(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2009-42910
(P2009-42910A)
(43)公開日　平成21年2月26日(2009.2.26)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 3/033	(2006.01)	G06F	3/033	310Y	3C007
G10L 17/00	(2006.01)	G10L	17/00	200C	5B087
B25J 13/08	(2006.01)	G10L	17/00	400	5D015
G06F 3/01	(2006.01)	B25J	13/08	Z	5E501
B25J 19/02	(2006.01)	G06F	3/01		

審査請求 未請求　請求項の数 25　ＯＬ　（全 50 頁）　最終頁に続く

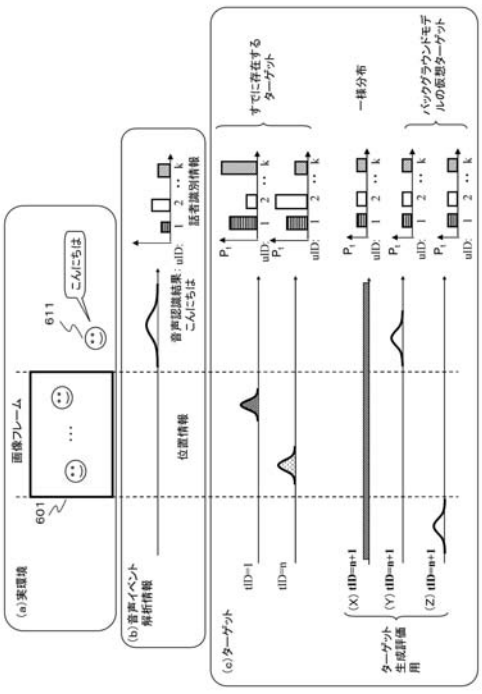
(21)出願番号　特願2007-205646 (P2007-205646)
(22)出願日　平成19年8月7日 (2007.8.7)

(71)出願人　000002185
ソニー株式会社
東京都港区港南１丁目７番１号
(74)代理人　100093241
弁理士　宮田　正昭
(74)代理人　100101801
弁理士　山田　英治
(74)代理人　100086531
弁理士　澤田　俊夫
(74)代理人　100095496
弁理士　佐々木　榮二
(72)発明者　澤田　務
東京都港区港南１丁目７番１号　ソニー株式会社内

最終頁に続く

(54)【発明の名称】情報処理装置、および情報処理方法、並びにコンピュータ・プログラム

(57)【要約】
【課題】不確実で非同期な入力情報に基づく情報解析により、精度の高いユーザ位置およびユーザ識別情報を効率的に生成する構成を実現する
【解決手段】カメラやマイクによって取得される画像情報や音声情報に基づいてユーザの推定位置および推定識別データを含むイベント情報を入力して、複数ターゲットを設定した複数パーティクルを適用したパーティクルフィルタリング処理を行い仮説の更新取捨選択によりユーザ位置および識別情報を生成する。また、カメラの画像フレーム外に仮想ターゲットを設定した暫定ターゲットとイベント検出部の生成するイベント情報との尤度を検証し、検証結果に応じて暫定ターゲットを各パーティクルに追加する。本構成により、フレーム外ユーザの音声入力に対応した処理が可能となり、ユーザ位置や識別の正確な推定処理が実現される。
【選択図】図１３



【特許請求の範囲】
【請求項１】
実空間における画像情報または音声情報のいずれかを含む情報を入力する複数の情報入力部と、
前記情報入力部から入力する情報の解析により、前記実空間に存在するユーザの推定位置情報および推定識別情報を含むイベント情報を生成するイベント検出部と、
ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、前記イベント情報に基づく仮説の更新および取捨選択により、前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報の生成を実行する情報統合処理部を有し、
前記情報統合処理部は、
前記イベント検出部の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットと前記イベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベント－ターゲット間尤度より大きい値である場合に、前記暫定ターゲットを各パーティクルに新規追加する処理を行うことを特徴とする情報処理装置。
【請求項２】
前記情報統合処理部は、
前記画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとして、画像フレームの異なる方向のフレーム外部位置に仮想ターゲットを設定した複数の異なる暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出して、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うことを特徴とする請求項１に記載の情報処理装置。
【請求項３】
前記情報統合処理部は、
前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットの他、均一データによって構成されるユーザ位置情報、ユーザ識別情報を持つ暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出し、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うことを特徴とする請求項１に記載の情報処理装置。
【請求項４】
前記イベント検出部は、
ガウス分布からなるユーザの推定位置情報と、ユーザ対応の確率値を示すユーザ確信度情報を含むイベント情報を生成する構成であり、
前記情報統合処理部は、
仮想的なユーザに対応するガウス分布からなるユーザ位置情報と、ユーザ対応の確率値を示すユーザの確信度情報を有するターゲットを複数設定したパーティクルを保持し、各パーティクルに設定されたターゲットと、前記イベント情報との類似度の指標値であるイベント－ターゲット間尤度を算出して、イベント－ターゲット間尤度の高いターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行する構成であることを特徴とする請求項１～３いずれかに記載の情報処理装置。
【請求項５】
前記情報統合処理部は、

前記イベントーターゲット間尤度と、各パーティクルに設定したパーティクル重みとの総和データをターゲット重みとして算出し、ターゲット重みの大きいターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行する構成であることを特徴とする請求項４に記載の情報処理装置。

【請求項６】

前記情報統合処理部は、

各パーティクルに設定したイベント発生源仮説ターゲットと、前記イベント検出部から入力するイベント情報との尤度を算出し、該尤度の大小に応じた値をパーティクル重みとして各パーティクルに設定する構成であることを特徴とする請求項４に記載の情報処理装置。

【請求項７】

前記情報統合処理部は、

前記パーティクル重みの大きいパーティクルを優先的に再選択するリサンプリング処理を実行して、パーティクルの更新処理を行う構成であることを特徴とする請求項６に記載の情報処理装置。

【請求項８】

前記情報統合処理部は、

各パーティクルに設定したターゲットについて、経過時間を考慮した更新処理を実行する構成であることを特徴とする請求項１～３いずれかに記載の情報処理装置。

【請求項９】

前記情報統合処理部は、

各パーティクルに設定したイベント発生源仮説ターゲットについて、前記イベント検出部の生成するイベント情報を適用した更新処理を行う構成であることを特徴とする請求項４に記載の情報処理装置。

【請求項１０】

前記情報統合処理部は、

前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和を、各ターゲット対応のユーザ位置情報およびユーザ識別情報としたターゲット情報を生成する構成であることを特徴とする請求項６に記載の情報処理装置。

【請求項１１】

前記情報統合処理部は、

前記パーティクルの各々に設定したイベント発生源仮説ターゲットの数に応じて、イベント発生源の確率値としてのシクナル情報の生成を行う構成であることを特徴とする請求項４に記載の情報処理装置。

【請求項１２】

前記情報統合処理部は、

前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和に含まれるユーザ位置情報としてのガウス分布データのピーク値が予め設定した閾値未満である場合に、該ターゲットを削除する処理を実行する構成であることを特徴とする請求項６に記載の情報処理装置。

【請求項１３】

情報処理装置において情報解析処理を実行する情報処理方法であり、

複数の情報入力部が、実空間における画像情報または音声情報のいずれかを含む情報を入力する情報入力ステップと、

イベント検出部が、前記情報入力ステップにおいて入力する情報の解析により、前記実空間に存在するユーザの推定位置情報および推定識別情報を含むイベント情報を生成するイベント検出ステップと、

情報統合処理部が、ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、前記イベント情報に基づく仮説の更新および取捨選択により、前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情

報の生成を実行する情報統合処理ステップを有し、
前記情報統合処理ステップは、
前記イベント検出部の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットと前記イベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベント－ターゲット間尤度より大きい値である場合に、前記暫定ターゲットを各パーティクルに新規追加する処理を行うステップであることを特徴とする情報処理方法。

【請求項１４】
前記情報統合処理ステップは、
前記画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとして、画像フレームの異なる方向のフレーム外部位置に仮想ターゲットを設定した複数の異なる暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出して、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うステップであることを特徴とする請求項１３に記載の情報処理方法。

【請求項１５】
前記情報統合処理ステップは、
前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットの他、均一データによって構成されるユーザ位置情報、ユーザ識別情報を持つ暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出し、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うステップであることを特徴とする請求項１３に記載の情報処理方法。

【請求項１６】
前記イベント検出ステップは、
ガウス分布からなるユーザの推定位置情報と、ユーザ対応の確率値を示すユーザ確信度情報を含むイベント情報を生成するステップであり、
前記情報統合処理部は、仮想的なユーザに対応するガウス分布からなるユーザ位置情報と、ユーザ対応の確率値を示すユーザの確信度情報を有するターゲットを複数設定したパーティクルを保持し、
前記情報統合処理ステップは、
各パーティクルに設定されたターゲットと、前記イベント情報との類似度の指標値であるイベント－ターゲット間尤度を算出して、イベント－ターゲット間尤度の高いターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行するステップであることを特徴とする請求項１３～１５いずれかに記載の情報処理方法。

【請求項１７】
前記情報統合処理ステップは、
前記イベント－ターゲット間尤度と、各パーティクルに設定したパーティクル重みとの総和データをターゲット重みとして算出し、ターゲット重みの大きいターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行するステップであることを特徴とする請求項１６に記載の情報処理方法。

【請求項１８】
前記情報統合処理ステップは、
各パーティクルに設定したイベント発生源仮説ターゲットと、前記イベント検出部から入力するイベント情報との尤度を算出し、該尤度の大小に応じた値をパーティクル重みと

して各パーティクルに設定するステップであることを特徴とする請求項１６に記載の情報処理方法。

【請求項１９】

前記情報統合処理ステップは、

前記パーティクル重みの大きいパーティクルを優先的に再選択するリサンプリング処理を実行して、パーティクルの更新処理を行うステップであることを特徴とする請求項１８に記載の情報処理方法。

【請求項２０】

前記情報統合処理ステップは、

各パーティクルに設定したターゲットについて、経過時間を考慮した更新処理を実行するステップであることを特徴とする請求項１３～１５いずれかに記載の情報処理方法。

【請求項２１】

前記情報統合処理ステップは、

各パーティクルに設定したイベント発生源仮説ターゲットについて、前記イベント検出部の生成するイベント情報を適用した更新処理を行うステップであることを特徴とする請求項１６に記載の情報処理方法。

【請求項２２】

前記情報統合処理ステップは、

前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和を、各ターゲット対応のユーザ位置情報およびユーザ識別情報としたターゲット情報を生成するステップであることを特徴とする請求項１８に記載の情報処理方法。

【請求項２３】

前記情報統合処理ステップは、

前記パーティクルの各々に設定したイベント発生源仮説ターゲットの数に応じて、イベント発生源の確率値としてのシクナル情報の生成を行うステップであることを特徴とする請求項１６に記載の情報処理方法。

【請求項２４】

前記情報統合処理ステップは、

前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和に含まれるユーザ位置情報としてのガウス分布データのピーク値が予め設定した閾値未満である場合に、該ターゲットを削除する処理を実行するステップを含むことを特徴とする請求項１８に記載の情報処理方法。

【請求項２５】

情報処理装置において情報解析処理を実行させるコンピュータ・プログラムであり、

複数の情報入力部に、実空間における画像情報または音声情報のいずれかを含む情報を入力させる情報入力ステップと、

イベント検出部に、前記情報入力ステップにおいて入力する情報の解析により、前記実空間に存在するユーザの推定位置情報および推定識別情報を含むイベント情報を生成させるイベント検出ステップと、

情報統合処理部に、ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、前記イベント情報に基づく仮説の更新および取捨選択により、前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報の生成を実行させる情報統合処理ステップを有し、

前記情報統合処理ステップは、

前記イベント検出部の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットと前記イベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応

するイベントーターゲット間尤度より大きい値である場合に、前記暫定ターゲットを各パーティクルに新規追加する処理を行わせるステップであることを特徴とするコンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、情報処理装置、および情報処理方法、並びにコンピュータ・プログラムに関する。さらに詳細には、外界からの入力情報、例えば画像、音声などの情報を入力し、入力情報に基づく外界環境の解析、具体的には言葉を発している人物の位置や誰であるか等の解析処理を実行する情報処理装置、および情報処理方法、並びにコンピュータ・プログラムに関する。

【背景技術】

【０００２】

人とＰＣやロボットなどの情報処理装置との相互間の処理、例えばコミュニケーションやインタラクティブ処理を行うシステムはマンーマシン　インタラクション　システムと呼ばれる。このマンーマシン　インタラクション　システムにおいて、ＰＣやロボット等の情報処理装置は、人のアクション例えば人の動作や言葉を認識するために画像情報や音声情報を入力して入力情報に基づく解析を行う。

【０００３】

人が情報を伝達する場合、言葉のみならずしぐさ、視線、表情など様々なチャネルを情報伝達チャネルとして利用する。このようなすべてのチャネルの解析をマシンにおいて行うことができれば、人とマシンとのコミュニケーションも人と人とのコミュニケーションと同レベルに到達することができる。このような複数のチャネル（モダリティ、モーダルとも呼ばれる）からの入力情報の解析を行うインタフェースは、マルチモーダルインタフェースと呼ばれ、近年、開発、研究が盛んに行われている。

【０００４】

例えばカメラによって撮影された画像情報、マイクによって取得された音声情報を入力して解析を行う場合、より詳細な解析を行うためには、様々なポイントに設置した複数のカメラおよび複数のマイクから多くの情報を入力することが有効である。

【０００５】

具体的なシステムとしては、例えば以下のようなシステムが想定される。情報処理装置（テレビ）が、カメラおよびマイクを介して、テレビの前のユーザ（父、母、姉、弟）の画像および音声を入力し、それぞれのユーザの位置やどのユーザが発した言葉であるか等を解析し、テレビが解析情報に応じた処理、例えば会話を行ったユーザに対するカメラのズームアップや、会話を行ったユーザに対する的確な応答を行うなどのシステムが実現可能となる。

【０００６】

従来の一般的なマンーマシン　インタラクション　システムの多くは、複数チャネル（モーダル）からの情報を決定論的に統合して、複数のユーザが、それぞれどこにいて、それらは誰で、誰がシグナルを発したのかを決定するという処理を行っていた。このようなシステムを開示した従来技術として、例えば特許文献１（特開２００５－２７１１３７号公報）、特許文献２（特開２００２－２６４０５１号公報）がある。

【０００７】

しかし、従来のシステムにおいて行われるマイクやカメラから入力される不確実かつ非同期なデータを利用した決定論的な統合処理方法ではロバスト性にかけ、精度の低いデータしか得られないという問題がある。実際のシステムにおいて、実環境で取得可能なセンサ情報、すなわちカメラからの入力画像やマイクから入力される音声情報には様々な余分な情報、例えばノイズや不要な情報が含まれる不確実なデータであり、画像解析や音声解析処理を行う場合には、このようなセンサ情報から有効な情報を効率的に統合する処理が重要となる。

【特許文献１】特開２００５－２７１１３７号公報
【特許文献２】特開２００２－２６４０５１号公報
【発明の開示】
【発明が解決しようとする課題】
【０００８】
本発明は、上述の問題点に鑑みてなされたものであり、複数のチャネル（モダリティ、モーダル）からの入力情報の解析、具体的には、例えば周囲にいる人物の位置などの特定処理を行うシステムにおいて、画像、音声情報などの様々な入力情報に含まれる不確実な情報に対する確率的な処理を行ってより精度の高いと推定される情報に統合する処理を行うことによりロバスト性を向上させ、精度の高い解析を行う情報処理装置、および情報処理方法、並びにコンピュータ・プログラム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】
【０００９】
本発明の第１の側面は、
実空間における画像情報または音声情報のいずれかを含む情報を入力する複数の情報入力部と、
前記情報入力部から入力する情報の解析により、前記実空間に存在するユーザの推定位置情報および推定識別情報を含むイベント情報を生成するイベント検出部と、
ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、前記イベント情報に基づく仮説の更新および取捨選択により、前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報の生成を実行する情報統合処理部を有し、
前記情報統合処理部は、
前記イベント検出部の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットと前記イベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベント－ターゲット間尤度より大きい値である場合に、前記暫定ターゲットを各パーティクルに新規追加する処理を行うことを特徴とする情報処理装置にある。
【００１０】
さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとして、画像フレームの異なる方向のフレーム外部位置に仮想ターゲットを設定した複数の異なる暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出して、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うことを特徴とする。
【００１１】
さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットの他、均一データによって構成されるユーザ位置情報、ユーザ識別情報を持つ暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出し、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うことを特徴とする。
【００１２】
さらに、本発明の情報処理装置の一実施態様において、前記イベント検出部は、ガウス

分布からなるユーザの推定位置情報と、ユーザ対応の確率値を示すユーザ確信度情報を含むイベント情報を生成する構成であり、前記情報統合処理部は、仮想的なユーザに対応するガウス分布からなるユーザ位置情報と、ユーザ対応の確率値を示すユーザの確信度情報を有するターゲットを複数設定したパーティクルを保持し、各パーティクルに設定されたターゲットと、前記イベント情報との類似度の指標値であるイベント－ターゲット間尤度を算出して、イベント－ターゲット間尤度の高いターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行する構成であることを特徴とする。

【００１３】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記イベント－ターゲット間尤度と、各パーティクルに設定したパーティクル重みとの総和データをターゲット重みとして算出し、ターゲット重みの大きいターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行する構成であることを特徴とする。

【００１４】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、各パーティクルに設定したイベント発生源仮説ターゲットと、前記イベント検出部から入力するイベント情報との尤度を算出し、該尤度の大小に応じた値をパーティクル重みとして各パーティクルに設定する構成であることを特徴とする。

【００１５】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記パーティクル重みの大きいパーティクルを優先的に再選択するリサンプリング処理を実行して、パーティクルの更新処理を行う構成であることを特徴とする。

【００１６】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、各パーティクルに設定したターゲットについて、経過時間を考慮した更新処理を実行する構成であることを特徴とする。

【００１７】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、各パーティクルに設定したイベント発生源仮説ターゲットについて、前記イベント検出部の生成するイベント情報を適用した更新処理を行う構成であることを特徴とする。

【００１８】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和を、各ターゲット対応のユーザ位置情報およびユーザ識別情報としたターゲット情報を生成する構成であることを特徴とする。

【００１９】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記パーティクルの各々に設定したイベント発生源仮説ターゲットの数に応じて、イベント発生源の確率値としてのシクナル情報の生成を行う構成であることを特徴とする。

【００２０】

さらに、本発明の情報処理装置の一実施態様において、前記情報統合処理部は、前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和に含まれるユーザ位置情報としてのガウス分布データのピーク値が予め設定した閾値未満である場合に、該ターゲットを削除する処理を実行する構成であることを特徴とする。

【００２１】

さらに、本発明の第２の側面は、

情報処理装置において情報解析処理を実行する情報処理方法であり、

複数の情報入力部が、実空間における画像情報または音声情報のいずれかを含む情報を入力する情報入力ステップと、

イベント検出部が、前記情報入力ステップにおいて入力する情報の解析により、前記実

空間に存在するユーザの推定位置情報および推定識別情報を含むイベント情報を生成するイベント検出ステップと、

情報統合処理部が、ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、前記イベント情報に基づく仮説の更新および取捨選択により、前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報の生成を実行する情報統合処理ステップを有し、

前記情報統合処理ステップは、

前記イベント検出部の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットと前記イベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベントーターゲット間尤度より大きい値である場合に、前記暫定ターゲットを各パーティクルに新規追加する処理を行うステップであることを特徴とする情報処理方法にある。

【００２２】

さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとして、画像フレームの異なる方向のフレーム外部位置に仮想ターゲットを設定した複数の異なる暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出して、算出した暫定ターゲットのイベントーターゲット間尤度の最大値が、既存ターゲットに対応するイベントーターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うステップであることを特徴とする。

