に、音量情報を送信しあう(step602)。すなわち、ユーザの発声の音量をマイクロホンから取り込まれた音声から計測し、ネットワーク上の他の音声入力装置に伝える。

【0051】次に、音声を検知した他の音声入力装置からの音量情報と自音声入力装置の音量情報を比較し、自音声入力装置が最も大きかった場合は音声を処理し(step603)、そうでなければ当該音声を処理しないという判断をする(step604)ことで、ユーザが指定しなくともユーザに最も近い音声入力装置が音声の処理を行う、もしくは元の発声を最もよく収録した音声で処理を行えるようになる。この音量情報としては音圧レベルや音響パワーレベル、あるいはphonやsoneなどの単位が挙げられる。

【0052】また、周囲の雑音に対するユーザの発声の信号対雑音比をマイクロホンから取り込まれた音声から計算して、処理の判断の参考にもすることもできる。信号対雑音比を利用した本発明の例として図7を用いて説明する。

【0053】図7ではビデオに搭載された音声入力装置(以下音声入力装置Gとする)とエアコンに搭載された音声入力装置(以下音声入力装置Hとする)に対して音声が入力される例を示しており、騒音源があり、音声入力装置Gは音声入力装置Hより騒音源が遠い位置にあるものとする。

【0054】始めに、各音声入力装置は常時音声を取り込んで周囲の雑音の情報を計測しておく(step701)。

【0055】次に、ユーザは音声入力装置G及び音声入力装置Hに対して発声し(step702)、それぞれの音声入力装置は信号処理部202においてマイクロホン201で取り込まれたユーザからの発声を検知し、ユーザの発声をマイクロホンから取り込んだときに雑音情報をもとに信号対雑音比を計算し、ネットワーク上の他の音声入力装置に伝える(step703)。

【0056】次に、音声を検知した他の音声入力装置からの信号対雑音比情報と自音声入力装置の信号対雑音比情報を比較し、自音声入力装置が最も大きかった場合は音声を処理し(step704)、そうでなければ当該音声を処理しないという判断をする(step705)。

【0057】これにより、ユーザが指定しなくともユーザに最も近い音声入力装置が音声の処理を行う、もしくは元の発声を最もよく収録した音声で処理を行えるようになる。ここでの例では、無発声中でも常時周囲音を取り込んで雑音を計算する例を挙げたが、他にも例えば発声を検知してから発声中の無音区間をもとに雑音を推定してもよい。

【0058】また、記憶部に使用状況に関する過去の履

歴を保持しておき、それを処理の判断に利用することもできる。過去の履歴を利用した本発明の例について図8を用いて説明する。

【0059】図8ではビデオに搭載された音声入力装置（以下音声入力装置Iとする）とエアコンに搭載された音声入力装置（以下音声入力装置Jとする）に対して音声が入力される例を示しており、音声入力装置Iは音声入力装置Jより使用頻度が高いものとする。

【0060】始めに、ユーザが両方の音声入力装置に対して発声（step801）し、この発声に対して最近の使用時間・使用回数等をネットワーク経由で他の音声入力装置に伝える（step802）。

【0061】一方、音声入力装置Iでは音声入力装置Jの使用履歴と比較して、音声入力装置Iが最もよく使われているなら音声の処理を行うよう判断する（step803）ことでユーザがわざわざ指定しなくてもよく使われている音声入力装置Iを利用できるようになる。

【0062】また、他方、音声入力装置Jでは音声入力装置Iの使用履歴と比較して、音声入力装置Jがあまり使われていないなら音声の処理は行わず（step804）、待機中のままになる。

【0063】また、音声認識をする手段を備えその認識結果を利用して取り込まれた音声の処理を判断することもできる。信号処理部からの情報は音声認識を行う機構で処理されその結果が中央処理部に渡される。このとき行われる音声認識は、演算処理を中央処理部で扱っても構わない。

【0064】また音声認識に使われる手法は混合正規分布をモデルに使ったHMMやDPマッチングのような一般に現実化されている手法で構わないとし、このとき使われるHMMや言語モデルは記憶部にあっても構わないとする。音声認識の語彙は音声入力装置毎に異なっているとしても共通化されていても構わないとする。さらにその語彙に制御命令を対応させることで音声コマンドを可能にすることもできる。この音声認識を利用した本発明の例について図9で説明する。

【0065】図9ではビデオに搭載された音声入力装置（以下音声入力装置Kとする）とエアコンに搭載された音声入力装置（以下音声入力装置Lとする）に対して音声が入力される例を示している。

【0066】始めに、各音声入力装置に対してユーザからの音声入力装置Kに関連する「再生」という発声があった（step901）場合に、各音声入力装置はその音声の検知と音声認識を行う（step902）。

【0067】その音声認識した結果を中央処理部は受け取り、認識結果から自音声入力装置に対する発声か否かを判断し（step903）、その判断結果と認識結果をネットワーク経由で他の音声入力装置に伝える（step904）。

【0068】一方、他の音声入力装置の判断結果と認識

結果をみて、音声入力装置Kでは自音声入力装置への発声と判断（step905）できたら当該音声に対する処理を行うことで、ユーザが特に指定しなくても発声対象の音声入力装置を使うことができるようになる。

【0069】他方、音声入力装置Lでは自音声入力装置への発声と判断しない（step906）ので、待機中のままである。

【0070】また、音源の識別を行う手段を備え、その識別結果を利用して音声の処理を判断することもできる。音源の種類としては人間、機械、動物など使用目的に応じて考えられるが、以降では例として人間の発声を音声とした場合について説明する。信号処理部からのユーザの音声情報に対して話者識別を行い、その結果を中央処理部に伝える。この話者識別を行う方法は話者毎に学習または適応されたHMMに対する尤度から判断するものや、性別や年齢層毎のモデルで最も近いカテゴリーを選ぶものなど、個人あるいは話者の特性（例えば性別や年齢層など）を識別できる手法ならば構わないものとする。