【００２３】

さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットの他、均一データによって構成されるユーザ位置情報、ユーザ識別情報を持つ暫定ターゲットを生成し、生成した複数の暫定ターゲットと前記イベント情報との尤度を個別に算出し、算出した暫定ターゲットのイベントーターゲット間尤度の最大値が、既存ターゲットに対応するイベントーターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行うステップであることを特徴とする。

【００２４】

さらに、本発明の情報処理方法の一実施態様において、前記イベント検出ステップは、ガウス分布からなるユーザの推定位置情報と、ユーザ対応の確率値を示すユーザ確信度情報を含むイベント情報を生成するステップであり、前記情報統合処理部は、仮想的なユーザに対応するガウス分布からなるユーザ位置情報と、ユーザ対応の確率値を示すユーザの確信度情報を有するターゲットを複数設定したパーティクルを保持し、前記情報統合処理ステップは、各パーティクルに設定されたターゲットと、前記イベント情報との類似度の指標値であるイベントーターゲット間尤度を算出して、イベントーターゲット間尤度の高いターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行するステップであることを特徴とする。

【００２５】

さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記イベントーターゲット間尤度と、各パーティクルに設定したパーティクル重みとの総和データをターゲット重みとして算出し、ターゲット重みの大きいターゲットを優先的にイベント発生源仮説ターゲットとしたパーティクル設定処理を実行するステップであることを特徴とする。

【００２６】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、各パーティクルに設定したイベント発生源仮説ターゲットと、前記イベント検出部から入力するイベント情報との尤度を算出し、該尤度の大小に応じた値をパーティクル重みとして各パーティクルに設定するステップであることを特徴とする。
【００２７】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記パーティクル重みの大きいパーティクルを優先的に再選択するリサンプリング処理を実行して、パーティクルの更新処理を行うステップであることを特徴とする。
【００２８】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、各パーティクルに設定したターゲットについて、経過時間を考慮した更新処理を実行するステップであることを特徴とする。
【００２９】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、各パーティクルに設定したイベント発生源仮説ターゲットについて、前記イベント検出部の生成するイベント情報を適用した更新処理を行うステップであることを特徴とする。
【００３０】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和を、各ターゲット対応のユーザ位置情報およびユーザ識別情報としたターゲット情報を生成するステップであることを特徴とする。
【００３１】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記パーティクルの各々に設定したイベント発生源仮説ターゲットの数に応じて、イベント発生源の確率値としてのシクナル情報の生成を行うステップであることを特徴とする。
【００３２】
さらに、本発明の情報処理方法の一実施態様において、前記情報統合処理ステップは、前記パーティクルの各々に設定したターゲットデータと前記パーティクル重みとの積算総和に含まれるユーザ位置情報としてのガウス分布データのピーク値が予め設定した閾値未満である場合に、該ターゲットを削除する処理を実行するステップを含むことを特徴とする。
【００３３】
さらに、本発明の第３の側面は、
情報処理装置において情報解析処理を実行させるコンピュータ・プログラムであり、
複数の情報入力部に、実空間における画像情報または音声情報のいずれかを含む情報を入力させる情報入力ステップと、
イベント検出部に、前記情報入力ステップにおいて入力する情報の解析により、前記実空間に存在するユーザの推定位置情報および推定識別情報を含むイベント情報を生成させるイベント検出ステップと、
情報統合処理部に、ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、前記イベント情報に基づく仮説の更新および取捨選択により、前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報の生成を実行させる情報統合処理ステップを有し、
前記情報統合処理ステップは、
前記イベント検出部の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して前記実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、前記情報入力部を構成するカメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットと前記イベント検出部の生成するイベ

ント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベント－ターゲット間尤度より大きい値である場合に、前記暫定ターゲットを各パーティクルに新規追加する処理を行わせるステップであることを特徴とするコンピュータ・プログラムにある。

【００３４】

なお、本発明のコンピュータ・プログラムは、例えば、様々なプログラム・コードを実行可能な汎用コンピュータ・システムに対して、コンピュータ可読な形式で提供する記憶媒体、通信媒体によって提供可能なコンピュータ・プログラムである。このようなプログラムをコンピュータ可読な形式で提供することにより、コンピュータ・システム上でプログラムに応じた処理が実現される。

【００３５】

本発明のさらに他の目的、特徴や利点は、後述する本発明の実施例や添付する図面に基づくより詳細な説明によって明らかになるであろう。なお、本明細書においてシステムとは、複数の装置の論理的集合構成であり、各構成の装置が同一筐体内にあるものには限らない。

【発明の効果】

【００３６】

本発明の一実施例の構成によれば、カメラやマイクによって取得される画像情報や音声情報に基づいてユーザの推定位置および推定識別データを含むイベント情報を入力して、複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を行い、フィルタリングによる仮説の更新および取捨選択に基づいてユーザの位置および識別情報を生成する。また、カメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとイベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベント－ターゲット間尤度より大きい値である場合に、暫定ターゲットを各パーティクルに新規追加する処理を行う。本構成により、カメラの取得する画像フレームの外部にいるユーザからの音声入力イベントに対応した正しい推定処理が可能となり、ユーザ位置やユーザ識別情報を効率的に確実に生成することが可能となる。

【発明を実施するための最良の形態】

【００３７】

以下、図面を参照しながら本発明の実施形態に係る情報処理装置、および情報処理方法、並びにコンピュータ・プログラムの詳細について説明する。

【００３８】

まず、図１を参照して本発明に係る情報処理装置の実行する処理の概要について説明する。本発明の情報処理装置１００は、環境情報を入力するセンサ、ここでは一例としてカメラ２１と、複数のマイク３１～３４から画像情報、音声情報を入力し、これらの入力情報に基づいて環境の解析を行う。具体的には、複数のユーザ１，１１～４，１４の位置の解析、およびその位置にいるユーザの識別を行う。

【００３９】

図に示す例において、例えばユーザ１，１１～ユーザ４，１４が家族である父、母、姉、弟であるとき、情報処理装置１００は、カメラ２１と、複数のマイク３１～３４から入力する画像情報、音声情報の解析を行い、４人のユーザ１～４の存在する位置、各位置にいるユーザが父、母、姉、弟のいずれであるかを識別する。識別処理結果は様々な処理に利用される。例えば、例えば会話を行ったユーザに対するカメラのズームアップや、会話を行ったユーザに対してテレビから応答を行うなどの処理に利用される。

【００４０】

なお、本発明に係る情報処理装置１００の主要な処理は、複数の情報入力部（カメラ２１，マイク３１～３４）からの入力情報に基づいて、ユーザの位置識別およびユーザの特定処理としてのユーザ識別処理を行うことである。この識別結果の利用処理については特に限定するものではない。カメラ２１と、複数のマイク３１～３４から入力する画像情報

、音声情報には様々な不確実な情報が含まれる。本発明の情報処理装置１００では、これらの入力情報に含まれる不確実な情報に対する確率的な処理を行って、精度の高いと推定される情報に統合する処理を行う。この推定処理によりロバスト性を向上させ、精度の高い解析を行う。

【００４１】

図２に情報処理装置１００の構成例を示す。情報処理装置１００は、入力デバイスとして画像入力部（カメラ）１１１、複数の音声入力部（マイク）１２１ａ～ｄを有する。画像入力部（カメラ）１１１から画像情報を入力し、音声入力部（マイク）１２１から音声情報を入力し、これらの入力情報に基づいて解析を行う。複数の音声入力部（マイク）１２１ａ～ｄの各々は、図１に示すように様々な位置に配置されている。

【００４２】

複数のマイク１２１ａ～ｄから入力された音声情報は、音声イベント検出部１２２を介して音声・画像統合処理部１３１に入力される。音声イベント検出部１２２は、複数の異なるポジションに配置された複数の音声入力部（マイク）１２１ａ～ｄから入力する音声情報を解析し統合する。具体的には、音声入力部（マイク）１２１ａ～ｄから入力する音声情報に基づいて、発生した音の位置およびどのユーザの発生させた音であるかのユーザ識別情報を生成して音声・画像統合処理部１３１に入力する。

【００４３】

なお、情報処理装置１００の実行する具体的な処理は、例えば図１に示すように複数のユーザが存在する環境で、ユーザＡ～Ｄがどの位置にいて、会話を行ったユーザがどのユーザであるかを識別すること、すなわち、ユーザ位置およびユーザ識別を行うことであり、さらに声を発した人物などのイベント発生源を特定する処理である。

【００４４】

音声イベント検出部１２２は、複数の異なるポジションに配置された複数の音声入力部（マイク）１２１ａ～ｄから入力する音声情報を解析し、音声の発生源の位置情報を確率分布データとして生成する。具体的には、音源方向に関する期待値と分散データＮ（m_e，σ_e）を生成する。また、予め登録されたユーザの声の特徴情報との比較処理に基づいてユーザ識別情報を生成する。この識別情報も確率的な推定値として生成する。音声イベント検出部１２２には、予め検証すべき複数のユーザの声についての特徴情報が登録されており、入力音声と登録音声との比較処理を実行して、どのユーザの声である確率が高いかを判定する処理を行い、全登録ユーザに対する事後確率、あるいはスコアを算出する。

【００４５】

このように、音声イベント検出部１２２は、複数の異なるポジションに配置された複数の音声入力部（マイク）１２１ａ～ｄから入力する音声情報を解析し、音声の発生源の位置情報を確率分布データと、確率的な推定値からなるユーザ識別情報とによって構成される［統合音声イベント情報］を生成して音声・画像統合処理部１３１に入力する。

【００４６】

一方、画像入力部（カメラ）１１１から入力された画像情報は、画像イベント検出部１１２を介して音声・画像統合処理部１３１に入力される。画像イベント検出部１１２は、画像入力部（カメラ）１１１から入力する画像情報を解析し、画像に含まれる人物の顔を抽出し、顔の位置情報を確率分布データとして生成する。具体的には、顔の位置や方向に関する期待値と分散データＮ（m_e，σ_e）を生成する。また、予め登録されたユーザの顔の特徴情報との比較処理に基づいてユーザ識別情報を生成する。この識別情報も確率的な推定値として生成する。画像イベント検出部１１２には、予め検証すべき複数のユーザの顔についての特徴情報が登録されており、入力画像から抽出した顔領域の画像の特徴情報と登録された顔画像の特徴情報との比較処理を実行して、どのユーザの顔である確率が高いかを判定する処理を行い、全登録ユーザに対する事後確率、あるいはスコアを算出する。

【００４７】

なお、音声イベント検出部１２２や画像イベント検出部１１２において実行する音声識

別や、顔検出、顔識別処理は従来から知られる技術を適用する。例えば顔検出、顔識別処理としては以下の文献に開示された技術の適用が可能である。

佐部　浩太郎，日台　健一，"ピクセル差分特徴を用いた実時間任意姿勢顔検出器の学習"，第１０回画像センシングシンポジウム講演論文集，ｐｐ．５４７－５５２，２００４

特開２００４－３０２６４４（Ｐ２００４－３０２６４４Ａ）［発明の名称：顔識別装置、顔識別方法、記録媒体、及びロボット装置］

【００４８】

音声・画像統合処理部１３１は、音声イベント検出部１２２や画像イベント検出部１１２からの入力情報に基づいて、複数のユーザが、それぞれどこにいて、それらは誰で、誰が音声等のシグナルを発したのかを確率的に推定する処理を実行する。この処理については後段で詳細に説明する。音声・画像統合処理部１３１は、音声・画像統合処理部１３１は、音声イベント検出部１２２や画像イベント検出部１１２からの入力情報に基づいて、

（ａ）複数のユーザが、それぞれどこにいて、それらは誰であるかの推定情報としての［ターゲット情報］

（ｂ）例えば話しをしたユーザなどのイベント発生源を［シグナル情報］として、処理決定部１３２に出力する。

【００４９】

これらの識別処理結果を受領した処理決定部１３２は、識別処理結果を利用した処理を実行する、例えば、例えば会話を行ったユーザに対するカメラのズームアップや、会話を行ったユーザに対してテレビから応答を行うなどの処理を行う。

【００５０】

上述したように、音声イベント検出部１２２は、音声の発生源の位置情報を確率分布データ、具体的には、音源方向に関する期待値と分散データＮ（m_e，σ_e）を生成する。また、予め登録されたユーザの声の特徴情報との比較処理に基づいてユーザ識別情報を生成して音声・画像統合処理部１３１に入力する。また、画像イベント検出部１１２は、画像に含まれる人物の顔を抽出し、顔の位置情報を確率分布データとして生成する。具体的には、顔の位置や方向に関する期待値と分散データＮ（m_e，σ_e）を生成する。また、予め登録されたユーザの顔の特徴情報との比較処理に基づいてユーザ識別情報を生成して音声・画像統合処理部１３１に入力する。

【００５１】

図３を参照して、音声イベント検出部１２２および画像イベント検出部１１２が生成し音声・画像統合処理部１３１に入力する情報の例について説明する。図３（Ａ）は図１を参照して説明したと同様のカメラやマイクが備えられた実環境の例を示し、複数のユーザ１～ｋ，２０１～２０ｋが存在する。この環境で、あるユーザが話しをしたとすると、マイクで音声が入力される。また、カメラは連続的に画像を撮影している。

【００５２】

音声イベント検出部１２２および画像イベント検出部１１２が生成し音声・画像統合処理部１３１に入力する情報は、基本的に同様の情報であり、図３（Ｂ）に示す２つの情報によって構成される。すなわち、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）

これらの２つの情報である。これらの２つの情報は、イベントの発生毎に生成される。音声イベント検出部１２２は、音声入力部（マイク）１２１ａ～ｄから音声情報が入力された場合に、その音声情報に基づいて上記の（ａ）ユーザ位置情報、（ｂ）ユーザ識別情報を生成して音声・画像統合処理部１３１に入力する。画像イベント検出部１１２は、例えば予め定めた一定のフレーム間隔で、画像入力部（カメラ）１１１から入力された画像情報に基づいて（ａ）ユーザ位置情報、（ｂ）ユーザ識別情報を生成して音声・画像統合処理部１３１に入力する。なお、本例では、画像入力部（カメラ）１１１は１台のカメラを設定した例を示しており、１つのカメラに複数のユーザの画像が撮影される設定であり

、この場合、１つの画像に含まれる複数の顔の各々について（ａ）ユーザ位置情報、（ｂ）ユーザ識別情報を生成して音声・画像統合処理部１３１に入力する。

【００５３】

音声イベント検出部１２２が音声入力部（マイク）１２１ａ～ｄから入力する音声情報に基づいて、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（話者識別情報）

これらの情報を生成する処理について説明する。

【００５４】

音声イベント検出部１２２による（ａ）ユーザ位置情報の生成処理

音声イベント検出部１２２は、音声入力部（マイク）１２１ａ～ｄから入力された音声情報に基づいて解析された声を発したユーザ、すなわち［話者］の位置の推定情報を生成する。すなわち、話者が存在すると推定される位置を、期待値（平均）［m_e］と分散情報［σ_e］からなるガウス分布（正規分布）データＮ（m_e，σｅ）として生成する。

【００５５】

音声イベント検出部１２２による（ｂ）ユーザ識別情報（話者識別情報）の生成処理

音声イベント検出部１２２は、音声入力部（マイク）１２１ａ～ｄから入力された音声情報に基づいて話者が誰であるかを、入力音声と予め登録されたユーザ１～ｋの声の特徴情報との比較処理により推定する。具体的には話者が各ユーザ１～ｋである確率を算出する。この算出値を（ｂ）ユーザ識別情報（話者識別情報）とする。例えば入力音声の特徴と最も近い登録された音声特徴を有するユーザに最も高いスコアを配分し、最も異なる特徴を持つユーザに最低のスコア（例えば０）を配分する処理によって各ユーザである確率を設定したデータを生成して、これを（ｂ）ユーザ識別情報（話者識別情報）とする。

【００５６】

画像イベント検出部１１２が画像入力部（カメラ）１１１から入力する画像情報に基づいて、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報）

これらの情報を生成する処理について説明する。

【００５７】

画像イベント検出部１１２による（ａ）ユーザ位置情報の生成処理

画像イベント検出部１１２は、画像入力部（カメラ）１１１から入力された画像情報に含まれる顔の各々について顔の位置の推定情報を生成する。すなわち、画像から検出された顔が存在すると推定される位置を、期待値（平均）［m_e］と分散情報［σ_e］からなるガウス分布（正規分布）データＮ（m_e，σ_e）として生成する。

【００５８】

画像イベント検出部１１２による（ｂ）ユーザ識別情報（顔識別情報）の生成処理

画像イベント検出部１１２は、画像入力部（カメラ）１１１から入力された画像情報に基づいて、画像情報に含まれる顔を検出し、各顔が誰であるかを、入力画像情報と予め登録されたユーザ１～ｋの顔の特徴情報との比較処理により推定する。具体的には抽出された各顔が各ユーザ１～ｋである確率を算出する。この算出値を（ｂ）ユーザ識別情報（顔識別情報）とする。例えば入力画像に含まれる顔の特徴と最も近い登録された顔の特徴を有するユーザに最も高いスコアを配分し、最も異なる特徴を持つユーザに最低のスコア（例えば０）を配分する処理によって各ユーザである確率を設定したデータを生成して、これを（ｂ）ユーザ識別情報（顔識別情報）とする。

【００５９】

なお、カメラの撮影画像から複数の顔が検出された場合には、各検出顔に応じて、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報）

これらの情報を生成して、音声・画像統合処理部１３１に入力する。

また、本例では、画像入力部１１１として１台のカメラを利用した例を説明するが、複数のカメラの撮影画像を利用してもよく、その場合は、画像イベント検出部１１２は、各カメラの撮影画像の各々に含まれる各顔について、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報）

これらの情報を生成して、音声・画像統合処理部１３１に入力する。

【００６０】

次に、音声・画像統合処理部１３１の実行する処理について説明する。音声・画像統合処理部１３１は、上述したように、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示す２つの情報、すなわち、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）

これらの情報を逐次入力する。なお、これらの各情報の入力タイミングは様々な設定が可能であるが、例えば、音声イベント検出部１２２は新たな音声が入力された場合に上記（ａ），（ｂ）の各情報を音声イベント情報として生成して入力し、画像イベント検出部１１２は、一定のフレーム周期単位で、上記（ａ），（ｂ）の各情報を音声イベント情報として生成して入力するといった設定が可能である。

【００６１】

音声・画像統合処理部１３１の実行する処理について、図４以下を参照して説明する。音声・画像統合処理部１３１は、ユーザの位置および識別情報についての仮説（Ｈｙｐｏｔｈｅｓｉｓ）の確率分布データを設定し、その仮説を入力情報に基づいて更新することで、より確からしい仮説のみを残す処理を行う。この処理手法として、パーティクル・フィルタ（Ｐａｒｔｉｃｌｅ　Ｆｉｌｔｅｒ）を適用した処理を実行する。

【００６２】

パーティクル・フィルタ（Ｐａｒｔｉｃｌｅ　Ｆｉｌｔｅｒ）を適用した処理は、様々な仮説、本例では、ユーザの位置と誰であるかの仮説に対応するパーティクルを多数設定し、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示す２つの情報、すなわち、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）

これらの入力情報に基づいて、より確からしいパーティクルのウェイトを高めていくという処理を行う。

【００６３】

パーティクル・フィルタ（Ｐａｒｔｉｃｌｅ　Ｆｉｌｔｅｒ）を適用した基本的な処理例について図４を参照して説明する。例えば、図４に示す例は、あるユーザに対応する存在位置をパーティクル・フィルタにより推定する処理例を示している。図４に示す例は、ある直線上の１次元領域におけるユーザ３０１の存在する位置を推定する処理である。

【００６４】

初期的な仮説（Ｈ）は、図４（ａ）に示すように均一なパーティクル分布データとなる。次に、画像データ３０２が取得され、取得画像に基づくユーザ３０１の存在確率分布データが図４（ｂ）のデータとして取得される。この取得画像に基づく確率分布データに基づいて、図４（ａ）のパーティクル分布データが更新され、図４（ｃ）の更新された仮説確率分布データが得られる。このような処理を、入力情報に基づいて繰り返し実行して、ユーザのより確からしい位置情報を得る。

【００６５】

なお、パーティクル・フィルタを用いた処理の詳細については、例えば［Ｄ．　Ｓｃｈｕｌｚ，　Ｄ．　Ｆｏｘ，　ａｎｄ　Ｊ．　Ｈｉｇｈｔｏｗｅｒ．　Ｐｅｏｐｌｅ　Ｔｒａｃｋｉｎｇ　ｗｉｔｈ　Ａｎｏｎｙｍｏｕｓ　ａｎｄ　ＩＤ－ｓｅｎｓｏｒｓ　Ｕｓｉｎｇ　Ｒａｏ－Ｂｌａｃｋｗｅｌｌｉｓｅｄ　Ｐａｒｔｉｃｌｅ　Ｆｉｌｔｅｒｓ．Ｐｒｏｃ．　ｏｆ　ｔｈｅ　Ｉｎｔｅｒｎａｔｉｏｎａｌ　Ｊｏｉｎｔ　Ｃｏｎｆｅｒｅｎｃ

ｅ ｏｎ Ａｒｔｉｆｉｃｉａｌ Ｉｎｔｅｌｌｉｇｅｎｃｅ （ＩＪＣＡＩ－０３）］に記載されている。

【００６６】

図４に示す処理例は、ユーザの存在位置のみについて、入力情報を画像データのみとした処理例として説明しており、パーティクルの各々は、ユーザ３０１の存在位置のみの情報を有している。

【００６７】

一方、本発明に従った処理は、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示す２つの情報、すなわち、

（ａ）ユーザ位置情報

（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）

これらの入力情報に基づいて、複数のユーザの位置と複数のユーザがそれぞれ誰であるかを判別する処理を行うことになる。従って、本発明におけるパーティクル・フィルタ（Ｐａｒｔｉｃｌｅ Ｆｉｌｔｅｒ）を適用した処理では、音声・画像統合処理部１３１が、ユーザの位置と誰であるかの仮説に対応するパーティクルを多数設定して、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示す２つの情報に基づいて、パーティクル更新を行うことになる。

【００６８】

図５を参照して、本処理例で設定するパーティクルの構成について説明する。音声・画像統合処理部１３１は、予め設定した数＝ｍのパーティクルを有する。図５に示すパーティクル１～ｍである。各パーティクルには識別子としてのパーティクルＩＤ（ＰＩＤ＝１～ｍ）が設定されている。

【００６９】

各パーティクルに、位置および識別を行うオブジェクトに対応する仮想的なオブジェクトに対応する複数のターゲットを設定する。本例では、例えば実空間に存在すると推定される人数以上の仮想のユーザに対応する複数のターゲットを各パーティクルに設定する。ｍ個のパーティクルの各々はターゲット単位でデータをターゲット数分保持する。図５に示す例では、１つのパーティクルにｎ個のターゲットが含まれる。各パーティクルに含まれるターゲット各々が有するターゲットデータの構成を図６に示す。

【００７０】

各パーティクルに含まれる各ターゲットデータについて図６を参照して説明する。図６は、図５に示すパーティクル１（ｐＩＤ＝１）に含まれる１つのターゲット（ターゲットＩＤ：ｔＩＤ＝ｎ）３１１のターゲットデータの構成である。ターゲット３１１のターゲットデータは、図６に示すように、以下のデータ、すなわち、

（ａ）各ターゲット各々に対応する存在位置の確率分布［ガウス分布：Ｎ（m_{1n}，σ_{1n}）］、

（ｂ）各ターゲットが誰であるかを示すユーザ確信度情報（ｕＩＤ）

uID_{1n1}＝０．０

uID_{1n2}＝０．１

：

uID_{1nk}＝０．５

これらのデータによって構成される。

【００７１】

なお、（ａ）に示すガウス分布：Ｎ（m_{1n}，σ_{1n}）における［m_{1n}，σ_{1n}］の（１ｎ）は、パーティクルＩＤ：ｐＩＤ＝１におけるターゲットＩＤ：ｔＩＤ＝ｎに対応する存在確率分布としてのガウス分布であることを意味する。

また、（ｂ）に示すユーザ確信度情報（ｕＩＤ）における、［uID_{1n1}］に含まれる（１ｎ１）は、パーティクルＩＤ：ｐＩＤ＝１におけるターゲットＩＤ：ｔＩＤ＝ｎの、ユーザ＝ユーザ１である確率を意味する。すなわちターゲットＩＤ＝ｎのデータは、

ユーザ１である確率が０．０、

ユーザ２である確率が０．１、
：
ユーザｋである確率が０．５、
であることを意味している。
【００７２】
図５に戻り、音声・画像統合処理部１３１の設定するパーティクルについての説明を続ける。図５に示すように、音声・画像統合処理部１３１は、予め決定した数＝ｍのパーティクル（ＰＩＤ＝１～ｍ）を設定し、各パーティクルは、実空間に存在すると推定されるターゲット（ｔＩＤ＝１～ｎ）各々について、
（ａ）各ターゲット各々に対応する存在位置の確率分布［ガウス分布：Ｎ（ｍ，σ）］、
（ｂ）各ターゲットが誰であるかを示すユーザ確信度情報（ｕＩＤ）
これらのターゲットデータを有する。
【００７３】
音声・画像統合処理部１３１は、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示すイベント情報、すなわち、
（ａ）ユーザ位置情報
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）
これらのイベント情報を入力してｍ個のパーティクル（ＰＩＤ＝１～ｍ）の更新処理を行う。
【００７４】
音声・画像統合処理部１３１、これらの更新処理を実行して、
（ａ）複数のユーザが、それぞれどこにいて、それらは誰であるかの推定情報としての［ターゲット情報］、
（ｂ）例えば話をしたユーザなどのイベント発生源を示す［シグナル情報］、
これらを生成して処理決定部１３２に出力する。
【００７５】
［ターゲット情報］は、図５の右端のターゲット情報３０５に示すように、各パーティクル（ＰＩＤ＝１～ｍ）に含まれる各ターゲット（ｔＩＤ＝１～ｎ）対応データの重み付き総和データとして生成される。各パーティクルの重みについては後述する。
【００７６】
ターゲット情報３０５は、音声・画像統合処理部１３１が予め設定した仮想的なユーザに対応するターゲット（ｔＩＤ＝１～ｎ）の
（ａ）存在位置
（ｂ）誰であるか（ｕＩＤ１～ｕＩＤｋのいずれであるか）
これらを示す情報である。このターゲット情報は、パーティクルの更新に伴い、順次更新されることになり、例えばユーザ１～ｋが実環境内で移動しない場合、ユーザ１～ｋの各々が、ｎ個のターゲット（ｔＩＤ＝１～ｎ）から選択されたｋ個にそれぞれ対応するデータとして収束することになる。
【００７７】
例えば、図５に示すターゲット情報３０５中の最上段のターゲット１（ｔＩＤ＝１）のデータ中に含まれるユーザ確信度情報（ｕＩＤ）は、ユーザ２（uID_{12}＝０．７）について最も高い確率を有している。従って、このターゲット１（ｔＩＤ＝１）のデータは、ユーザ２に対応するものであると推定されることになる。なお、ユーザ確信度情報（ｕＩＤ）を示すデータ［uID_{12}＝０．７］中の（uID_{12}）内の（１２）は、ターゲットＩＤ＝１のユーザ＝２のユーザ確信度情報（ｕＩＤ）に対応する確率であることを示している。
【００７８】
このターゲット情報３０５中の最上段のターゲット１（ｔＩＤ＝１）のデータは、ユーザ２である確率が最も高く、このユーザ２は、その存在位置が、ターゲット情報３０５中

の最上段のターゲット１（ｔＩＤ＝１）のデータに含まれる存在確率分布データに示す範囲にいると推定されることなる。

【００７９】

このように、ターゲット情報３０５は、初期的に仮想的なオブジェクト（仮想ユーザ）として設定した各ターゲット（ｔＩＤ＝１～ｎ）の各々について、

（ａ）存在位置

（ｂ）誰であるか（ｕＩＤ１～ｕＩＤｋのいずれであるか）

の各情報を示す。従って、各ターゲット（ｔＩＤ＝１～ｎ）のｋ個のターゲット情報の各々は、ユーザが移動しない場合は、ユーザ１～ｋに対応するように収束する。

【００８０】

ターゲット（ｔＩＤ＝１～ｎ）の数がユーザ数ｋより大きい場合、どのユーザにも対応しないターゲットが発生する。例えば、ターゲット情報３０５中の最下段のターゲット（ｔＩＤ＝ｎ）は、ユーザ確信度情報（ｕＩＤ）も最大で０．５であり、存在確率分布データも大きなピークを有していない。このようなデータは特定のユーザに対応するデータではないと判定される。なお、このようなターゲットについては、削除するような処理が行われる場合もある。ターゲットの削除処理については後述する。

【００８１】

先に説明したように、音声・画像統合処理部１３１は、入力情報に基づくパーティクルの更新処理を実行して、

（ａ）複数のユーザが、それぞれどこにいて、それらは誰であるかの推定情報としての［ターゲット情報］、

（ｂ）例えば話をしたユーザなどのイベント発生源を示す［シグナル情報］、

これらを生成して処理決定部１３２に出力する。

【００８２】

ターゲット情報は、図５のターゲット情報３０５を参照して説明した情報である。音声・画像統合処理部１３１は、このターゲット情報の他に話をしたユーザなどのイベント発生源を示す［シグナル情報］についても生成して出力する。イベント発生源を示す［シグナル情報］は、音声イベントについては、誰が話をしたか、すなわち［話者］を示すデータであり、画像イベントについては、画像に含まれる顔が誰であるかを示すデータである。なお、画像イベントの場合のシグナル情報は、本例では結果としてターゲット情報のユーザ確信度情報（ｕＩＤ）から得られるものと一致することになる。

【００８３】

音声・画像統合処理部１３１が、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示すイベント情報、すなわち、ユーザ位置情報と、ユーザ識別情報（顔識別情報または話者識別情報）、これらのイベント情報を入力して、

（ａ）複数のユーザが、それぞれどこにいて、それらは誰であるかの推定情報としての［ターゲット情報］、

（ｂ）例えば話をしたユーザなどのイベント発生源を示す［シグナル情報］、

これらの情報を生成して処理決定部１３２に出力する処理について、図７以下を参照して説明する。

【００８４】

図７は、音声・画像統合処理部１３１の実行する処理シーケンスを説明するフローチャートを示す図である。まず、ステップＳ１０１において、音声・画像統合処理部１３１は、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示すイベント情報、すなわち、ユーザ位置情報と、ユーザ識別情報（顔識別情報または話者識別情報）、これらのイベント情報を入力する。

【００８５】

イベント情報の取得に成功した場合は、ステップＳ１０２に進み、イベント情報の取得に失敗した場合は、ステップＳ１２１に進む。ステップＳ１２１の処理については後段で説明する。

【００８６】
イベント情報の取得に成功した場合は、音声・画像統合処理部１３１は、ステップＳ１０２以下において、入力情報に基づくパーティクル更新処理を行うことになるが、パーティクル更新処理の前にステップＳ１０２において、図５に示すｍ個のパーティクル（ｐＩＤ＝１～ｍ）の各々にイベントの発生源の仮説を設定する。イベント発生源とは、例えば、音声イベントであれば、話をしたユーザがイベント発生源であり、画像イベントであれば、抽出した顔を持つユーザがイベント発生源である。
【００８７】
図５に示す例では、各パーティクルの最下段にイベント発生源の仮設データ（ｔＩＤ＝ｘｘ）を示している。図５の例では、
パーティクル１（ｐＩＤ＝１）は、ｔＩＤ＝２、
パーティクル２（ｐＩＤ＝２）は、ｔＩＤ＝ｎ、
：
パーティクルｍ（ｐＩＤ＝ｍ）は、ｔＩＤ＝ｎ、
このように各パーティクルについて、イベント発生源がターゲット１～ｎのいずれであるかの仮説を設定する。図５に示す例では、各パーティクルについて、仮説として設定したイベント発生源のターゲットデータを二重線で囲んで示している。
【００８８】
このイベント発生源の仮説設定は、入力イベントに基づくパーティクル更新処理を行う前に毎回実行する。すなわち、各パーティクル１～ｍ各々にイベントの発生源仮説を設定して、その仮説の下で、イベントとして音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示すイベント情報、すなわち、
（ａ）ユーザ位置情報
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）
これらのイベント情報を入力してｍ個のパーティクル（ＰＩＤ＝１～ｍ）の更新処理を行う。
【００８９】
パーティクル更新処理が行われた場合は、各パーティクル１～ｍ各々に設定されていたイベントの発生源の仮説はリセットされて、各パーティクル１～ｍ各々に新たな仮説の設定が行われる。この仮説の設定態様としては、
（１）ランダムな設定、
（２）音声・画像統合処理部１３１の有する内部モデルに従って設定、
上記（１），（２）のいずれかの手法で設定することが可能である。なお、パーティクルの数：ｍは、ターゲットの数：ｎより大きく設定されているので、複数のパーティクルが同一のターゲットをイベント発生源とした仮設に設定される。例えば、ターゲットの数：ｎが１０とした場合、パーティクル数：ｍ＝１００～１０００程度に設定した処理などが行われる。
【００９０】
上記の（２）音声・画像統合処理部１３１の有する内部モデルに従って仮説を設定する処理の具体的処理例について説明する。
音声・画像統合処理部１３１は、まず、音声イベント検出部１２２および画像イベント検出部１１２から取得したイベント情報、すなわち、図３（Ｂ）に示す２つの情報、すなわち、
（ａ）ユーザ位置情報
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）
これらのイベント情報と、
音声・画像統合処理部１３１の保持するパーティクルのターゲットの持つデータとの比較によって、各ターゲットの重み［W_{tID}］を算出し、算出した各ターゲットの重み［W_{tID}］に基づいて、各パーティクル（ｐＩＤ＝１～ｍ）に対するイベント発生源の仮説を設定する。以下、具体的な処理例について説明する。

【００９１】
なお、初期状態では、各パーティクル（ｐＩＤ＝１～ｍ）に設定されるイベント発生源の仮説は均等な設定とする。すなわちｎ個のターゲット（ｔＩＤ＝１～ｎ）を持つｍ個のパーティクル（ｐＩＤ＝１～ｍ）が設定されている構成では、
ターゲット１（ｔＩＤ＝１）をイベント発生源とするパーティクルをｍ／ｎ個、
ターゲット２（ｔＩＤ＝２）をイベント発生源とするパーティクルをｍ／ｎ個、
：
ターゲットｎ（ｔＩＤ＝ｎ）をイベント発生源とするパーティクルをｍ／ｎ個、
というように、各パーティクル（ｐＩＤ＝１～ｍ）に設定する初期的なイベント発生源の仮説ターゲット（ｔＩＤ＝１～ｎ）を均等に割り振る設定とする。
【００９２】
図７に示すフローのステップＳ１０１において、音声・画像統合処理部１３１が音声イベント検出部１２２および画像イベント検出部１１２からイベント情報、すなわち、図３（Ｂ）に示す２つの情報、すなわち、
（ａ）ユーザ位置情報
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）
これらのイベント情報を取得して、イベント情報の取得に成功すると、ステップＳ１０２において、音声・画像統合処理部１３１は、ｍ個のパーティクル（ＰＩＤ＝１～ｍ）の各々に対して、イベント発生源の仮説ターゲット（ｔＩＤ＝１～ｎ）を設定する。
【００９３】
ステップＳ１０２におけるパーティクル対応の仮説ターゲットの設定の詳細について説明する。音声・画像統合処理部１３１は、まず、ステップＳ１０１で入力したイベント情報と、音声・画像統合処理部１３１の保持するパーティクルのターゲットの持つデータとの比較を行い、比較結果を用いて、各ターゲットのターゲット重み［W_{tID}］を算出する。
【００９４】
ターゲット重み［W_{tID}］の算出処理の詳細について図８を参照して説明する。ターゲット重みの算出は、図８の右端に示すように、各パーティクルに設定されるターゲット１～ｎの各々に対応するｎ個のターゲット重みの算出処理として実行される。このｎ個のターゲット重みの算出に際しては、まず、図８（１）に示す入力イベント情報、すなわち、音声・画像統合処理部１３１が、音声イベント検出部１２２および画像イベント検出部１１２から入力したイベント情報と、各パーティクルの各ターゲットデータとの類似度の指標値としての尤度算出を行う。
【００９５】
図８（２）に示す尤度算出処理例は、（１）入力イベント情報と、パーティクル１の１つのターゲットデータ（ｔＩＤ＝ｎ）との比較によるイベント－ターゲット間尤度の算出例を説明する図である。なお、図８には、１つのターゲットデータとの比較例を示しているが、各パーティクルの各ターゲットデータについて、同様の尤度算出処理を実行する。
【００９６】
図８の下段に示す（２）尤度算出処理について説明する。図８（２）に示すように、尤度算出処理は、まず、
（ａ）ユーザ位置情報についてのイベントと、ターゲットデータとの類似度データとしてのガウス分布間尤度［ＤＬ］、
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）についてのイベントと、ターゲットデータとの類似度データとしてのユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］
これらを個別に算出する。
【００９７】
まず、（ａ）ユーザ位置情報についてのイベントと、ターゲットデータとの類似度データとしてのガウス分布間尤度［ＤＬ］の算出処理について説明する。
図８（１）に示す入力イベント情報中の、ユーザ位置情報に対応するガウス分布をＮ（

m_e，σ_e）とし、
音声・画像統合処理部１３１の保持する内部モデルのあるパーティクルが持つあるターゲットのユーザ位置情報に対応するガウス分布をＮ（m_t，σ_t）とする。図８に示す例では、パーティクル１（ｐＩＤ＝１）のターゲットｎ（ｔＩＤ＝ｎ）のターゲットデータに含まれるガウス分布をＮ（m_t，σ_t）とする。
【００９８】
これら２つのデータのガウス分布の類似度を判定する指標としてのガウス分布間尤度［ＤＬ］は、以下の式によって算出する。
$DL = N(m_t, \sigma_t + \sigma_e) x \mid m_e$
上記式は、中心m_tで分散$\sigma_t + \sigma_e$のガウス分布において$x = m_e$の位置の値を算出する式である。
【００９９】
次に、（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）についてのイベントと、ターゲットデータとの類似度データとしてのユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］の算出処理について説明する。
図８（１）に示す入力イベント情報中の、ユーザ確信度情報（ｕＩＤ）の各ユーザ１～ｋの確信度の値（スコア）を$P_e[i]$とする。なお、ｉはユーザ識別子１～ｋに対応する変数である。
音声・画像統合処理部１３１の保持する内部モデルのあるパーティクルが持つあるターゲットのユーザ確信度情報（ｕＩＤ）の各ユーザ１～ｋの確信度の値（スコア）を$P_t[i]$とする。図８に示す例では、パーティクル１（ｐＩＤ＝１）のターゲットｎ（ｔＩＤ＝ｎ）のターゲットデータに含まれるユーザ確信度情報（ｕＩＤ）の各ユーザ１～ｋの確信度の値（スコア）を$P_t[i]$とする。
【０１００】
これら２つのデータのユーザ確信度情報（ｕＩＤ）の類似度を判定する指標としてのユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］は、以下の式によって算出する。
$UL = \Sigma P_e[i] \times P_t[i]$
上記式は、２つのデータのユーザ確信度情報（ｕＩＤ）に含まれる各対応ユーザの確信度の値（スコア）の積の総和を求める式であり、この値をユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］とする。
【０１０１】
もしくは、ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］として、各積の最大値、すなわち、
$UL = arg\ max(P_e[i] \times P_t[i])$
上記の値を算出し、この値をユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］として利用する構成としてもよい。
【０１０２】
入力イベント情報とあるパーティクル（ｐＩＤ）が持つ１つのターゲット（ｔＩＤ）との類似度の指標としてのイベントーターゲット間尤度［$L_{pID,tID}$］は、上記の２つの尤度、すなわち、
ガウス分布間尤度［ＤＬ］と、
ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］
これら２つの尤度を利用して算出する。すなわち重みα（α＝０～１）を用いて、イベントーターゲット間尤度［$L_{pID,tID}$］は下式によって算出する。
$[L_{pID,tID}] = UL^{\alpha} \times DL^{1-\alpha}$
としてイベントとターゲットとの類似度の指標であるイベントーターゲット間尤度［$L_{pID,tID}$］を算出する。
ただし、α＝０～１とする。
【０１０３】
このイベントーターゲット間尤度［$L_{pID,tID}$］は、各パーティクルの各ターゲ