【0071】この話者識別を使った本発明の例を次の図10を用いて説明する。

【0072】図10ではビデオに搭載された音声入力装置（以下音声入力装置Mとする）とエアコンに搭載された音声入力装置（以下音声入力装置Nとする）に対して音声が入力され、あるユーザは片方の音声入力装置Mでのみ音声の処理が可能である場合の例を示している。

【0073】始めに、各音声入力装置に対してユーザからの発声があった（step1001）場合に、ユーザの発声を検知した音声入力装置は話者識別を行い（step1002）、自音声入力装置で処理すべき発声か否かを判断（step1003）をして、その判断結果と話者識別結果をネットワーク経由で他の音声入力装置に伝える（step1004）。

【0074】そして自音声入力装置と他の音声入力装置における判断結果と話者認識結果をみて、自音声入力装置への発声と判断（step1005）できたら当該音声に対する処理を行い、逆に他方の音声入力装置Nは自音声入力装置への発声ではないと判断（step1006）できたら処理を行わないとすることで、ある音声入力装置が特定のユーザに利用可能である場合に、ユーザが特に指定しなくても発声対象の音声入力装置を使うことができるようになる。

【0075】また、話者識別の信頼性が低い場合や複数話者が候補となった場合に、システム側からさらに暗証番号や定型句あるいは自由発声を促してさらにデータを得ることによって識別精度を上げてから話者識別以降の処理をおこなってもよい。

【0076】また、ここでは人物の話者認識について述べているが、前記のように故障者や動物の音に応じて識別とその後の処理を行うことも可能である。

【0077】また、音声入力装置やネットワーク上の他

の機器と共通の命令を持ち、お互いに許された範囲で制御することもできる。こうすることで、他の音声入力装置の働きを抑制したり、音声入力装置同士の互換性をよくしたりすることができる。

【0078】この例を図11で説明する。

【0079】例えばネットワーク1102に接続されている全ての音声入力装置1101が「電源ON」「電源OFF」「省電力」といった共通の電源管理命令を持っているときに、ネットワーク1102に繋がっているパーソナルコンピュータ1103から一度に複数システムも含めた任意の音声入力装置1101の電源を操作する命令をネットワーク経由で送信し、各音声入力装置がその命令を実行することが出来る。

【0080】また、音声入力装置やネットワーク上の他の機器と共通の音声による制御命令と、入力された音声とその命令をマッチングする手段を備えることで、より平易で確実な音声による制御命令の実行を可能とする。この例について図12のフロー図を用いて説明する。

【0081】図12の例では音声入力装置を有しているビデオ（音声入力装置O）と音声入力装置を有しているエアコン（音声入力装置P）があったときに、ユーザが「ビデオ」「エアコン」と命令対象の名称を発声した後で、「電源ON」「電源OFF」といったように共通の動作について共通化された命令を発声する。

【0082】ここで、ユーザから「ビデオ」「電源ON」という発声があった場合（step1201）、音声入力装置O及び音声入力装置Pは前述の音声認識で使われるマッチング手段で機器名称と機器命令を認識（step1202）し、自身のシステムへの命令か、処理可能かについて判断する（step1203）。

【0083】その結果をネットワーク上の他の音声入力装置や制御可能機器にその結果を伝達（step1204）し、その結果と他の音声入力装置からの結果から自音声入力装置が処理すべき発声か判断（step1205）してその制御命令に対応した処理を行うことができる。

【0084】共通化された命令に対して複数の音声入力装置から得られた結果を使うことが、これまでの音声によるリモコンや音声によって命令する機器とは異なる点である。

【0085】また、ネットワーク上に音声による制御可能機器が複数ある場合に、記憶部でその制御命令の全てまたは一部に関する情報を記憶できるような仕組みと、入力された音声とそれらの命令をマッチングさせる手段を備えることで、より平易で確実な音声による制御命令の実行が可能となる。

【0086】この例を次の図13、図14を用いて説明する。ネットワーク上に音声入力装置で制御可能なビデオ（音声入力装置Q）とエアコン（音声入力装置R）があるとして、音声命令について音声入力装置Qが「再生」「停止」、音声入力装置Rが「温度あげて」「温度さげ

て」等であった場合に、ネットワーク上のそれぞれの音声入力装置では認識単語と対象機器を関連付けて記憶できるようにしているとする。

【0087】図13はこの認識単語と対象機器、そして処理内容を結びつける概念を表している。この図13のような認識単語と処理内容との結びつきは、単純な表引きやオブジェクト指向や高次の知識処理によって実現できるものとし、ここではその詳細は省略する。

【0088】図14のフロー図に示すようにユーザが「ビデオ」「再生」と発声した場合（step1401）、音声入力装置Qと音声入力装置Rは発声の検知と認識を行う（step1402）。

【0089】さらに、図13に示した概念を用いて発声内容を判断し（step1403）、その結果をネットワーク上の他の音声入力装置に伝達し（step1404）、その結果と他の音声入力装置から送られてきた結果をもとに自音声入力装置が処理すべき発声だったかを判断して（step1405）、その制御命令に対応した処理を行う。

【0090】上述の「ビデオ」「再生」の場合、図13のような認識単語と対象機器、処理内容の結びつきによってどちらの音声入力装置も発声がビデオに対して再生の命令であったと判断できる。さらにネットワーク経由で送信した情報により、発声が一意に解釈でき、音声入力装置は認識結果に対応する処理内容を行うことが出来る。

【0091】またこれまでの音声認識を用いた例では基本的に単語認識による例を挙げてきたが、ワードスポッティングや連続音声認識の技術を使っても、各音声入力装置での音声認識のスペックに差があっても、図13のような認識結果と処理内容の対応づけの概念がされれば構わないとする。

【0092】また、上述の図14で示した例については、音声入力装置以外のネットワークに接続された制御対象機器についても処理できるものとする。その例について図15を用いて説明する。