ットについて各々算出し、このイベントーターゲット間尤度［$L_{pID,tID}$］に基づいて各ターゲットのターゲット重み［W_{tID}］を算出する。

【０１０４】

なお、イベントーターゲット間尤度［$L_{pID,tID}$］の算出に適用する重み［α］は、予め固定された値としてもよいし、入力イベントに応じて値を変更する設定としてもよい。例えば入力イベントが画像である場合において、顔検出に成功し位置情報は取得できたが顔識別に失敗した場合などは、α＝０の設定として、ユーザ確信度情報（ｕＩＤ）間尤度：ＵＬ＝１としてガウス分布間尤度［ＤＬ］のみに依存してイベントーターゲット間尤度［$L_{pID,tID}$］を算出して、ガウス分布間尤度［ＤＬ］のみに依存したターゲット重み［W_{tID}］を算出する構成としてもよい。

【０１０５】

また、入力イベントが音声である場合において、話者識別に成功し話者情報破取得できたが、位置情報の取得に失敗した場合などは、α＝０の設定として、ガウス分布間尤度［ＤＬ］＝１として、ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］のみに依存してイベントーターゲット間尤度［$L_{pID,tID}$］を算出して、ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］のみに依存したターゲット重み［W_{tID}］を算出する構成としてもよい。

【０１０６】

イベントーターゲット間尤度［$L_{pID,tID}$］に基づく、ターゲット重み［W_{tID}］の算出式は、以下の通りである。

【数１】

$$W_{tID} = \sum_{pID}^{m} W_{pID} L_{pID,tID}$$

【０１０７】

とする。なお、上記式において、［W_{pID}］は、各パーティクル各々に設定されるパーティクル重みである。パーティクル重み［W_{pID}］の算出処理については後段で説明する。パーティクル重み［W_{pID}］は初期状態では、すべてのパーティクル（ｐＩＤ＝１～ｍ）において均一な値が設定される。

【０１０８】

図７に示すフローにおけるステップＳ１０１の処理、すなわち、各パーティクル対応のイベント発生源仮説の生成は、上記のイベントーターゲット間尤度［$L_{pID,tID}$］に基づいて算出したターゲット重み［W_{tID}］に基づいて実行する。ターゲット重み［W_{tID}］は、パーティクルに設定されるターゲット１～ｎ（ｔＩＤ＝１～ｎ）に対応したｎ個のデータが算出される。

【０１０９】

ｍ個のパーティクル（ｐＩＤ＝１～ｍ）各々に対するイベント発生源仮説ターゲットは、ターゲット重み［W_{tID}］の比率に応じて割り振る設定とする。

例えばｎ＝４で、ターゲット１～４（ｔＩＤ＝１～４）に対応して算出されたターゲット重み［W_{tID}］が、

ターゲット１：ターゲット重み＝３

ターゲット２：ターゲット重み＝２

ターゲット３：ターゲット重み＝１

ターゲット４：ターゲット重み＝５

である場合、ｍ個のパーティクルのイベント発生源仮説ターゲットを

ｍ個のパーティクル中の３０％をイベント発生源仮説ターゲット１、

ｍ個のパーティクル中の２０％をイベント発生源仮説ターゲット２、

ｍ個のパーティクル中の１０％をイベント発生源仮説ターゲット３、

m個のパーティクル中の５０％をイベント発生源仮説ターゲット４、
このような設定とする。
すなわちパーティクルに設定するイベント発生源仮説ターゲットをターゲットの重みに応じた配分比率とする。

【０１１０】

この仮説設定の後、図７に示すフローのステップＳ１０３に進む。ステップＳ１０３では、各パーティクル対応の重み、すなわちパーティクル重み［W_{pID}］の算出を行う。このパーティクル重み［W_{pID}］は前述したように、初期的には各パーティクルに均一な値が設定されるが、イベント入力に応じて更新される。

【０１１１】

図９、図１０を参照して、パーティクル重み［W_{pID}］の算出処理の詳細について説明する。パーティクル重み［W_{pID}］は、イベント発生源の仮説ターゲットを生成した各パーティクルの仮説の正しさの指標に相当する。パーティクル重み［W_{pID}］は、m個のパーティクル（ｐＩＤ＝１～m）の各々において設定されたイベント発生源の仮説ターゲットと、入力イベントとの類似度であるイベントーターゲット間尤度として算出される。

【０１１２】

図９には、音声・画像統合処理部１３１が、音声イベント検出部１２２および画像イベント検出部１１２から入力するイベント情報４０１と、音声・画像統合処理部１３１が、が保持するパーティクル４１１～４１３を示している。核パーティクル４１１｜４１３には、前述した処理、すなわち、図７に示すフローのステップＳ１０２におけるイベント発生源の仮説設定において設定された仮説ターゲットが１つずつ設定されている。図９中に示す例では、
パーティクル１（ｐＩＤ＝１）４１１におけるターゲット２（ｔＩＤ＝２）４２１、
パーティクル２（ｐＩＤ＝２）４１２におけるターゲットｎ（ｔＩＤ＝ｎ）４２２、
パーティクルｍ（ｐＩＤ＝ｍ）４１３におけるターゲットｎ（ｔＩＤ＝ｎ）４２３、
これらの仮説ターゲットである。

【０１１３】

図９の例において、各パーティクルのパーティクル重み［W_{pID}］は、
パーティクル１：イベント情報４０１とターゲット２（ｔＩＤ＝２）４２１とのイベントーターゲット間尤度、
パーティクル２：イベント情報４０１とターゲットｎ（ｔＩＤ＝ｎ）４２２とのイベントーターゲット間尤度、
パーティクルｍ：イベント情報４０１とターゲットｎ（ｔＩＤ＝ｎ）４２３とのイベントーターゲット間尤度、
これらのイベントーターゲット間尤度に対応することになる。

【０１１４】

図１０は、パーティクル１（ｐＩＤ＝１）のパーティクル重み［W_{pID}］算出処理例を示している。図１０（２）に示すパーティクル重み［W_{pID}］算出処理は、先に、図８（２）を参照して説明したと同様の尤度算出処理であり、本例では、（１）入力イベント情報と、パーティクルから選択された唯一の仮説ターゲットとの類似度指標としてのイベントーターゲット間尤度の算出として実行される。

【０１１５】

図１０の下段に示す（２）尤度算出処理も、先に図８（２）を参照して説明したと同様、
（ａ）ユーザ位置情報についてのイベントと、ターゲットデータとの類似度データとしてのガウス分布間尤度［ＤＬ］、
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）についてのイベントと、ターゲットデータとの類似度データとしてのユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］
これらを個別に算出する。

【０１１６】
（ａ）ユーザ位置情報についてのイベントと、仮説ターゲットとの類似度データとしてのガウス分布間尤度［ＤＬ］の算出処理は以下の処理となる。
入力イベント情報中の、ユーザ位置情報に対応するガウス分布をＮ（m_e，σ_e）、
パーティクルから選択された仮説ターゲットのユーザ位置情報に対応するガウス分布をＮ（m_t，σ_t）、
として、ガウス分布間尤度［ＤＬ］を、以下の式によって算出する。
$DL = N(m_t, \sigma_t + \sigma_e)\ x \mid m_e$
上記式は、中心m_tで分散$\sigma_t + \sigma_e$のガウス分布において$x = m_e$の位置の値を算出する式である。

【０１１７】
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）についてのイベントと、仮説ターゲットとの類似度データとしてのユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］の算出処理は以下の処理となる。
入力イベント情報中の、ユーザ確信度情報（ｕＩＤ）の各ユーザ１～ｋの確信度の値（スコア）をＰｅ［ｉ］とする。なお、ｉはユーザ識別子１～ｋに対応する変数である。
パーティクルから選択された仮説ターゲットのユーザ確信度情報（ｕＩＤ）の各ユーザ１～ｋの確信度の値（スコア）をＰｔ［ｉ］として、ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］は、以下の式によって算出する。
$UL = \Sigma P_e[i] \times P_t[i]$
上記式は、２つのデータのユーザ確信度情報（ｕＩＤ）に含まれる各対応ユーザの確信度の値（スコア）の積の総和を求める式であり、この値をユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］とする。

【０１１８】
パーティクル重み［W_{pID}］は、上記の２つの尤度、すなわち、
ガウス分布間尤度［ＤＬ］と、
ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］
これら２つの尤度を利用し、重みα（α＝０～１）を用いて下式によって算出する。
パーティクル重み［W_{pID}］$= UL^{\alpha} \times DL^{1-\alpha}$
上記式により、パーティクル重み［W_{pID}］を算出する。
ただし、α＝０～１とする。
このパーティクル重み［W_{pID}］は、各パーティクルについて各々算出する。

【０１１９】
なお、パーティクル重み［W_{pID}］の算出に適用する重み［α］は、前述したイベントーターゲット間尤度［$L_{pID,tID}$］の算出処理と同様、予め固定された値としてもよいし、入力イベントに応じて値を変更する設定としてもよい。例えば入力イベントが画像である場合において、顔検出に成功し位置情報は取得できたが顔識別に失敗した場合などは、α＝０の設定として、ユーザ確信度情報（ｕＩＤ）間尤度：ＵＬ＝１としてガウス分布間尤度［ＤＬ］のみに依存してパーティクル重み［W_{pID}］を算出する構成としてもよい。また、入力イベントが音声である場合において、話者識別に成功し話者情報破取得できたが、位置情報の取得に失敗した場合などは、α＝０の設定として、ガウス分布間尤度［ＤＬ］＝１として、ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］のみに依存してパーティクル重み［W_{pID}］を算出する構成としてもよい。

【０１２０】
図７のフローにおけるステップＳ１０３の各パーティクル対応の重み［W_{pID}］の算出は、このように図９、図１０を参照して説明した処理として実行される。次に、ステップＳ１０４において、ステップＳ１０３で設定した各パーティクルのパーティクル重み［W_{pID}］に基づくパーティクルのリサンプリング処理を実行する。

【０１２１】
このパーティクルリサンプリング処理は、ｍ個のパーティクルから、パーティクル重み

［W_{pID}］に応じてパーティクルを取捨選択する処理として実行される。具体的には、例えば、パーティクル数：ｍ＝５のとき、
パーティクル１：パーティクル重み［W_{pID}］＝０．４０
パーティクル２：パーティクル重み［W_{pID}］＝０．１０
パーティクル３：パーティクル重み［W_{pID}］＝０．２５
パーティクル４：パーティクル重み［W_{pID}］＝０．０５
パーティクル５：パーティクル重み［W_{pID}］＝０．２０
これらのパーティクル重みが各々設定されていた場合、
パーティクル１は、４０％の確率でリサンプリングされ、パーティクル２は１０％の確率でリサンプリングされる。なお、実際にはｍ＝１００～１０００といった多数であり、リサンプリングされた結果は、パーティクルの重みに応じた配分比率のパーティクルによって構成されることになる。

【０１２２】
この処理によって、パーティクル重み［W_{pID}］の大きなパーティクルがより多く残存することになる。なお、リサンプリング後もパーティクルの総数［ｍ］は変更されない。また、リサンプリング後は、各パーティクルの重み［W_{pID}］はリセットされ、新たなイベントの入力に応じてステップＳ１０１から処理が繰り返される。

【０１２３】
ステップＳ１０５では、各パーティクルに含まれるターゲットデータ（ユーザ位置およびユーザ確信度）の更新処理を実行する。各ターゲットは、先に図６等を参照して説明したように、
（ａ）ユーザ位置：各ターゲット各々に対応する存在位置の確率分布［ガウス分布：N（m_t，σ_t）］、
（ｂ）ユーザ確信度：各ターゲットが誰であるかを示すユーザ確信度情報（ｕＩＤ）として各ユーザ１～ｋである確立値（スコア）：Ｐｔ［ｉ］（ｉ＝１～ｋ）、すなわち、
uID_{t1}＝Ｐｔ［１］
uID_{t2}＝Ｐｔ［２］
：
uID_{tk}＝Ｐｔ［ｋ］
これらのデータによって構成される。

【０１２４】
ステップＳ１０５におけるターゲットデータの更新は、（ａ）ユーザ位置、（ｂ）ユーザ確信度の各々について実行する。まず、（ａ）ユーザ位置の更新処理について説明する。

【０１２５】
ユーザ位置の更新は、
（ａ１）全パーティクルの全ターゲットを対象とする更新処理、
（ａ２）各パーティクルに設定されたイベント発生源仮説ターゲットを対象とした更新処理、
これらの２段階の更新処理として実行する。

【０１２６】
（ａ１）全パーティクルの全ターゲットを対象とする更新処理は、イベント発生源仮説ターゲットとして選択されたターゲットおよびその他のターゲットのすべてを対象として実行する。この処理は、時間経過に伴うユーザ位置の分散が拡大するという仮定に基づいて実行され、前回の更新処理からの経過時間とイベントの位置情報によってカルマン・フィルタ（Ｋａｌｍａｎ　Ｆｉｌｔｅｒ）を用い更新される。

【０１２７】
以下、位置情報が１次元の場合の更新処理例について説明する。まず、前回の更新処理時間からの経過時間［ｄｔ］とし、全ターゲットについての、ｄｔ後のユーザ位置の予測分布を計算する。すなわち、ユーザ位置の分布情報としてのガウス分布：N（m_t，σ_t

）の期待値（平均）：［m_t］、分散［σ_t］について、以下の更新を行う。

$m_t = m_t + xc \times dt$

$\sigma_t^2 = \sigma_t^2 + \sigma c^2 \times dt$

なお、

m_t：予測期待値（ｐｒｅｄｉｃｔｅｄ　ｓｔａｔｅ）

σ_t^2：予測共分散（ｐｒｅｄｉｃｔｅｄ　ｅｓｔｉｍａｔｅ　ｃｏｖａｒｉａｎｃｅ）

ｘｃ：移動情報（ｃｏｎｔｒｏｌ　ｍｏｄｅｌ）

σc^2：ノイズ（ｐｒｏｃｅｓｓ　ｎｏｉｓｅ）

である。

なお、ユーザが移動しない条件の下で処理する場合は、ｘｃ＝０として更新処理を行うことができる。

上記の算出処理により、全ターゲットに含まれるユーザ位置情報としてのガウス分布：Ｎ（m_t，σ_t）を更新する。

【０１２８】

さらに、各パーティクルに１つ設定されているイベント発生源の仮説となったターゲットに関しては、音声イベント検出部１２２や画像イベント検出部１１２から入力するイベント情報に含まれるユーザ位置を示すガウス分布：Ｎ（m_e，σ_e）を用いた更新処理を実行する。

Ｋ：カルマンゲイン（Ｋａｌｍａｎ　Ｇａｉｎ）

m_e：入力イベント情報：Ｎ（m_e，σ_e）に含まれる観測値（Ｏｂｓｅｒｖｅｄ　ｓｔａｔｅ）

σ_e^2：入力イベント情報：Ｎ（m_e，σ_e）に含まれる観測値（Ｏｂｓｅｒｖｅｄ　ｃｏｖａｒｉａｎｃｅ）

として、以下の更新処理を行う。

$K = \sigma_t^2 / (\sigma_t^2 + \sigma_e^2)$

$m_t = m_t + K(xc - m_t)$

$\sigma_t^2 = (1 - K)\sigma_t^2$

【０１２９】

次に、ターゲットデータの更新処理として実行する（ｂ）ユーザ確信度の更新処理について説明する。ターゲットデータには上記のユーザ位置情報の他に、各ターゲットが誰であるかを示すユーザ確信度情報（ｕＩＤ）として各ユーザ１～ｋである確立値（スコア）：Ｐｔ［ｉ］（ｉ＝１～ｋ）が含まれている。ステップＳ１０５では、このユーザ確信度情報（ｕＩＤ）についても更新処理を行う。

【０１３０】

各パーティクルに含まれるターゲットのユーザ確信度情報（ｕＩＤ）：Ｐｔ［ｉ］（ｉ＝１～ｋ）についての更新は、登録ユーザ全員分の事後確率と、音声イベント検出部１２２や画像イベント検出部１１２から入力するイベント情報に含まれるユーザ確信度情報（ｕＩＤ）：Ｐｅ［ｉ］（ｉ＝１～ｋ）によって、予め設定した０～１の範囲の値を持つ更新率［β］を適用して更新する。

【０１３１】

ターゲットのユーザ確信度情報（ｕＩＤ）：Ｐｔ［ｉ］（ｉ＝１～ｋ）についての更新は、以下の式によって実行する。

$Pt[i] = (1 - \beta) \times Pt[i] + \beta * Pe[i]$

ただし、

ｉ＝１～ｋ

β：０～１

である。なお、更新率［β］は、０～１の範囲の値であり予め設定する。

【０１３２】

ステップＳ１０５では、この更新されたターゲットデータに含まれる以下のデータ、す

なわち、
（ａ）ユーザ位置：各ターゲット各々に対応する存在位置の確率分布［ガウス分布：Ｎ（m_t，σ_t）］、
（ｂ）ユーザ確信度：各ターゲットが誰であるかを示すユーザ確信度情報（ｕＩＤ）として各ユーザ１～ｋである確立値（スコア）：Ｐｔ［ｉ］（ｉ＝１～ｋ）、すなわち、
uID_{t1}＝Ｐｔ［１］
uID_{t2}＝Ｐｔ［２］
：
uID_{tk}＝Ｐｔ［ｋ］
これらのデータと、各パーティクル重み［W_{pID}］とに基づいて、ターゲット情報を生成して、処理決定部１３２に出力する。

【０１３３】
なお、ターゲット情報の生成は、図５を参照して説明したように、各パーティクル（ＰＩＤ＝１～ｍ）に含まれる各ターゲット（ｔＩＤ＝１～ｎ）対応データの重み付き総和データとして生成される。図５の右端のターゲット情報３０５に示すデータである。ターゲット情報は、各ターゲット（ｔＩＤ＝１～ｎ）各々の
（ａ）ユーザ位置情報、
（ｂ）ユーザ確信度情報、
これらの情報を含む情報として生成される。

【０１３４】
例えば、ターゲット（ｔＩＤ＝１）に対応するターゲット情報中の、ユーザ位置情報は、

【数２】

$$\sum_{i=1}^{m} W_i \cdot N(m_{i1}, \sigma_{i1})$$

【０１３５】
上記式で表される。上記式において、W_iは、パーティクル重み［W_{pID}］を示している。

【０１３６】
また、ターゲット（ｔＩＤ＝１）に対応するターゲット情報中の、ユーザ確信度情報は、

【数３】

$$\sum_{i=1}^{m} W_i \cdot uID_{i11}$$

$$\sum_{i=1}^{m} W_i \cdot uID_{i12}$$

$$\vdots$$

$$\sum_{i=1}^{m} W_i \cdot uID_{i1k}$$

【０１３７】

上記式で表される。上記式において、W_iは、パーティクル重み［W_{pID}］を示している。

音声・画像統合処理部１３１は、これらのターゲット情報をｎ個の各ターゲット（ｔＩＤ＝１～ｎ）各々について算出し、算出したターゲット情報を処理決定部１３２に出力する。

【０１３８】

次に、図７に示すフローのステップＳ１０６の処理について説明する。音声・画像統合処理部１３１は、ステップＳ１０６において、ｎ個のターゲット（ｔＩＤ＝１～ｎ）の各々がイベントの発生源である確率を算出し、これをシグナル情報として処理決定部１３２に出力する。

【０１３９】

先に説明したように、イベント発生源を示す［シグナル情報］は、音声イベントについては、誰が話をしたか、すなわち［話者］を示すデータであり、画像イベントについては、画像に含まれる顔が誰であるかを示すデータである。

【０１４０】

音声・画像統合処理部１３１は、各パーティクルに設定されたイベント発生源の仮説ターゲットの数に基づいて、各ターゲットがイベント発生源である確率を算出する。すなわち、ターゲット（ｔＩＤ＝１～ｎ）の各々がイベント発生源である確率を［Ｐ（ｔＩＤ＝ｉ）とする。ただしｉ＝１～ｎである。このとき、各ターゲットがイベント発生源である確率は、以下のように算出される。

Ｐ（ｔＩＤ＝１）：ｔＩＤ＝１を割り当てた数／ｍ
Ｐ（ｔＩＤ＝２）：ｔＩＤ＝２を割り当てた数／ｍ
：
Ｐ（ｔＩＤ＝ｎ）：ｔＩＤ＝ｎを割り当てた数／ｍ

音声・画像統合処理部１３１は、この算出処理によって、生成した情報、すなわち、各ターゲットがイベント発生源である確率を［シグナル情報］として、処理決定部１３２に出力する。

【０１４１】

ステップＳ１０６の処理が終了したら、ステップＳ１０１に戻り、音声イベント検出部１２２および画像イベント検出部１１２からのイベント情報の入力の待機状態に移行する。

【０１４２】

以上が、図７に示すフローのステップＳ１０１～Ｓ１０６の説明である。ステップＳ１０１において、音声・画像統合処理部１３１が、音声イベント検出部１２２および画像イベント検出部１１２から、図３（Ｂ）に示すイベント情報を取得できなかった場合も、ステップＳ１２１において、各パーティクルに含まれるターゲットの構成データの更新が実行される。この更新は、時間経過に伴うユーザ位置の変化を考慮した処理である。

【０１４３】

このターゲット更新処理は、先に、ステップＳ１０５の説明において（ａ１）全パーティクルの全ターゲットを対象とする更新処理と同様の処理であり、時間経過に伴うユーザ位置の分散が拡大するという仮定に基づいて実行され、前回の更新処理からの経過時間とイベントの位置情報によってカルマン・フィルタ（Ｋａｌｍａｎ　Ｆｉｌｔｅｒ）を用い更新される。

【０１４４】

位置情報が１次元の場合の更新処理例について説明する。まず、前回の更新処理時間からの経過時間［ｄｔ］とし、全ターゲットについての、ｄｔ後のユーザ位置の予測分布を計算する。すなわち、ユーザ位置の分布情報としてのガウス分布：Ｎ（m_t，σ_t）の期待値（平均）：［m_t］、分散［σ_t］について、以下の更新を行う。

$m_t = m_t + xc \times dt$

$\sigma_t^2 = \sigma_t^2 + \sigma c^2 \times dt$

なお、

m_t：予測期待値（ｐｒｅｄｉｃｔｅｄ　ｓｔａｔｅ）

σ_t^2：予測共分散（ｐｒｅｄｉｃｔｅｄ　ｅｓｔｉｍａｔｅ　ｃｏｖａｒｉａｎｃｅ）

ｘｃ：移動情報（ｃｏｎｔｒｏｌ　ｍｏｄｅｌ）

σc^2：ノイズ（ｐｒｏｃｅｓｓ　ｎｏｉｓｅ）

である。

なお、ユーザが移動しない条件の下で処理する場合は、ｘｃ＝０として更新処理を行うことができる。

上記の算出処理により、全ターゲットに含まれるユーザ位置情報としてのガウス分布：Ｎ（m_t，σ_t）を更新する。

【０１４５】

なお、各パーティクルのターゲットに含まれるユーザ確信度情報（ｕＩＤ）については、イベントの登録ユーザ全員分の事後確率、もしくはイベント情報からスコア［Ｐｅ］が取得できない限りは更新しない。

【０１４６】

ステップＳ１２１の処理が終了したら、ステップＳ１０１に戻り、音声イベント検出部１２２および画像イベント検出部１１２からのイベント情報の入力の待機状態に移行する。

【０１４７】

以上、図７を参照して音声・画像統合処理部１３１の実行する処理について説明した。音声・画像統合処理部１３１は、図７に示すフローに従った処理を音声イベント検出部１２２および画像イベント検出部１１２からのイベント情報の入力ごとに繰り返し実行する。この繰り返し処理により、より信頼度の高いターゲットを仮説ターゲットとして設定したパーティクルの重みが大きくなり、パーティクル重みに基づくリサンプリング処理により、より重みの大きいパーティクルが残存することになる。結果として音声イベント検出部１２２および画像イベント検出部１１２から入力するイベント情報に類似する信頼度の高いデータが残存することになり、最終的に信頼度の高い以下の各情報、すなわち、

（ａ）複数のユーザが、それぞれどこにいて、それらは誰であるかの推定情報としての［ターゲット情報］、

（ｂ）例えば話をしたユーザなどのイベント発生源を示す［シグナル情報］、

これらが生成されて処理決定部１３２に出力される。

【０１４８】
［ターゲットの生成および削除］
上述した実施例において、音声・画像統合処理部１３１では、予めｍ個のパーティクルにそれぞれｎ個のターゲットを設定して処理を行う構成を説明したが、ターゲットの数は、適宜変更する設定としてよい、すなわち、必要に応じて、新たなターゲットの生成や、ターゲットの削除を行う構成としてもよい。
【０１４９】
（ターゲットの生成）
まず、音声・画像統合処理部１３１における新たなターゲットの生成処理について、図１１を参照して説明する。新たなターゲットの生成は、例えば各パーティクルに対するイベント発生源仮説の設定時に行う。
【０１５０】
イベントと既存のｎ個の各ターゲットとのイベントーターゲット間尤度を計算する際、暫定的にｎ＋１番目のターゲットとして図１１に示すような「位置情報」、「識別情報」に一様分布（「分散が十分大きいガウス分布」と「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」）に設定した新たな暫定新規ターゲット５０１を生成する。
【０１５１】
この暫定的な新規ターゲット（ｔＩＤ＝ｎ＋１）を設定した後、新たなイベントの入力に基づいて、図７を参照して説明したフローにおけるステップＳ１０２のイベント発生源仮説の設定が行われ、この処理の際に、入力イベント情報と各ターゲット間の尤度算出が実行されて、各ターゲットのターゲット重み［W_{tID}］の算出が行われる。このとき、図１１に示す暫定ターゲット（ｔＩＤ＝ｎ＋１）についても、入力イベント情報との尤度算出を実行して、暫定的なｎ＋１番目のターゲットのターゲット重み（W_{n+1}）を算出する。
【０１５２】
この暫定的なｎ＋１番目のターゲットのターゲット重み（W_{n+1}）が、既存のｎ個のターゲットのターゲット重み（W_1～W_n）より大きいと判断された場合は、その新規ターゲットを全パーティクルに対して設定する。
【０１５３】
なお、例えばカメラの撮影する１つの画像中に複数の顔イベントがあり、１つ１つの顔イベントに対して、図７に示すフローの処理を行う構成において、１画像中の顔の数（＝イベント数）が、各パーティクルに設定されたターゲット数（ｎ）より少ない場合、ｔＩＤ＝ｎ＋１の暫定ターゲットの重みＷｎ＋１が、他のターゲットの重み（Ｗ１～Ｗｎ）より大きくなくても、そのまま新規ターゲットとして全パーティクルに対して生成する処理を行う構成としてしても良い。
【０１５４】
なお、新規ターゲットが生成された場合、イベント発生源の仮説の生成は事前に計算したターゲット重み［W_{tID}］に基づいて確率的に行っても良いし、全てのパーティクルにおいてイベント発生源の仮説を新規ターゲットにしても良い。
【０１５５】
（ターゲットの削除）
次に、音声・画像統合処理部１３１におけるターゲットの削除処理について、図１２を参照して説明する。ターゲットの削除は、例えば図７に示す処理フローにおけるステップＳ１０５のターゲットデータの更新処理に際して実行する。
【０１５６】
ステップＳ１０５では、先に説明したように、ターゲットデータの更新を実行して更新されたターゲットデータと、各パーティクル重み［W_{pID}］とに基づいて、ターゲット情報を生成して、処理決定部１３２に出力する処理が行われる。例えば図１２に示すターゲット情報５２０が生成される。ターゲット情報は、各ターゲット（ｔＩＤ＝１～ｎ）各々の

（ａ）ユーザ位置情報、
（ｂ）ユーザ確信度情報、
これらの情報を含む情報として生成される。

【０１５７】

音声・画像統合処理部１３１は、このように更新ターゲットに基づいてして生成したターゲット情報中のユーザ位置情報に着目する。ユーザ位置情報は、ガウス分布Ｎ（ｍ，σ）として設定される。このガウス分布に一定のピークが検出されない場合は、特定のユーザの位置を示す有効な情報とはならない。音声・画像統合処理部１３１は、このようなピークを持たない分布データとなるターゲットを削除対象として選択する。

【０１５８】

例えば、図１２に示すターゲット情報５２０には、ターゲット１，２，ｎの３つのターゲット情報５２１，５２２，５２３を示しているが、これらのターゲット情報中のユーザ位置を示すガウス分布データのピークと予め定めた閾値５３１との比較を実行し、閾値５３１以上のピークを持たないデータ、すなわち、図１２の例では、ターゲット情報５２３を削除ターゲットとする。

【０１５９】

この例ではターゲット（ｔＩＤ＝ｎ）が削除ターゲットとして選択され。すべてのパーティクルから削除される。このようにユーザ位置を示すガウス分布（確率密度分布）の最大値が、削除の閾値よりも小さいときに、全パーティクルに対してそのターゲットを削除する。なお、適用する閾値は、固定値でも良いし、インタラクション対象ターゲットに関しては閾値を下げて削除されにくくするなど、ターゲット毎に変える構成としてもよい。

【０１６０】

［画像フレーム外に仮想ターゲットを生成する処理例］

上述した［ターゲットの生成および削除］の説明では、新たなターゲットの生成および削除構成について説明したが、図２に示す画像イベント検出部１１２において、画像入力部（カメラ）１１１から入力する画像情報、すなわちカメラの撮影している画像フレームの外にユーザが存在する場合は、そのユーザに対する画像イベントを取得できないため、画像イベントからそのターゲットを生成することはできないという問題がある。

【０１６１】

そのような状態でそのユーザが音声イベントを発生しても、そのユーザ対応のターゲットが生成されず、カメラフレーム内の他のターゲットから音声イベントが発生したと推定してしまい、この場合、誤った推定結果を生成することになる。

【０１６２】

すなわち、暫定的にｎ＋１番目のターゲットとして図１１に示すような「位置情報」と、「識別情報」として一様分布（「分散が十分大きいガウス分布」と「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」）のデータを設定した新たな暫定新規ターゲット５０１を生成し、この暫定的な新規ターゲット（ｔＩＤ＝ｎ＋１）を設定した後、新たなイベントの入力に基づいて、図７を参照して説明したフローにおけるステップＳ１０２のイベント発生源仮説の設定が行われ、この処理の際に、入力イベント情報と各ターゲット間の尤度算出を実行して、各ターゲットのターゲット重み［W_{tID}］の算出を行う。

【０１６３】

このとき、図１１に示す暫定ターゲット（ｔＩＤ＝ｎ＋１）についても、入力イベント情報との尤度算出を実行して、暫定的なｎ＋１番目のターゲットのターゲット重み（W_{n+1}）を算出する。この暫定的なｎ＋１番目のターゲットのターゲット重み（W_{n+1}）が、既存のｎ個のターゲットのターゲット重み（W_1～W_n）より大きいと判断された場合は、その新規ターゲットを全パーティクルに対して設定する構成である。

【０１６４】

しかし、この方法を適用した場合、カメラフレーム外からの音声イベントのようにその位置情報の平均値と既に存在するターゲットの位置情報の平均値がある程度離れていても、位置情報の分散が大きい場合はガウス分布間尤度がそれほど小さくならない傾向がある

。

【０１６５】

その結果、システムがターゲットとして認識していないユーザ、すなわち、図２に示す画像入力部（カメラ）１１１から入力する画像フレームの外のユーザからの音声イベントであっても、「イベント」と「一様分布のターゲット（ｎ＋１）」間の尤度が最大にならずターゲットを生成することができないため、既に存在するターゲットのみで音声イベント発生源である確率の計算を行ってしまうことがある。

【０１６６】

そこで、各パーティクルでのイベント発生源の仮説生成においてターゲットの生成を確認する際、画像フレーム外に仮想ターゲットを生成する。以下、この処理例について説明する。

【０１６７】

本処理例では、バックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ　Ｍｏｄｅｌ）として画像入力部（カメラ）１１１から入力する画像フレーム外に仮想のターゲットを配置し、「イベント」と「既に存在するターゲット（１～ｎ）と一様分布のターゲット」間の尤度計算に加え、画像フレーム外に仮想のターゲットを配置したバックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ　Ｍｏｄｅｌ）の仮想ターゲットとも尤度計算を行う。なお、ユーザＩＤ間尤度の計算においては、一様分布のターゲットと同様、図１１に示す「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」を持つ一様のデータを用いる。

【０１６８】

新たなターゲットの生成は、例えば各パーティクルに対するイベント発生源仮説の設定時に行う。イベントと既存のｎ個の各ターゲットとのイベントーターゲット間尤度を計算する際、暫定的にｎ＋１番目のターゲットとして、画像フレーム外に仮想のターゲットを配置したバックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ　Ｍｏｄｅｌ）の暫定的な仮想ターゲット（ｔＩＤ＝ｎ＋１）を生成する。

【０１６９】

この暫定的な新規ターゲット（ｔＩＤ＝ｎ＋１）を設定した後、新たなイベントの入力に基づいて、図７を参照して説明したフローにおけるステップＳ１０２のイベント発生源仮説の設定を行う。

【０１７０】

すなわち、入力イベント情報と各ターゲット間の尤度算出を実行して各ターゲットのターゲット重み［W_{tID}］の算出を行う際に、バックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ　Ｍｏｄｅｌ）の暫定的な仮想ターゲット（ｔＩＤ＝ｎ＋１）についても、入力イベント情報との尤度算出を実行して、暫定的なｎ＋１番目のターゲットのターゲット重み（W_{n+1}）を算出する。

【０１７１】

この暫定的なｎ＋１番目のターゲットのターゲット重み（W_{n+1}）が、既存のｎ個のターゲットのターゲット重み（W_1～W_n）より大きいと判断された場合は、その新規ターゲットを全パーティクルに対して設定する。

【０１７２】

図１３に、画像フレーム外に仮想のターゲットを配置したバックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ　Ｍｏｄｅｌ）を含めたイベントーターゲット間尤度の計算例を示す。

【０１７３】

図１３（ａ）はイベント検出を行う実環境を示している。画像入力部（カメラ）１１１から入力する画像情報、すなわちカメラの撮影している画像フレーム６０１の外に声を発したユーザ６１１が存在する。

【０１７４】

図１３（ｂ）は、図２に示す音声イベント検出部１２２において検出された音声イベント情報を示している。音声イベント検出部１２２は、複数の異なるポジションに配置された複数の音声入力部（マイク）１２１ａ～ｄから入力する音声情報を解析し、音声の発生

源の位置情報を確率分布データとして生成する。具体的には、音源方向に関する期待値と分散データＮ（m_e，σ_e）を生成する。また、予め登録されたユーザの声の特徴情報との比較処理に基づいてユーザ識別情報を生成する。この識別情報も確率的な推定値として生成する。音声イベント検出部１２２には、予め検証すべき複数のユーザの声についての特徴情報が登録されており、入力音声と登録音声との比較処理を実行して、どのユーザの声である確率が高いかを判定する処理を行い、全登録ユーザに対する事後確率、あるいはスコアを算出する。

【０１７５】

図１３（ｃ）は、音声画像統合処理部１３１が保持する既存のｎ個のターゲット（ｔＩＤ＝１～ｎ）と、暫定的にｎ＋１番目のターゲットとして生成した（Ｘ），（Ｙ），（Ｚ）の３つのターゲットを示している。

【０１７６】

ターゲット（Ｘ）は、先に図１１を参照して説明した暫定的な新規ターゲット（ｔＩＤ＝ｎ＋１）であり、「位置情報」、「識別情報」に一様分布（「分散が十分大きいガウス分布」と「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」）に設定した新たな暫定新規ターゲットである。

【０１７７】

ターゲット（Ｙ），（Ｚ）は、上述したバックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ Ｍｏｄｅｌ）のターゲットであり、画像フレーム外に仮想のターゲットを配置した新規ターゲット（ＩＤ＝ｎ＋１）である。ターゲット（Ｙ）は、「位置情報」が、画像フレームの外の左側の位置に高い存在確率を持つ情報であり、「識別情報」は、（Ｘ）の一様分布ターゲットと同様、「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」を持つ一様のデータである。

【０１７８】

ターゲット（Ｚ）は、「位置情報」が、画像フレームの外の右側の位置に高い存在確率を持つ情報であり、「識別情報」は、（Ｘ）の一様分布ターゲットと同様、「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」を持つ一様のデータである。

【０１７９】

これらの暫定的な新規ターゲット（ｔＩＤ＝ｎ＋１）を設定した後、新たなイベントの入力に基づいて、図７を参照して説明したフローにおけるステップＳ１０２のイベント発生源仮説の設定が行われ、この処理の際に、入力イベント情報と各ターゲット間の尤度算出が実行されて、各ターゲットのターゲット重み［W_{tID}］の算出が行われる。このとき、図１３に示す３つの暫定ターゲット（Ｘ），（Ｙ），（Ｚ）についても、入力イベント情報との尤度算出を実行して、暫定的なｎ＋１番目のターゲットとしてのターゲット重み（W_{n+1}）を算出する。

【０１８０】

この暫定的なｎ＋１番目のターゲット（Ｘ），（Ｙ），（Ｚ）のいずれかのターゲット重み（W_{n+1}）が、既存のｎ個のターゲットのターゲット重み（W_1～W_n）より大きいと判断された場合は、その新規ターゲットを全パーティクルに対して設定する。

【０１８１】

ターゲット重みの算出例を図１４に示す。ターゲット重みは、図１４の右端に示すように、各パーティクルに設定されるターゲット１～ｎの各々に対応するｎ個のターゲット重みの算出処理として実行される。このｎ個のターゲット重みの算出処理に際しては、先に図８を参照して説明したように、まず、入力イベント情報、すなわち、音声・画像統合処理部１３１が、音声イベント検出部１２２および画像イベント検出部１１２から入力したイベント情報と、各パーティクルの各ターゲットデータとの類似度の指標値としての尤度算出を行う。