【0093】図15に示すように音声入力装置のついたエアコン1501と単体の音声入力装置1502及びビデオ1503がネットワーク1504に接続されており、ここでユーザがビデオ1503を操作する発声をする。

【0094】この音声入力装置は図14のフロー図に示す流れで音声の検知及び認識を行い、図13のような概念で認識結果と処理内容を結びつける。そして認識結果と処理内容の判断をしてからネットワーク1504上の他のシステムに送信する。

【0095】その結果、ビデオ1503は認識結果に応じた処理内容を受け、発声を実行することができる。よってビデオ1503自体に音声入力装置がなくても自分が制御可能な情報についてネットワークに情報を流し、各音声入力装置に図13のような認識結果と処理内容の概念をつくることで音声による制御が可能となる。

【0096】図12から図15までで説明した音声認識を用いた本発明の例については、これまでブロードキャスト方式で音声認識と判断の結果をネットワークの全てのシステムに送信する例を挙げてきたが、認識結果によって直接その対象機器にのみ認識結果と判断の結果を伝えてもよいものとする。

【0097】また、音声入力装置において、マイクロホンによる音声入力以外のセンサがある場合に、そのセンサ情報を利用して検知した音声の処理内容を判断することもできる。この例について、図16を用いて説明する。

【0098】図16に示すように音声入力装置を有したエアコン1601と単体の音声入力装置1602がネットワーク1603に接続されている。また、この単体の音声入力装置にはカメラを有しておりカメラから周辺の画像情報を取り入れることができる。なお、このカメラの入力は図2の信号処理部202に輸入され画像処理される。

【0099】この音声入力システムにおいて、ユーザがエアコン1601の音声入力装置に対して発声する。ここで、単体の音声入力装置1602に付いているカメラにより話者がどの方向を向いているかを推定する。なお、この話者がどの音声入力装置を向いているかについては、画像から人間を抽出する技術、顔部分を推定してその向きを推定する技術、口の動きから検知した発声かどの人間からのものか推定する技術等の組み合わせで実現できるものとするが、ここでは詳細は省略する。

【0100】推定された話者の顔向きから話者がエアコン1601の方を向いていると判断すると、発声の対象機器をエアコンと判断して、各音声入力装置は結果をネットワーク1603で他の音声入力装置に通知し、これまで述べてきた例のように処理を判断する。

【0101】ここではカメラを使った画像情報を利用した例を挙げたが、スイッチ等の直接的なセンサ・デバイスや音源定位のためのマイクロホンアレイなどが考えられるが、どのような計測技術を使うかは限定しない。

【0102】また、図2の音声入力装置の構成で述べたようにマイクロホン201、情報表出部206、信号処理部202、中央処理部203、記憶部204、ネットワーク接続部205は音声入力装置においてそれぞれその働きをするものの総称であるので、ネットワークを通した形や直接接続された形でそれぞれが物理的に複数に分かれていても構わないとする。この例を図17で説明する。

【0103】図17に示すように音声入力装置は物理的には2つの音声入力装置（1701、1702）に分かれていてもネットワーク1703で接続されており適切な情報のやり取りが出来るものとする。このときユーザの発声に対して、2つの音声入力装置（1701、1702）で一つの音声入力装置として働くことが出来る。

【0104】また、これまで述べたような音声入力装置に対する判断の基準は、他の音声入力装置の情報やユーザの設定によって変えられるものとする。例えば、音声

入力装置は音声を検知したときの検知や認識結果等の情報以外に、一定時間ごとに他の音声処理システムの処理状態、処理性能、認識可能語彙やそれに対する処理内容をやり取りして、自音声入力装置の記憶部に蓄えておけるとする。

【0105】そのような情報を利用して、現在はある音声入力装置は処理出来ないから自音声入力装置で処理可能な場合は代わって処理するとか、自音声入力装置より性能のいい音声入力装置の認識結果を自分の結果より重視することで認識誤りを補正するとか、ユーザが自分の好みに合わせて上述のような判断の制御を可能とすることが出来る。

【0106】また、これまで述べたような音声入力装置に対する入力判断の手段は、上述のものを組み合わせても構わないとする。例えば、検知時間が早い音声入力装置が発声を扱うとするが、ある許容時間内では時間差がないものとし、同じ時間の場合は音量で判断するか、音声認識の尤度と音声入力装置の順位を重み付けして最もスコアの高い音声入力装置で音声を扱うなどが考えられる。

【0107】また、上述のような判断の手段の組み合わせにより得られた情報を利用して高次のエージェントシステムや知識処理システムで判断する場合も考えられる。

【0108】また、これまで述べたような音声入力装置における処理の判断の手段は、ネットワーク上の音声入力装置間で同一であることを必須とはしないものとする。例えば音声入力装置が2つあり、一つは音声の検知時間のみで、他方は音量情報のみで判断する場合には、音声を検知したあとに相互に授受する情報は必ずしも対応は取れないが、各々の音声入力装置でその場合における処理を装置の目的に応じて設定しておけば、音声入力システムとして処理が破綻せずに各々の音声入力装置で処理の判断が可能である。また、上述のような音声入力装置の判断の手段が各々の音声入力装置において異なっている場合に、ネットワークを通して授受した情報をもとに音声入力装置より高次のエージェントシステムや知識処理システムで処理を判断する場合も考えられる。

【0109】また、これまで述べたような音声入力装置に対する入力判断において、音声検知時間や音量といった発声に関する情報や、音声認識結果や識別結果といった情報から、ユーザがどの機器に対してどのような音声入力を行ったのかが一意には判断できなかった場合は、音声入力装置の一つがユーザと対話処理を行って決定したり、マイクロホン以外のセンサ情報といった他の条件を使って決定したりすることもできる。

【0110】次に、これまで述べたような音声入力装置において先に説明した図2の情報表出部206とこれまでの説明の補足となる例を次の図18を用いて説明する。

【0111】図18に示すように音声入力装置を有したエ

アコン1801、単体の音声入力装置1802及び音声入力装置を有したビデオ1803がネットワーク1804に接続されている。また、これらの音声入力装置は図2の情報表出部206を有している。

【0112】この音声入力システムでは前述したように、待機中に各音声入力装置は自音声入力装置の情報、すなわち認識語彙、処理内容やここでは特に情報表出部の有無と表現可能なメディアの情報をやり取りして記憶部に保存してあるものとする。

【0113】この例での各音声入力装置の情報表出部は全てスピーカを備え、中央処理部と信号処理部によって合成された任意文の音声ユーザに返すことができる。そしてその情報表出部への制御命令の一部は音声入力装置で共通化されているとする。つまりある音声入力装置が自分の情報表出部からユーザに応答を返す代わりに、ネットワーク上の別な音声入力装置の情報表出部からユーザへの応答を可能とする。

【0114】ここでユーザから「ビデオ」「再生」という発声があったときに、エアコンに接続された音声入力装置と単体の音声入力装置がその音声を検知したとする。なお、ユーザの位置は単体の音声入力装置に一番近いところにあるとする。

【0115】これまで述べてきたような手順により両音声入力装置は音声の検知、認識、自音声入力装置への命令か判断して、結局「ビデオ」への「再生命令」と判断し、それぞれネットワーク上の他の音声入力装置へ伝える。ビデオに接続された音声入力装置は直接音声を検知しないが、ネットワーク上の別な音声入力装置からの情報を受け、自音声入力装置への命令と解釈して、再生命令がされた場合の処理を実行する。

【0116】またこのとき、単体の音声入力装置の方がユーザに近いため、ネットワーク上に送られた音量や信号対雑音比の情報で判断したときに、ビデオに接続された音声入力装置よりも単体の音声入力装置の方が音声処理に適していることを各音声入力装置は判断できる。

【0117】よって、単体の音声入力装置とビデオの音声入力装置はそれぞれ単体の音声入力装置がユーザとの音声の授受を行う音声入力装置と判断できる。

【0118】再生命令を受けたビデオの音声入力装置は、ビデオに対して再生の制御命令を送る一方、再生を始めたことをユーザに伝えるために、「再生を開始しました」という合成音声を単体の音声入力装置からユーザに返すよう命令を生成して、ネットワークを介して単体の音声入力装置へ伝える。このときビデオの音声入力装置から送信される制御命令はこれまでのネットワークへの情報送信と同様に単体の音声入力装置一つへ直接送信してもよいし、単体の音声入力装置への命令という情報を含んだ形で、ブロードキャスト形式で全ての音声入力装置へ伝えられてもよい。

【0119】このようにしてビデオの音声入力装置から

送られたユーザへの応答命令を解釈して、単体の音声入力装置は合成音声で「再生を開始しました」というメッセージをユーザに伝えることができる。

【0120】また、この処理を通して単体の音声入力装置とビデオの音声入力装置は、ユーザと対話処理中であるというフラグを一定時間立てることで、ユーザの次の発声を優先的に処理し、エアコンの音声入力装置で処理しなくてもよいように出来るという例については既に述べてある。

【0121】次に、これまで述べてきたような音声入力装置において音声入力装置が何らかの基準でグループ化されている場合の例について図19を用いて説明する。

【0122】この例では音声入力装置の場所を基準としグループ「キッチン」1901、グループ「ウェアラブル」1902、グループ「リビング」1903のグループは全てネットワーク1904で接続されている。また、それぞれのグループ内に音声入力装置があり、これらの各グループ内におけるそれぞれの音声入力装置は他グループを同定できる情報を持っているものとする。

【0123】ただし、自グループの他の音声入力装置に関して記憶部が持つ情報と、他グループに関して持つ情報の種類は必ずしも同一でなくてよい。具体的にはここでは他グループにおける各々の音声入力装置の認識語彙やそれに対応する対象機器や処理内容の情報までは持たないとする。

【0124】ここでユーザが「リビング」ビデオ、「再生」と発声し、それがグループ「キッチン」とグループ「ウェアラブル」の音声入力装置で検知されたとする。これまで述べてきた例と同様に、検知した音声入力装置で認識と自音声入力装置で処理すべきか判断した結果、自グループへの発声でなくグループ「リビング」への発声と判断し、その音声情報や判断結果をグループ「リビング」の音声入力装置へ伝える。

【0125】このとき基本的に同定できたグループにのみ情報を送ることで、多くの音声入力装置がネットワークに接続されたときに必要な音声入力装置のみが情報のやり取りをできるようになることがグループ化することの利点である。