【０１８２】

図８を参照して説明したように、尤度算出処理は、

（ａ）ユーザ位置情報についてのイベントと、ターゲットデータとの類似度データとし

てのガウス分布間尤度［ＤＬ］、
（ｂ）ユーザ識別情報（顔識別情報または話者識別情報）についてのイベントと、ターゲットデータとの類似度データとしてのユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］
これらを個別に算出する。

【０１８３】

次に、入力イベント情報とあるパーティクル（ｐＩＤ）が持つ１つのターゲット（ｔＩＤ）との類似度の指標としてのイベント－ターゲット間尤度［$L_{pID,tID}$］は、上記の２つの尤度、すなわち、
ガウス分布間尤度［ＤＬ］と、
ユーザ確信度情報（ｕＩＤ）間尤度［ＵＬ］
これら２つの尤度を利用して算出する。すなわち重みα（α＝０～１）を用いて、イベント－ターゲット間尤度［$L_{pID,tID}$］は下式によって算出する。

$$[L_{pID,tID}] = UL^{\alpha} \times DL^{1-\alpha}$$

としてイベントとターゲットとの類似度の指標であるイベント－ターゲット間尤度［$L_{pID,tID}$］を算出する。
ただし、α＝０～１とする。

【０１８４】

このイベント－ターゲット間尤度［$L_{pID,tID}$］を、各パーティクルの各ターゲットについて各々算出し、このイベント－ターゲット間尤度［$L_{pID,tID}$］に基づいて各ターゲットのターゲット重み［W_{tID}］を算出する。

【０１８５】

イベント－ターゲット間尤度［$L_{pID,tID}$］に基づく、ターゲット重み［W_{tID}］の算出式は、先に説明した通り、以下の算出式である。

【数４】

$$W_{tID} = \sum_{pID}^{m} W_{pID} L_{pID,tID}$$

【０１８６】

図１４に示すターゲット重みの算出例において、上段に記載のW_1～W_nは、すでに設定済みのターゲットについて算出したイベント－ターゲット間尤度である。下段の（Ｘ），（Ｙ），（Ｚ）として示す３つのW_{n+1}は、図１３を参照して説明した暫定的な新規ターゲット（ｔＩＤ＝ｎ＋１）に対応するイベント－ターゲット間尤度である。

【０１８７】

すなわち、（Ｘ）は、「位置情報」、「識別情報」に一様分布（「分散が十分大きいガウス分布」と「全Ｐｔ［ｉ］が等しいＵｓｅｒＩＤ分布」）に設定した新たな暫定新規ターゲット、（Ｙ），（Ｚ）は、上述したバックグランドモデル（Ｂａｃｋｇｒｏｕｎｄ Ｍｏｄｅｌ）のターゲットであり、画像フレーム外に仮想のターゲットを配置した新規ターゲット（ＩＤ＝ｎ＋１）であり、これらに対応するイベント－ターゲット間尤度も算出する。

【０１８８】

この暫定的なｎ＋１番目のターゲット（Ｘ），（Ｙ），（Ｚ）のいずれかのターゲット重み（W_{n+1}）が、既存のｎ個のターゲットのターゲット重み（W_1～W_n）より大きいと判断された場合は、その新規ターゲットを全パーティクルに対して設定する。

【０１８９】

なお、ターゲット生成確認時に用いたカメラフレーム外に仮想のターゲットは、他の処理では用いない。この処理例に従えば、カメラによって撮影された画像フレーム外のユーザからの音声イベントに対して、各ターゲットがイベント発生源である確率推定の性能が

向上する。

【０１９０】

このように、本処理例では、図１に示す情報処理装置１００の音声・画像統合処理部１３１がイベント検出部１１２，１２２の生成するイベント情報を入力し、仮想的なユーザに対応する複数のターゲットを設定した複数のパーティクルを適用したパーティクルフィルタリング処理を実行して実空間に存在するユーザのユーザ位置情報およびユーザ識別情報を含む解析情報を生成する構成を有し、カメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとイベント検出部１１２，１２２の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベント－ターゲット間尤度より大きい値である場合に、暫定ターゲットを各パーティクルに新規追加する処理を行う。

【０１９１】

また、音声・画像統合処理部１３１は、暫定ターゲットとして、図１３、図１４を参照して説明したように、

（Ｘ）均一データによって構成されるユーザ位置情報、ユーザ識別情報を持つ暫定ターゲット

（Ｙ），（Ｚ）画像フレームの異なる方向のフレーム外部位置に仮想ターゲットを設定した複数の異なる暫定ターゲット

これらの異なるタイプの暫定ターゲットを生成し、生成した複数の暫定ターゲットとイベント情報との尤度を個別に算出して、算出した暫定ターゲットのイベント－ターゲット間尤度の最大値が、既存ターゲットに対応するイベント－ターゲット間尤度より大きい値を有する場合に、その最大値に対応する暫定ターゲットを各パーティクルに新規追加する処理を行う。本構成により、カメラの取得する画像フレームの外部にいるユーザからの音声入力イベントに対応した正しい推定処理が可能となり、ユーザ位置やユーザ識別情報を効率的に確実に生成することが可能となる。

【０１９２】

以上、特定の実施例を参照しながら、本発明について詳解してきた。しかしながら、本発明の要旨を逸脱しない範囲で当業者が実施例の修正や代用を成し得ることは自明である。すなわち、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本発明の要旨を判断するためには、特許請求の範囲の欄を参酌すべきである。

【０１９３】

また、明細書中において説明した一連の処理はハードウェア、またはソフトウェア、あるいは両者の複合構成によって実行することが可能である。ソフトウェアによる処理を実行する場合は、処理シーケンスを記録したプログラムを、専用のハードウェアに組み込まれたコンピュータ内のメモリにインストールして実行させるか、あるいは、各種処理が実行可能な汎用コンピュータにプログラムをインストールして実行させることが可能である。例えば、プログラムは記録媒体に予め記録しておくことができる。記録媒体からコンピュータにインストールする他、ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、インターネットといったネットワークを介してプログラムを受信し、内蔵するハードディスク等の記録媒体にインストールすることができる。

【０１９４】

なお、明細書に記載された各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。また、本明細書においてシステムとは、複数の装置の論理的集合構成であり、各構成の装置が同一筐体内にあるものには限らない。

【産業上の利用可能性】

【０１９５】

以上、説明したように、本発明の一実施例の構成によれば、カメラやマイクによって取得される画像情報や音声情報に基づいてユーザの推定位置および推定識別データを含むイベント情報を入力して、複数のターゲットを設定した複数のパーティクルを適用したパー

ティクルフィルタリング処理を行い、フィルタリングによる仮説の更新および取捨選択に基づいてユーザの位置および識別情報を生成する。また、カメラの取得する画像フレームの外部に仮想ターゲットを設定した暫定ターゲットとイベント検出部の生成するイベント情報との尤度が、画像フレームの内部にターゲットを設定した既存ターゲットに対応するイベントーターゲット間尤度より大きい値である場合に、暫定ターゲットを各パーティクルに新規追加する処理を行う。本構成により、カメラの取得する画像フレームの外部にいるユーザからの音声入力イベントに対応した正しい推定処理が可能となり、ユーザ位置やユーザ識別情報を効率的に確実に生成することが可能となる。

【図面の簡単な説明】

【０１９６】

【図１】本発明に係る情報処理装置の実行する処理の概要について説明する図である。

【図２】本発明の一実施例の情報処理装置の構成および処理について説明する図である。

【図３】音声イベント検出部１２２および画像イベント検出部１１２が生成し音声・画像統合処理部１３１に入力する情報の例について説明する図である。

【図４】パーティクル・フィルタ（Ｐａｒｔｉｃｌｅ　Ｆｉｌｔｅｒ）を適用した基本的な処理例について説明する図である。

【図５】本処理例で設定するパーティクルの構成について説明する図である。

【図６】各パーティクルに含まれるターゲット各々が有するターゲットデータの構成について説明する図である。

【図７】音声・画像統合処理部１３１の実行する処理シーケンスを説明するフローチャートを示す図である。

【図８】ターゲット重み［W_{tID}］の算出処理の詳細について説明する図である。

【図９】パーティクル重み［W_{pID}］の算出処理の詳細について説明する図である。

【図１０】パーティクル重み［W_{pID}］の算出処理の詳細について説明する図である。

【図１１】音声・画像統合処理部１３１における新たなターゲットの生成処理について説明する図である。

【図１２】音声・画像統合処理部１３１におけるターゲットの削除処理について説明する図である。

【図１３】画像フレーム外に仮想ターゲットを生成する処理例について説明する図である。

【図１４】画像フレーム外に仮想ターゲットを生成する処理におけるイベントーターゲット間尤度の算出処理例について説明する図である。

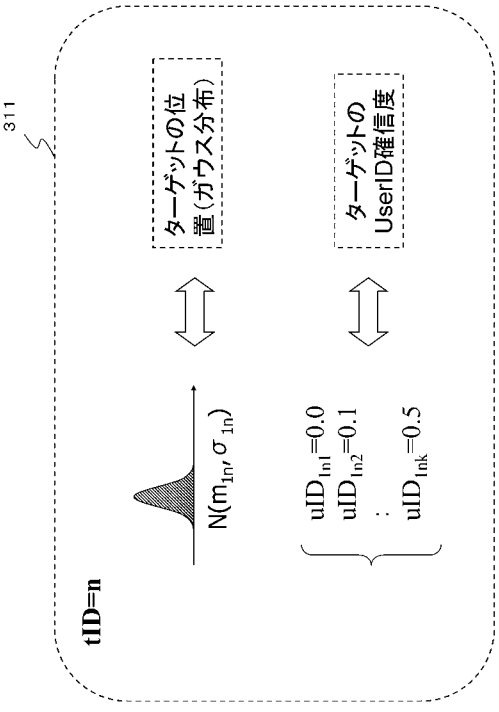
【符号の説明】

【０１９７】

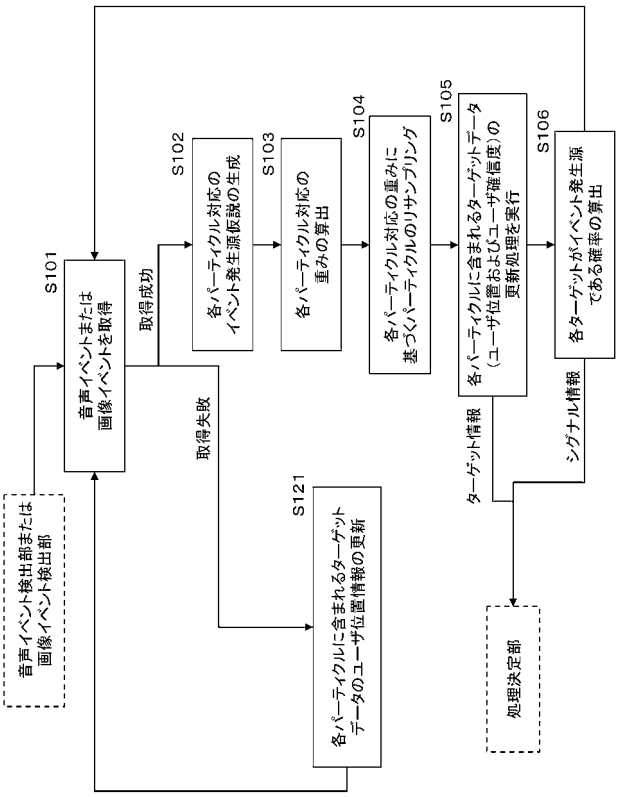
１１～１４　ユーザ
２１　カメラ
３１～３４　マイク
１００　情報処理装置
１１１　画像入力部
１１２　画像イベント検出部
１２１　音声入力部
１２２　音声イベント検出部
１３１　音声・画像統合処理部
１３２　処理決定部
２０１～２０ｋ　ユーザ
３０１　ユーザ
３０２　画像データ
３０５　ターゲット情報
３１１　ターゲットデータ
４０１　イベント情報

４１１～４１３　パーティクル
４２１～４２３　ターゲット
５０１　暫定新規ターゲット
５２０　ターゲット情報
５２１～５２３　ターゲット情報
５３１　閾値
６０１　画像フレーム
６１１　ユーザ