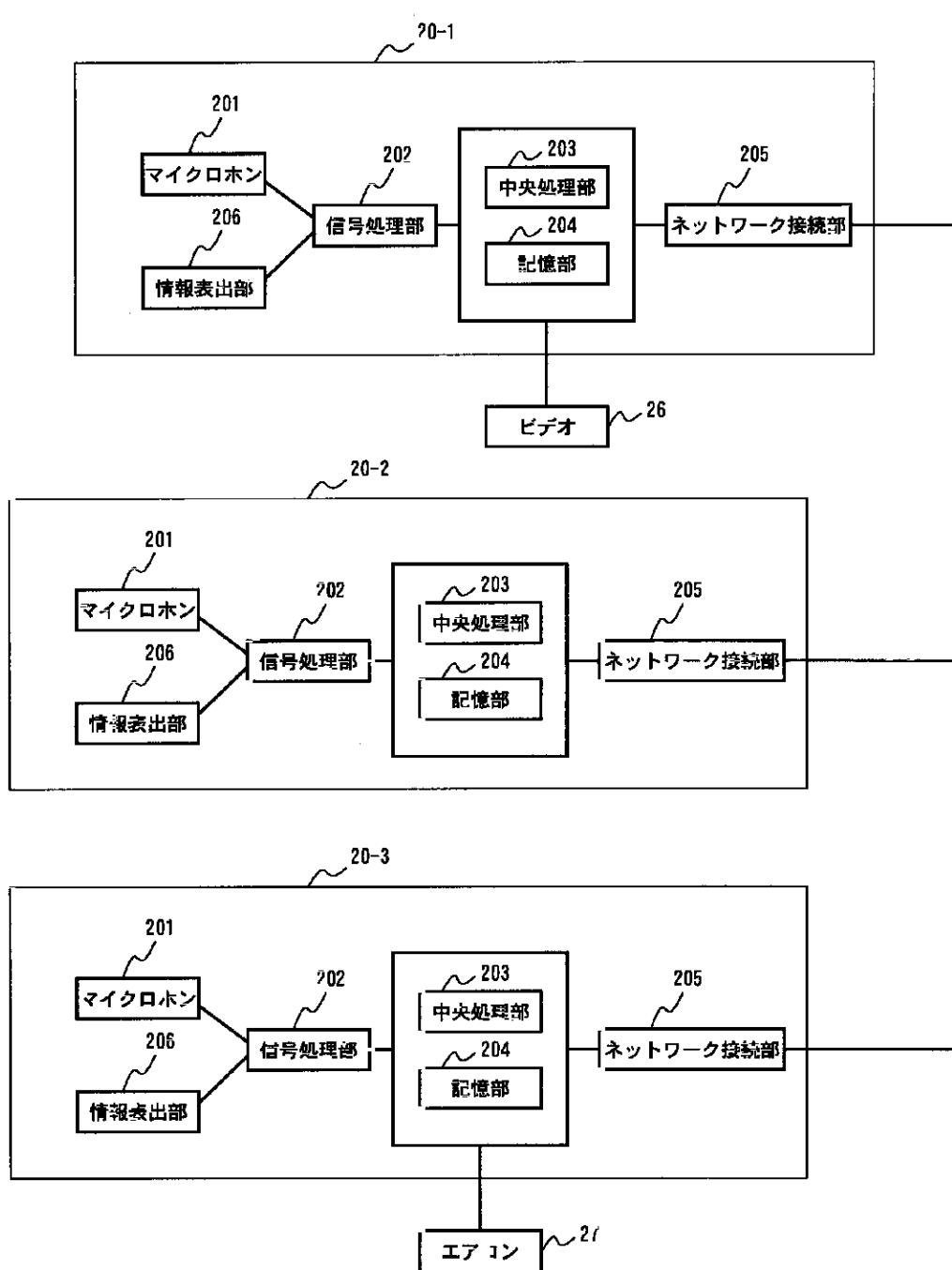
【0126】したがって、グループ「リビング」の音声入力装置は自グループ宛の音声に関する情報を受け取ることで自グループ内の「ビデオ」に対する「再生」の命令と判断してそれに対応する処理をすることができる。なお、本発明は音声入力プログラムに適用することも言うまでもない。

【0127】

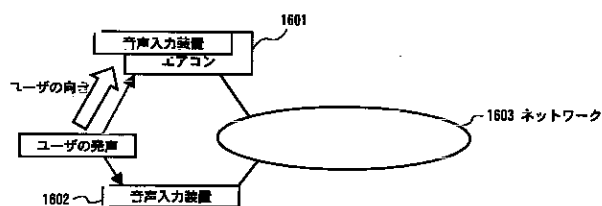
【発明の効果】以上説明したように、本発明はユーザの発声に対して他の音声入力装置からの情報を利用することで、ユーザに負担をかけずに音声に対する処理を決定することができる。