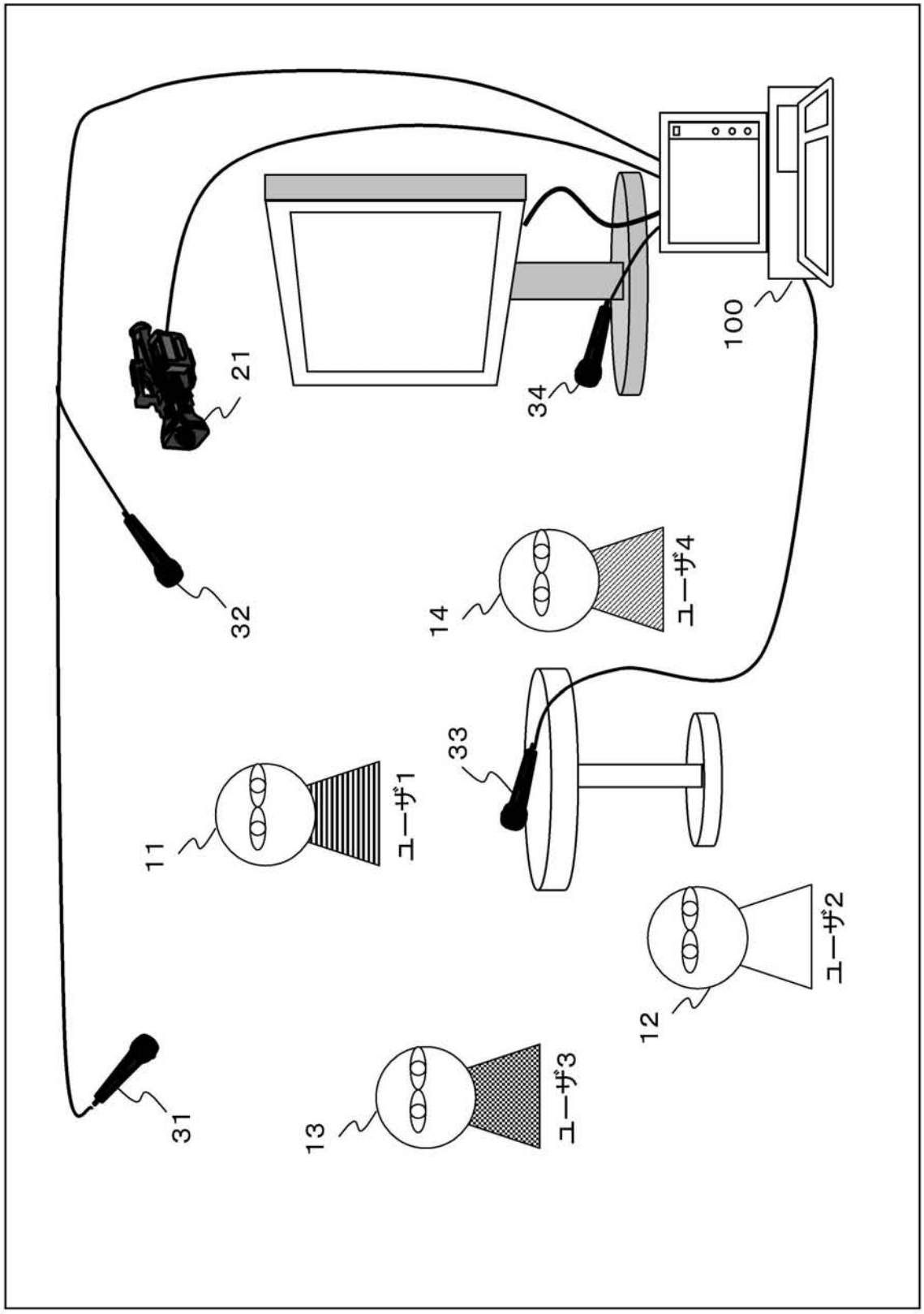
【図６】



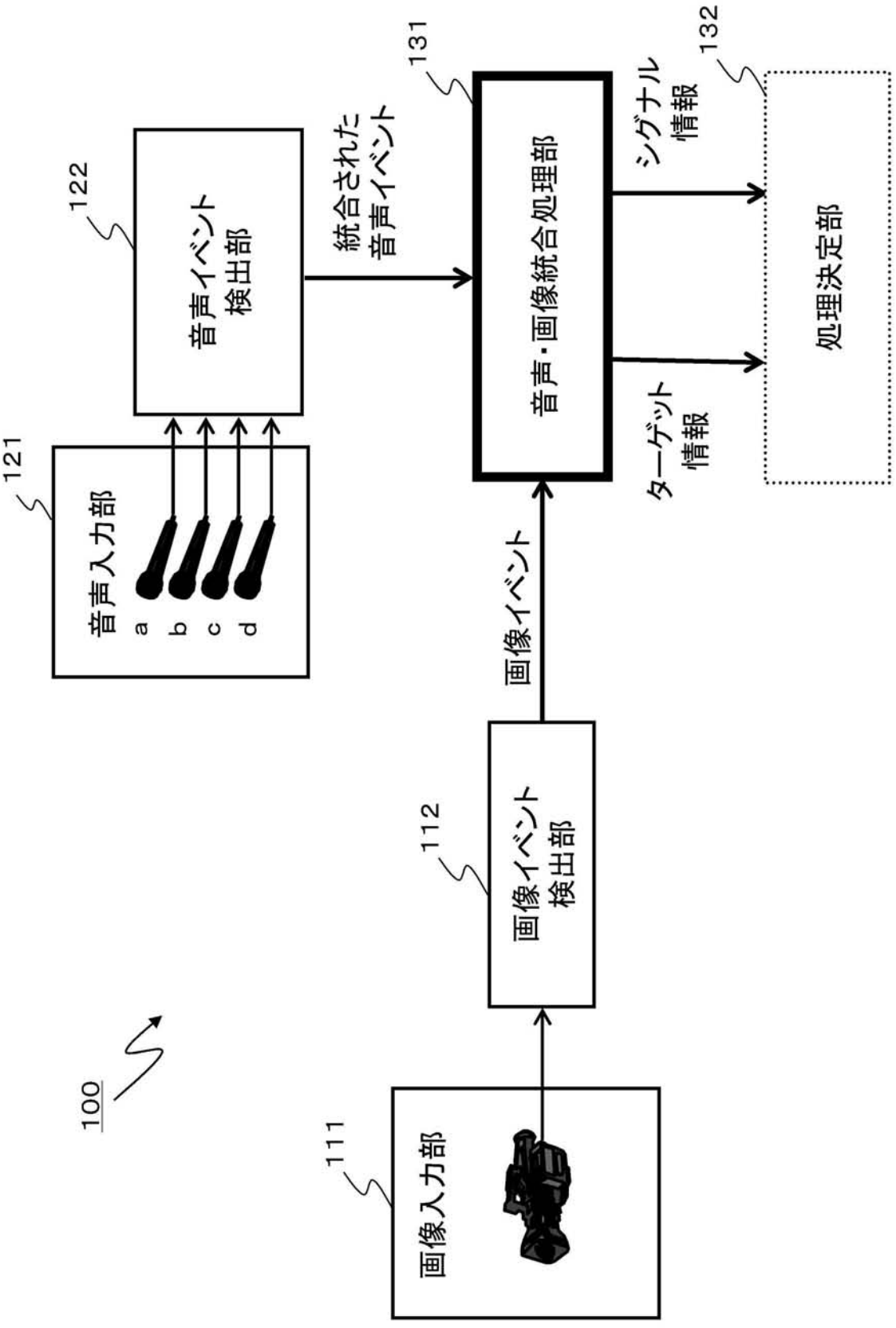
【図７】



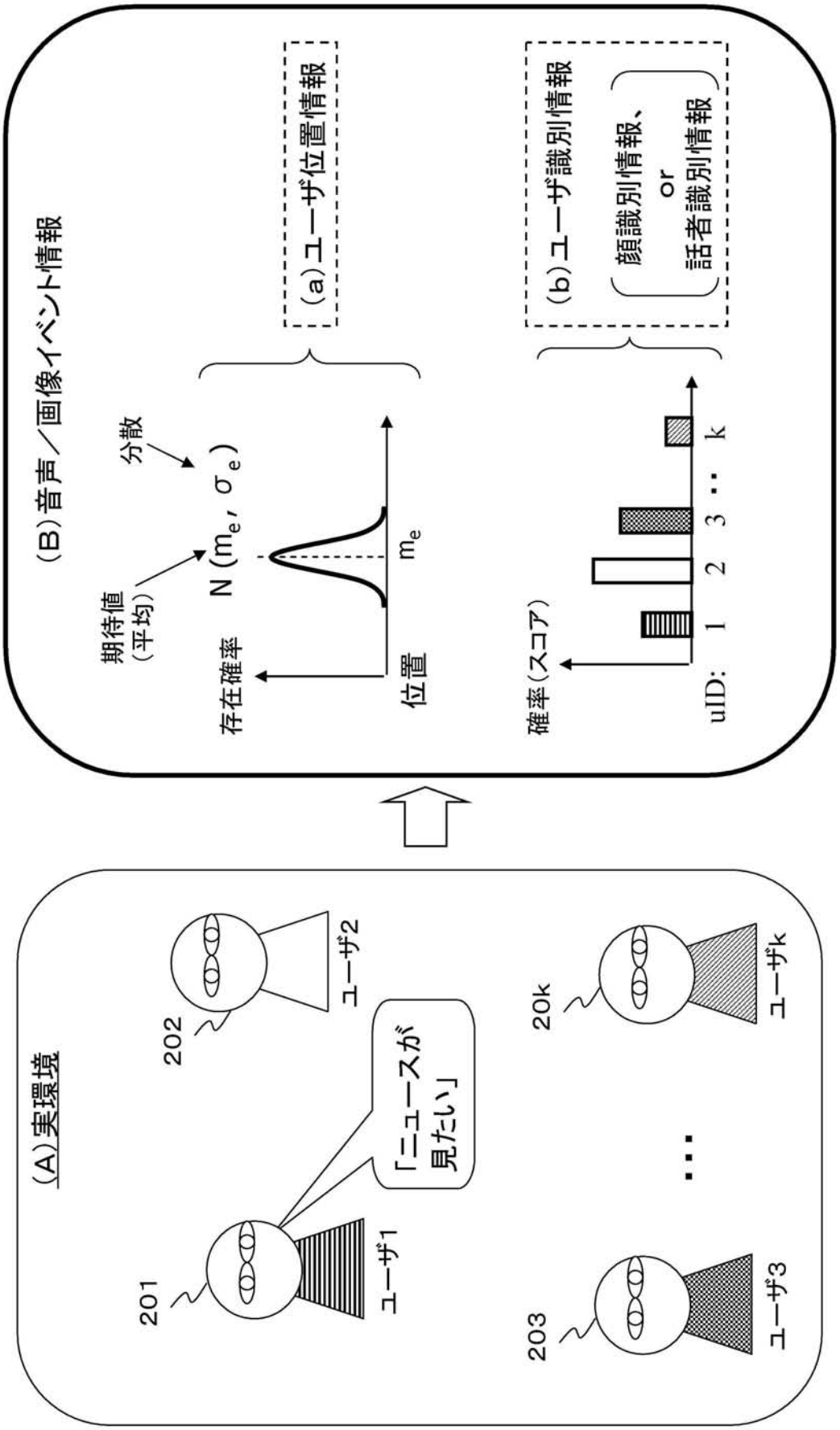
【図１】



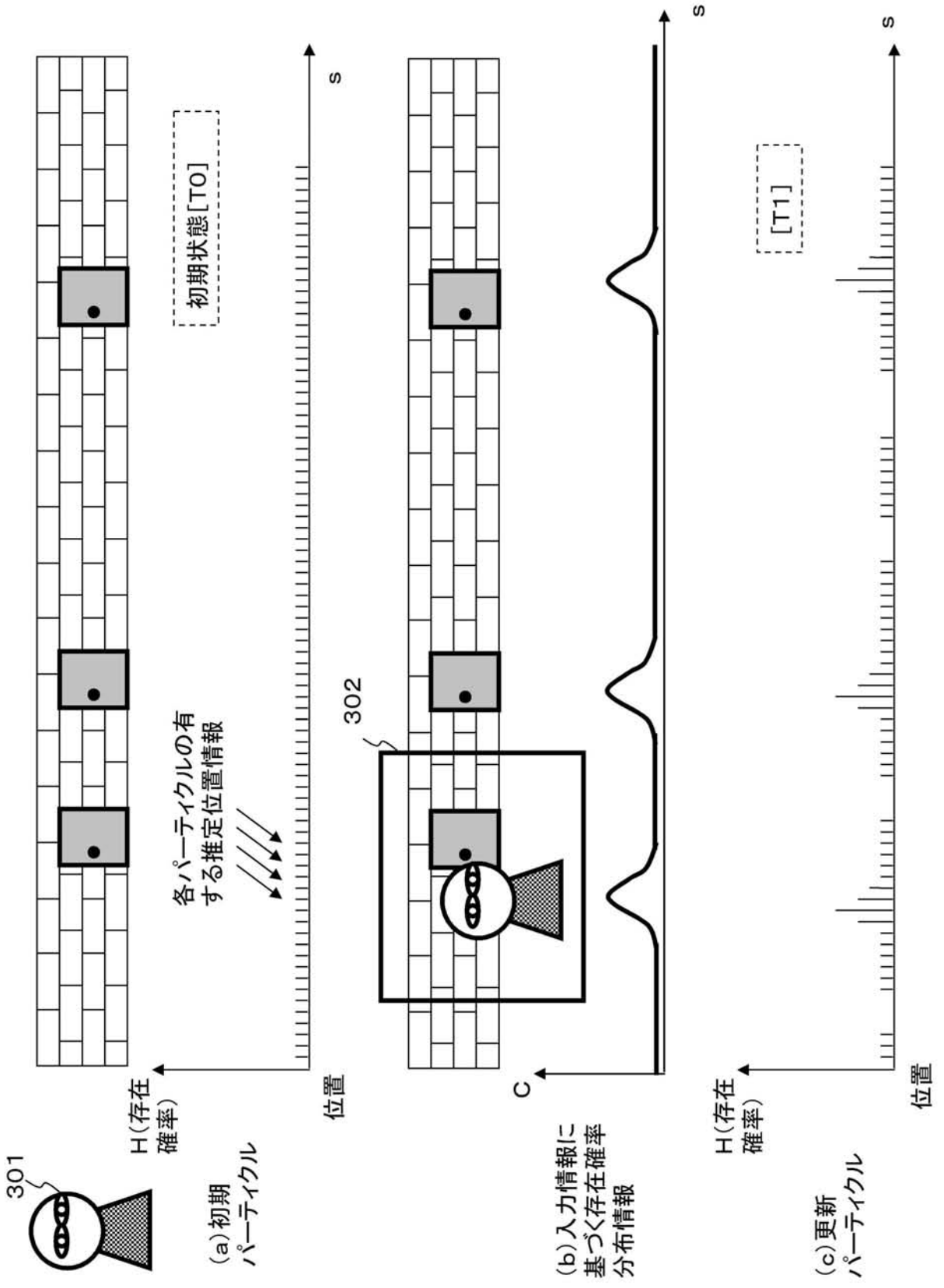
【図２】


100
111
画像入力部
112
画像イベント
検出部
画像イベント
121
音声入力部
a
b
c
d
122
音声イベント
検出部
統合された
音声イベント
131
音声・画像統合処理部
シグナル
情報
ターゲット
情報
132
処理決定部

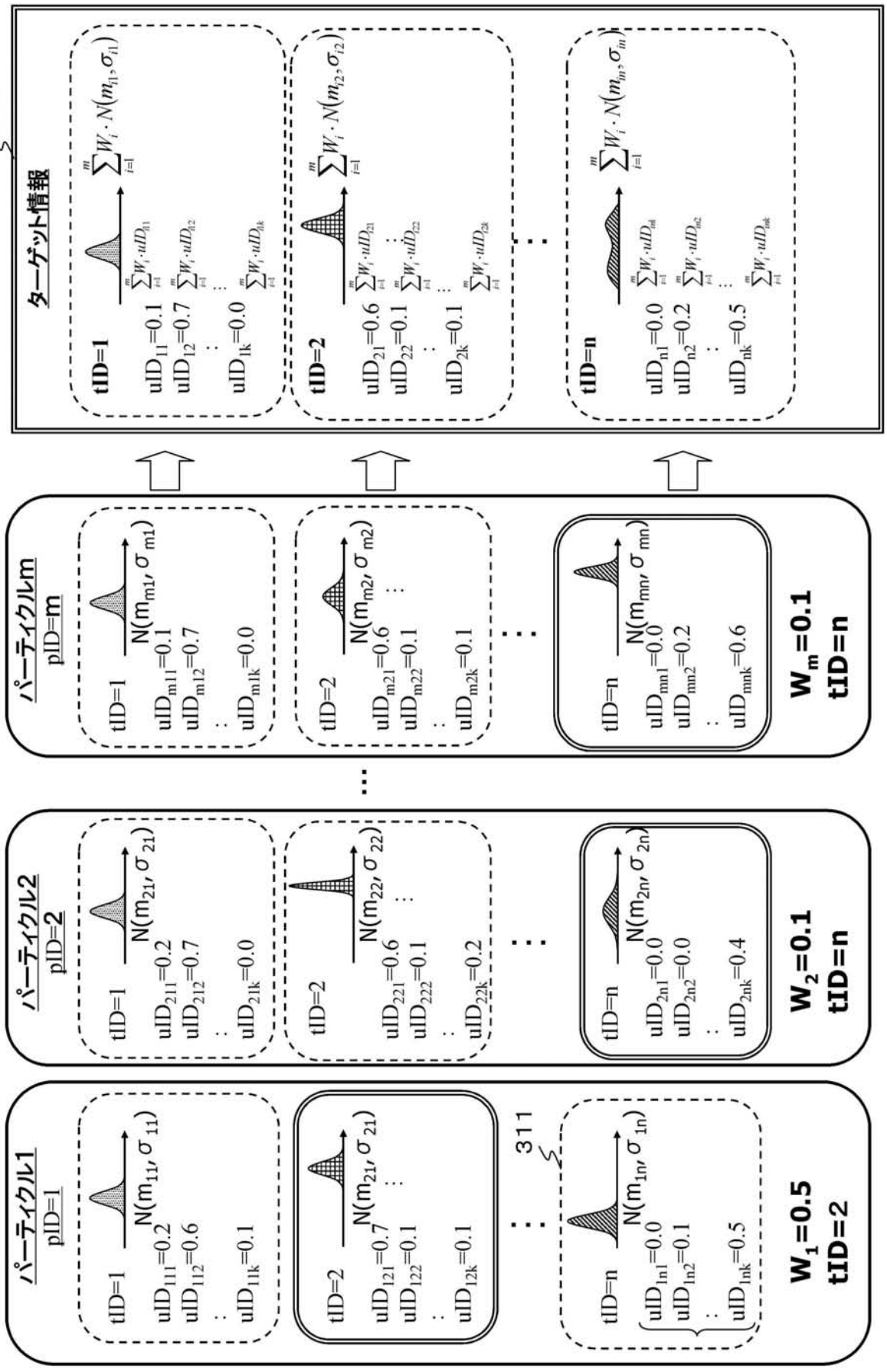
【図３】


(A)実環境
201
ユーザ1
「ニュースが
見たい」
202
ユーザ2
203
ユーザ3
20k
ユーザk
(B)音声／画像イベント情報
(a)ユーザ位置情報
期待値
(平均)
分散
N (m_e, σ_e)
存在確率
位置
m_e
(b)ユーザ識別情報
顔識別情報、
or
話者識別情報
確率(スコア)
uID: 1 2 3 ･･ k

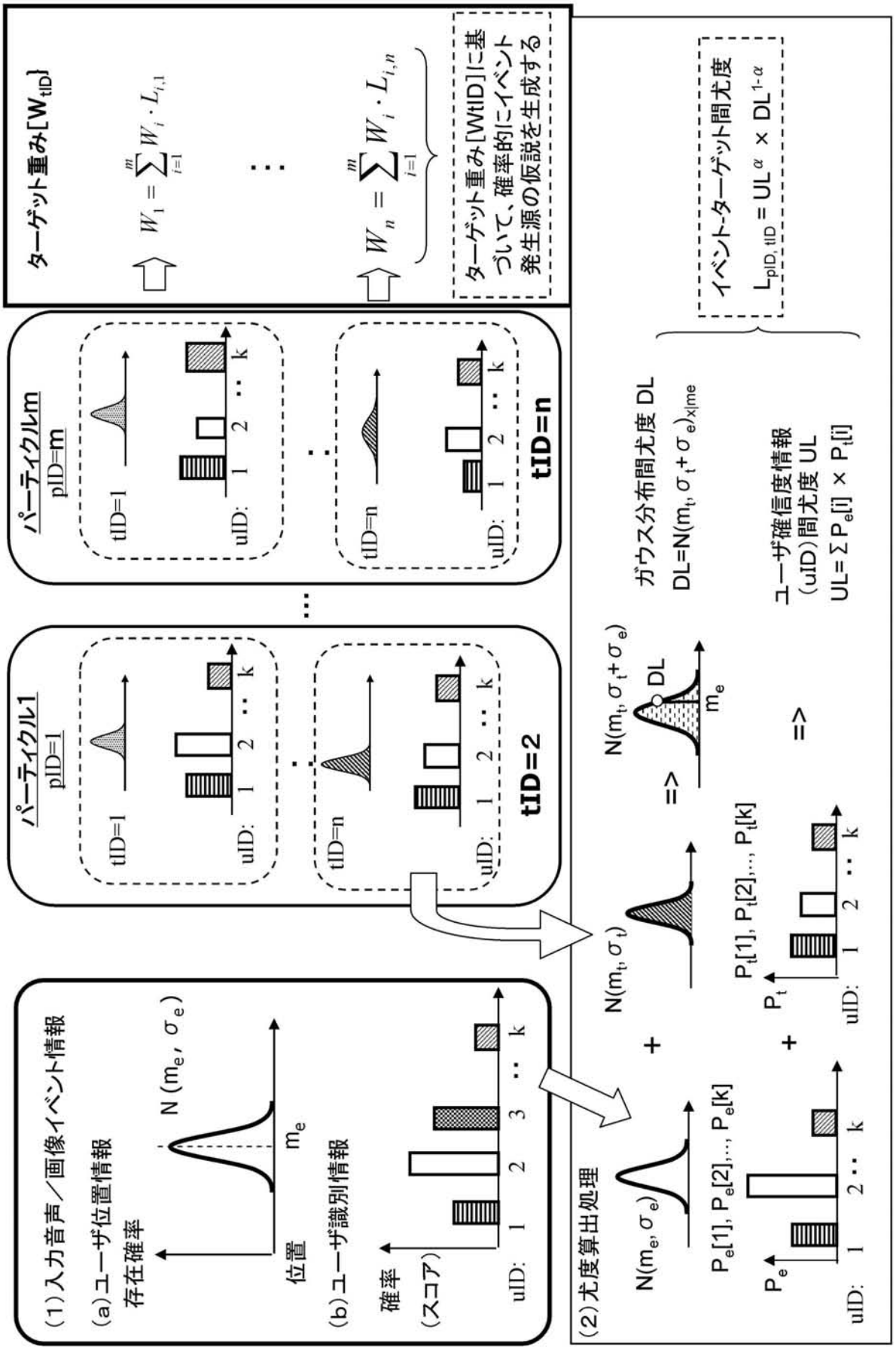
【図４】


301
H(存在確率)
(a)初期パーティクル
位置
s
初期状態[T0]
各パーティクルの有する推定位置情報
302
C
(b)入力情報に基づく存在確率分布情報
[T1]
(c)更新パーティクル

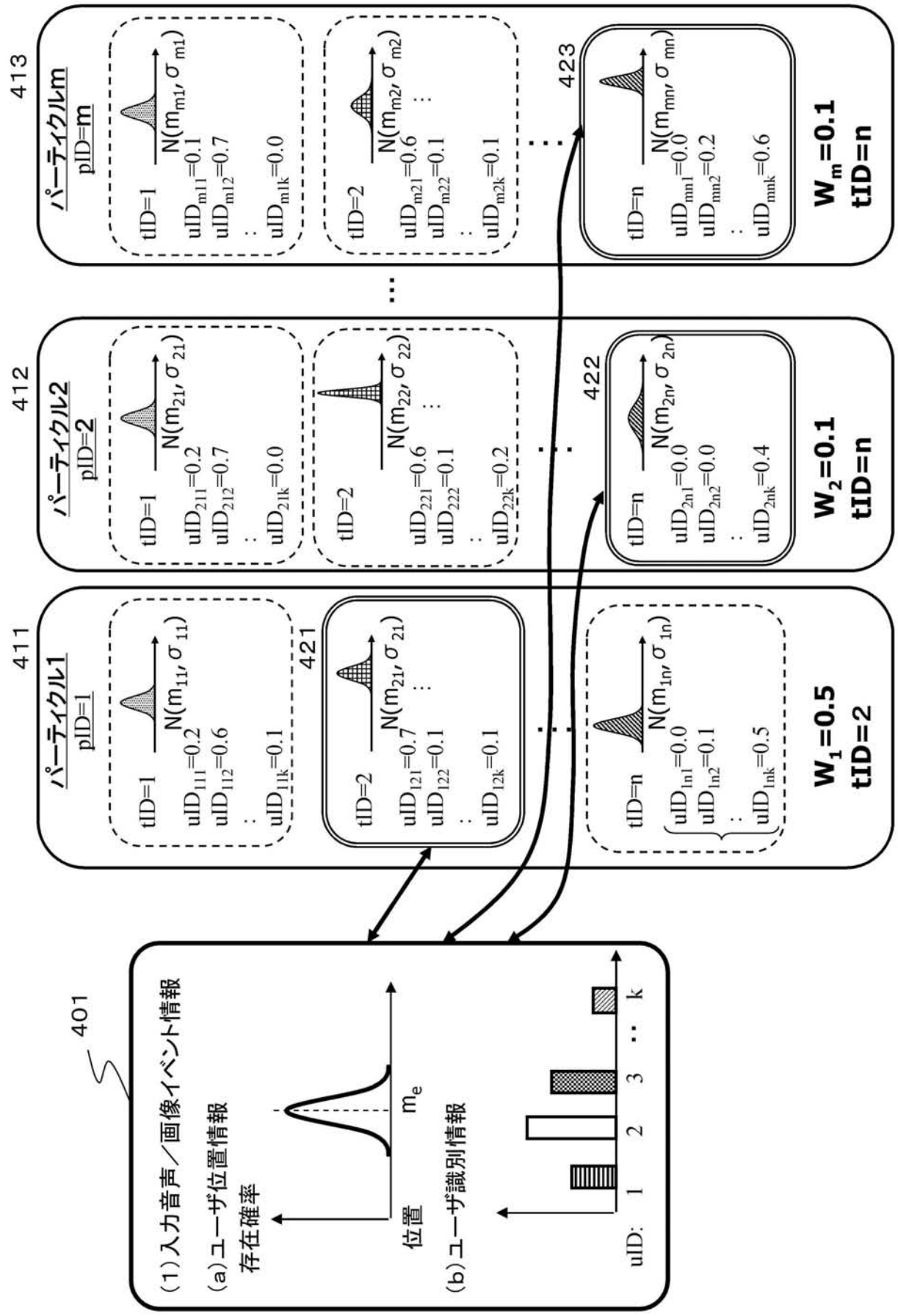
【図５】


305
ターゲット情報
tID=1
tID=2
tID=n
311
パーティクル1
pID=1
パーティクル2
pID=2
パーティクルm
pID=m
W1=0.5
tID=2
W2=0.1
tID=n
Wm=0.1
tID=n