【図面の簡単な説明】

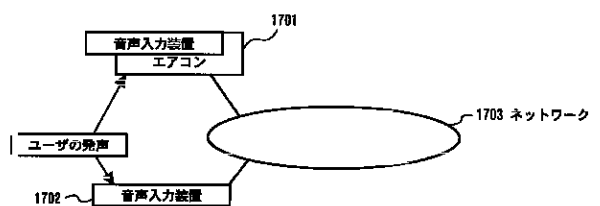
【図2】



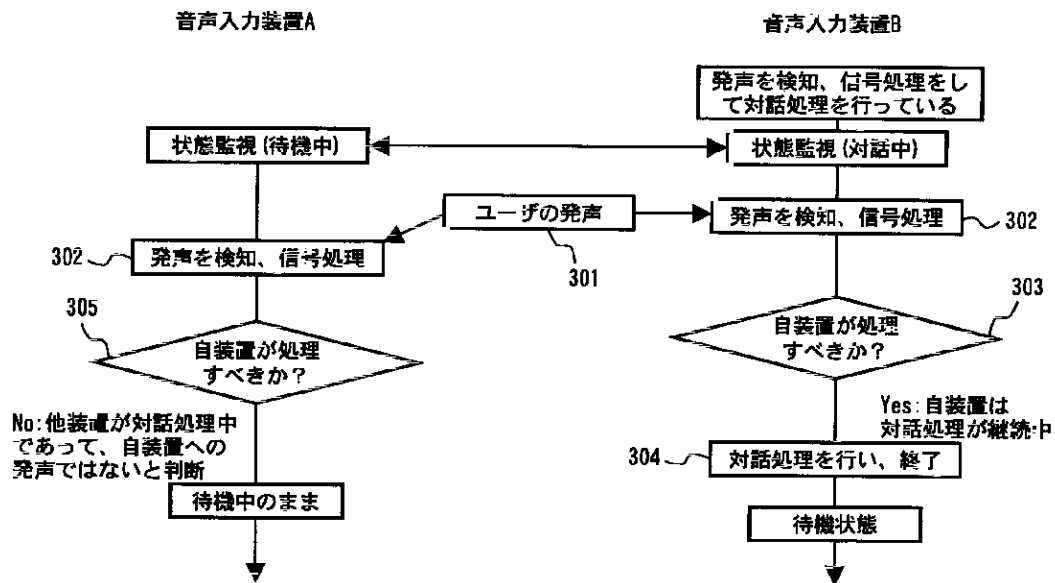
【図16】



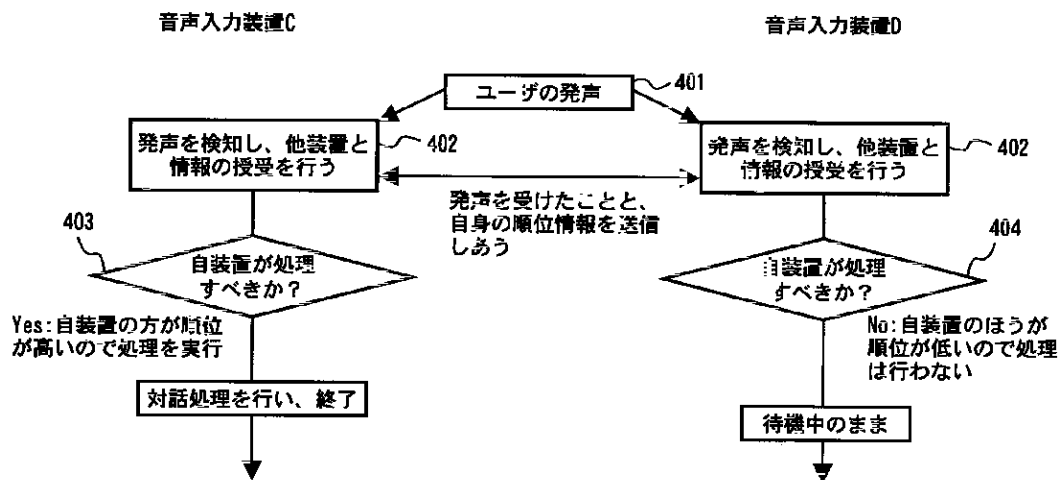
【図17】



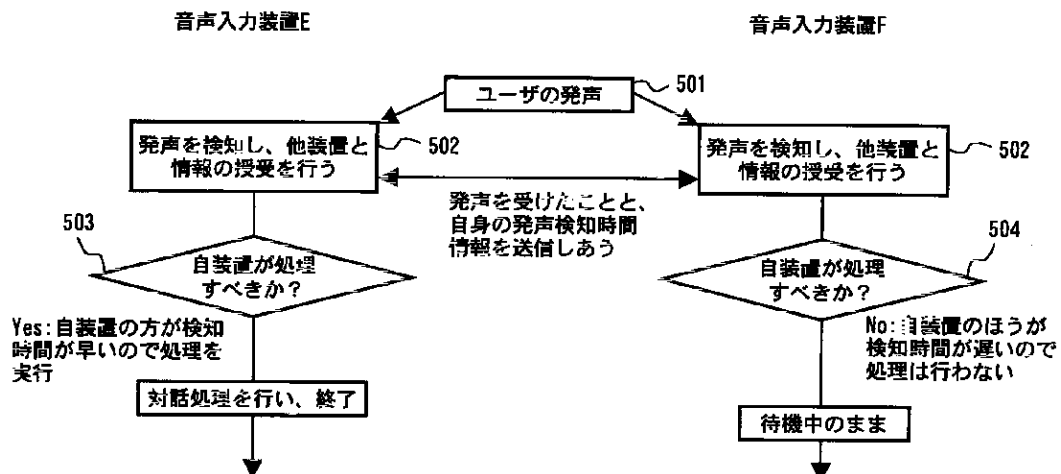
【図3】



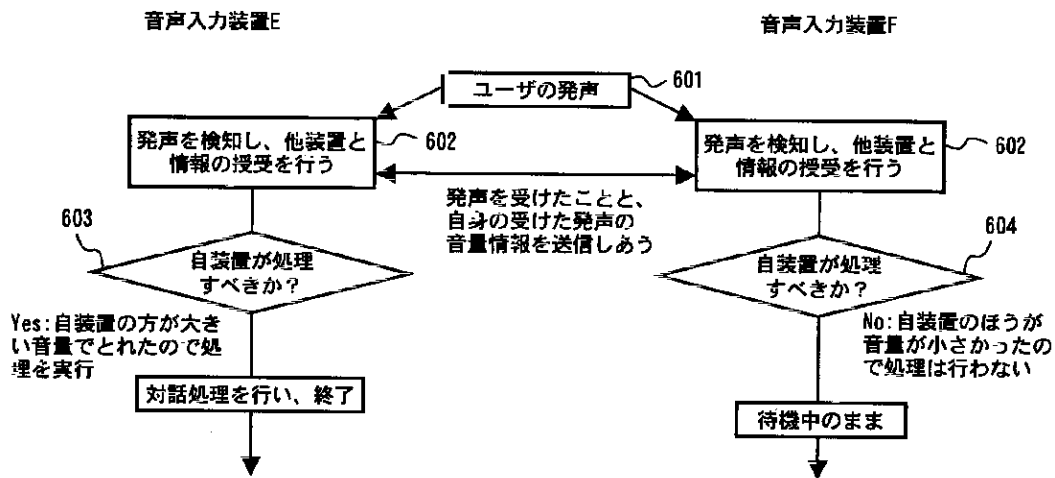
【図4】



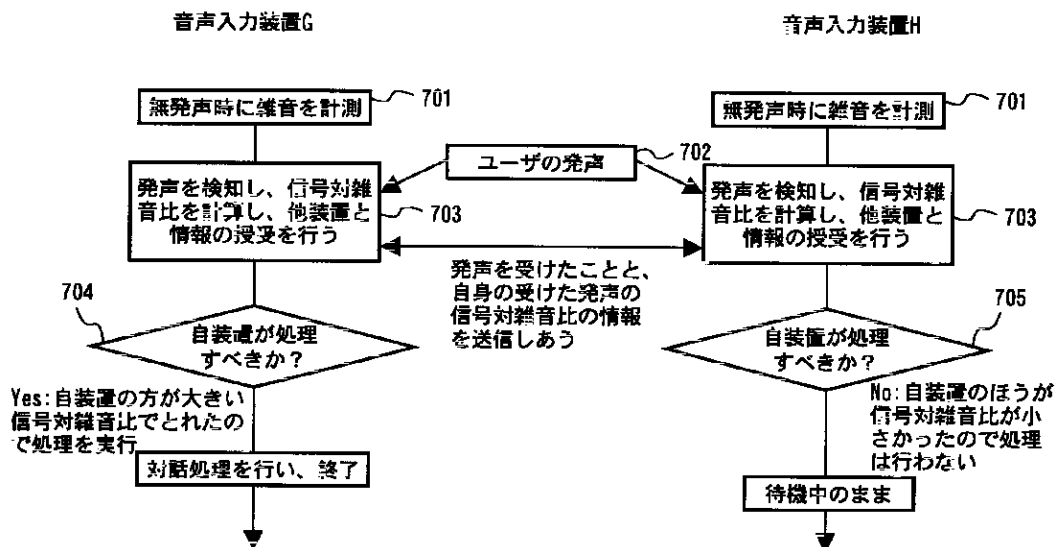
【図5】



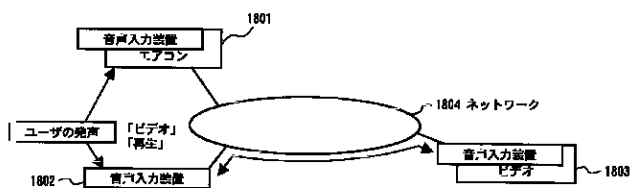
【図6】



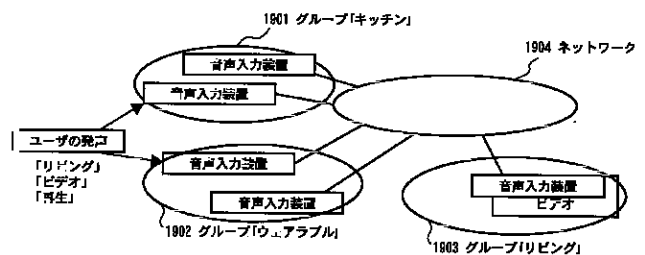
【図7】



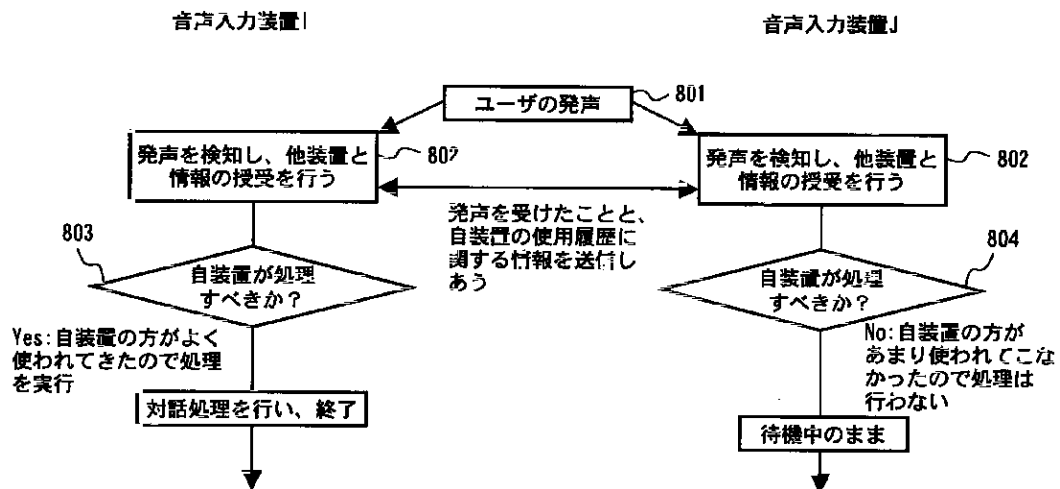
【図18】



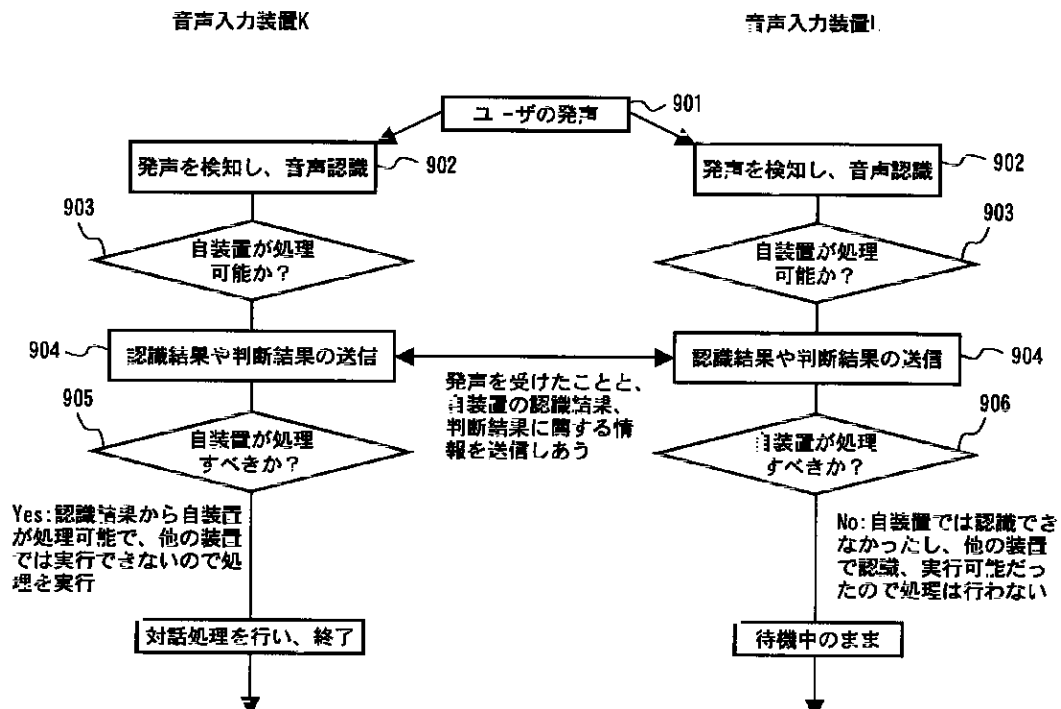
【図19】



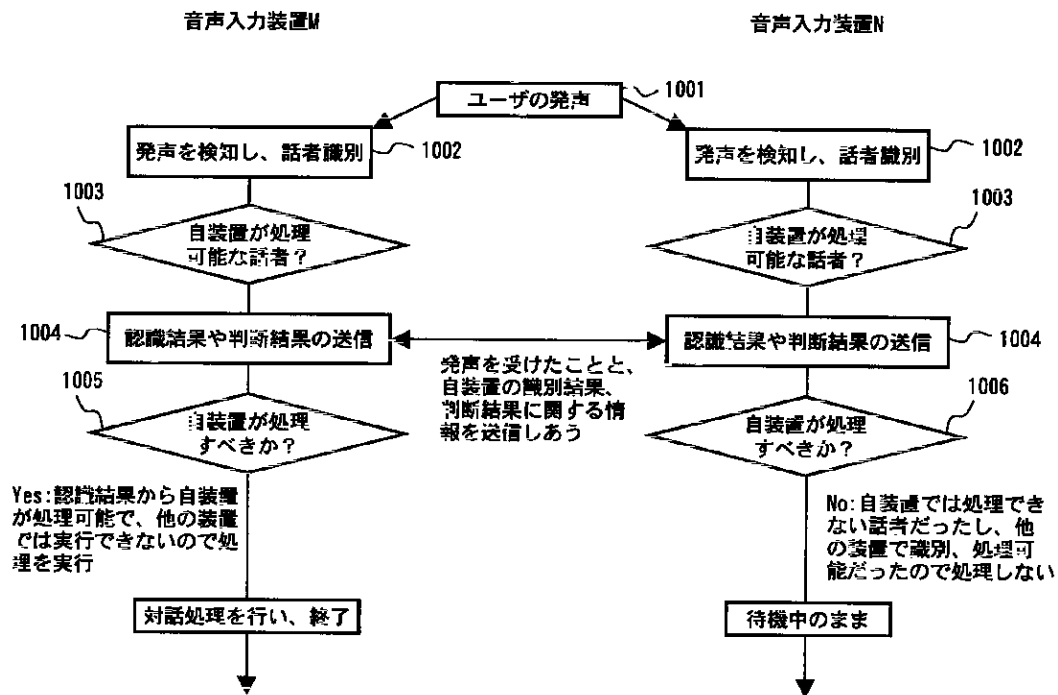
【図8】



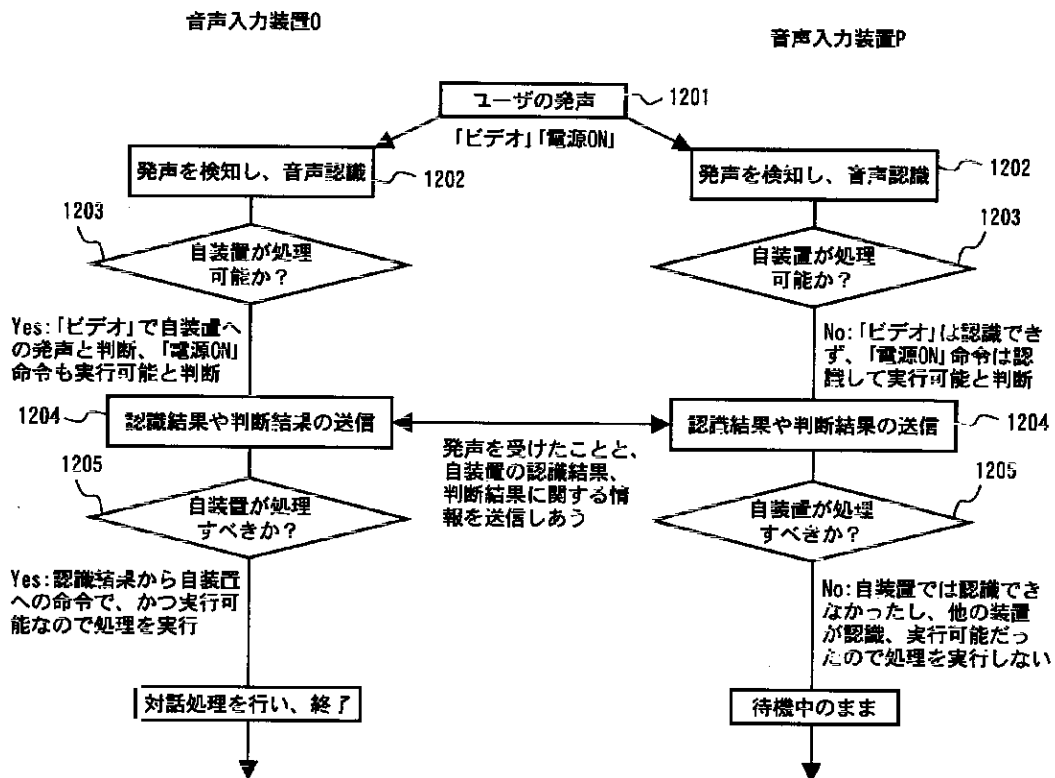
【図9】



【図10】



【図12】



【図14】