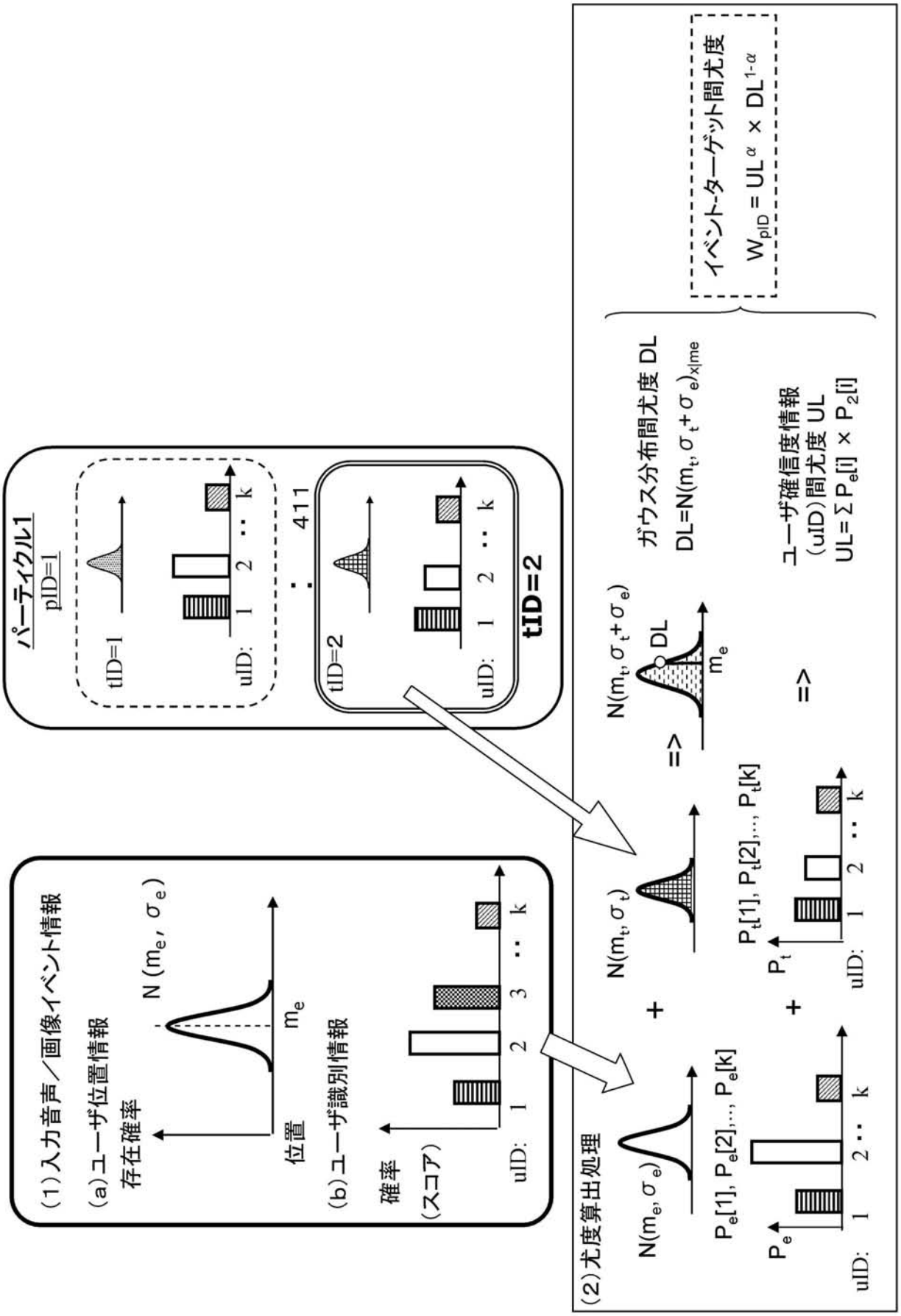
【図８】


(1)入力音声／画像イベント情報
(a)ユーザ位置情報
存在確率
N (m_e, σ_e)
位置
m_e
(b)ユーザ識別情報
確率
(スコア)
uID: 1 2 3 ‥ k
(2)尤度算出処理
N(m_e, σ_e)
P_e[1], P_e[2],.., P_e[k]
P_e
uID: 1 2 ‥ k
+
N(m_t, σ_t)
P_t[1], P_t[2],.., P_t[k]
P_t
uID: 1 2 ‥ k
=>
N(m_t, σ_t+σ_e)
DL
m_e
=>
ガウス分布間尤度 DL
DL=N(m_t, σ_t+σ_e)_x|me
ユーザ確信度情報
(uID)間尤度 UL
UL=Σ P_e[i] × P_t[i]
イベント-ターゲット間尤度
L_pID, tID = UL^α × DL^1-α
パーティクル1
pID=1
tID=1
tID=n
uID: 1 2 ‥ k
tID=2
パーティクルm
pID=m
tID=1
tID=n
uID: 1 2 ‥ k
tID=n
ターゲット重み[W_tID]
W_1 = Σ_{i=1}^{m} W_i · L_{i,1}
W_n = Σ_{i=1}^{m} W_i · L_{i,n}
ターゲット重み[WtID]に基づいて、確率的にイベント発生源の仮説を生成する

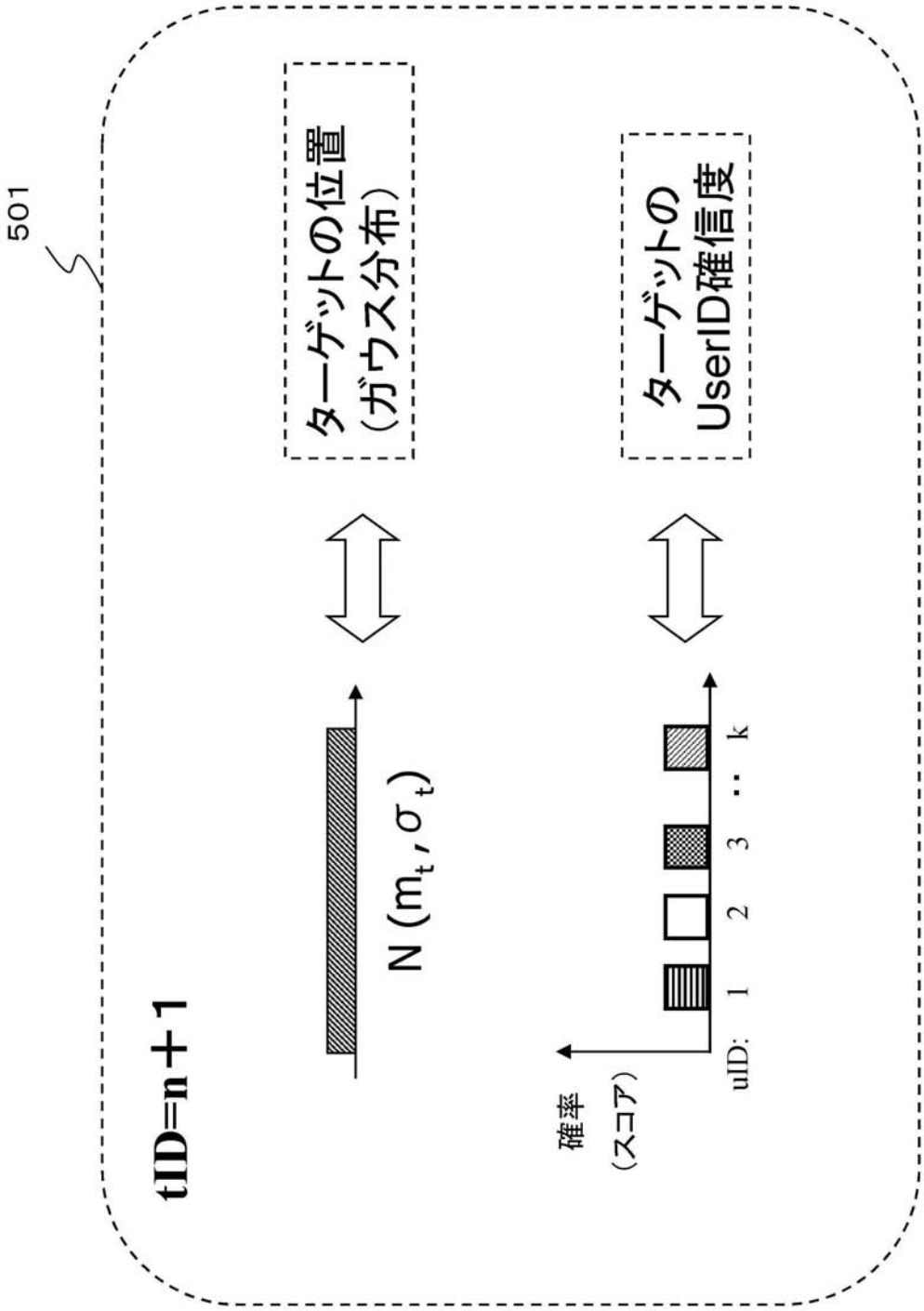
【図９】


401
(1)入力音声／画像イベント情報
(a)ユーザ位置情報
存在確率
位置
m_e
(b)ユーザ識別情報
uID: 1 2 3 .. k
411
パーティクル1
pID=1
412
パーティクル2
pID=2
413
パーティクルm
pID=m
tID=1 $N(m_{11}, \sigma_{11})$
uID_{111}=0.2
uID_{112}=0.6
uID_{11k}=0.1
tID=2 $N(m_{21}, \sigma_{21})$
uID_{121}=0.7
uID_{122}=0.1
uID_{12k}=0.1
421
tID=n $N(m_{1n}, \sigma_{1n})$
uID_{1n1}=0.0
uID_{1n2}=0.1
uID_{1nk}=0.5
W_1=0.5
tID=2
tID=1 $N(m_{21}, \sigma_{21})$
uID_{211}=0.2
uID_{212}=0.7
uID_{21k}=0.0
tID=2 $N(m_{22}, \sigma_{22})$
uID_{221}=0.6
uID_{222}=0.1
uID_{22k}=0.2
422
tID=n $N(m_{2n}, \sigma_{2n})$
uID_{2n1}=0.0
uID_{2n2}=0.0
uID_{2nk}=0.4
W_2=0.1
tID=n
tID=1 $N(m_{m1}, \sigma_{m1})$
uID_{m11}=0.1
uID_{m12}=0.7
uID_{m1k}=0.0
tID=2 $N(m_{m2}, \sigma_{m2})$
uID_{m21}=0.6
uID_{m22}=0.1
uID_{m2k}=0.1
423
tID=n $N(m_{mn}, \sigma_{mn})$
uID_{mn1}=0.0
uID_{mn2}=0.2
uID_{mnk}=0.6
W_m=0.1
tID=n

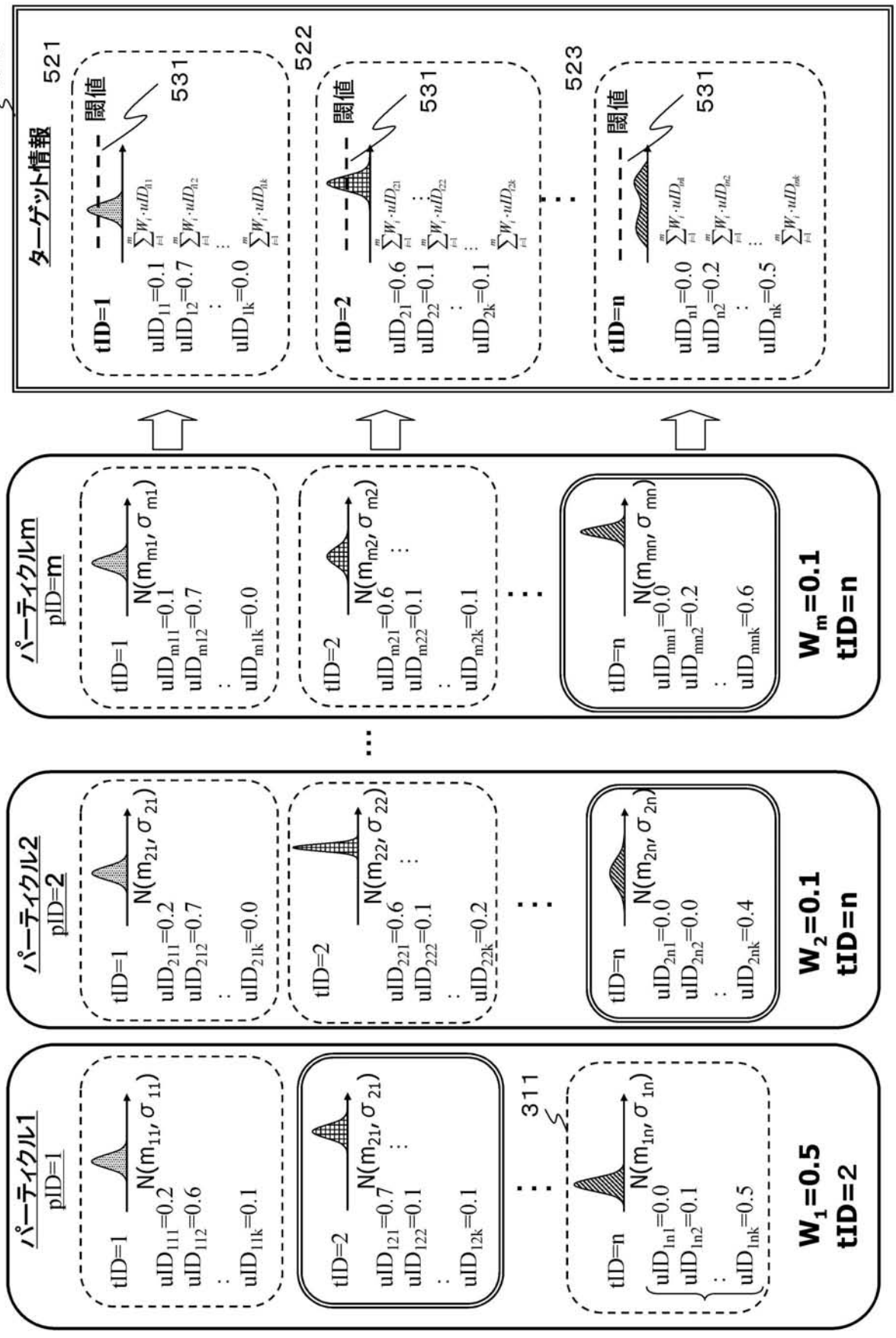
【図１０】


パーティクル1
pID=1
tID=1
uID: 1 2 ‥ k
‥ 411
tID=2
uID: 1 2 ‥ k
tID=2
(1)入力音声／画像イベント情報
(a)ユーザ位置情報
存在確率
N (me, σe)
位置
me
(b)ユーザ識別情報
確率
(スコア)
uID: 1 2 3 ‥ k
(2)尤度算出処理
N(me, σe)
Pe[1], Pe[2],‥, Pe[k]
Pe
uID: 1 2 ‥ k
+
N(mt, σt)
Pt[1], Pt[2],‥, Pt[k]
Pt
uID: 1 2 ‥ k
=>
N(mt, σt+σe)
DL
me
ガウス分布間尤度 DL
DL=N(mt, σt+σe)x|me
=>
ユーザ確信度情報
(uID)間尤度 UL
UL=Σ Pe[i] × P2[i]
イベント-ターゲット間尤度
WplD = UL^α × DL^1-α

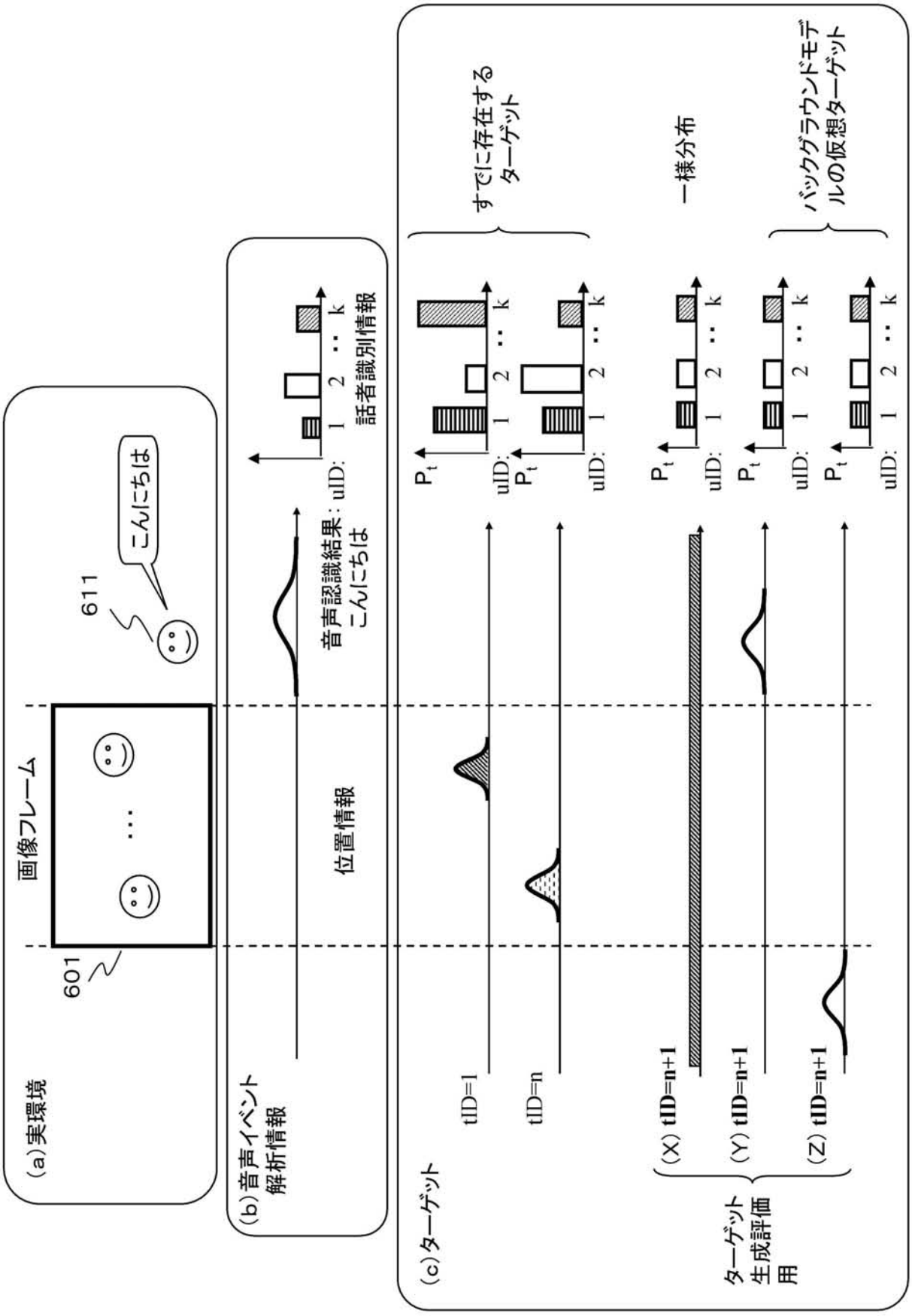
【図１１】


501
tID=n+1
ターゲットの位置
（ガウス分布）
N (m_t , σ_t)
ターゲットの
UserID確信度
確率
（スコア）
uID: 1 2 3 .. k

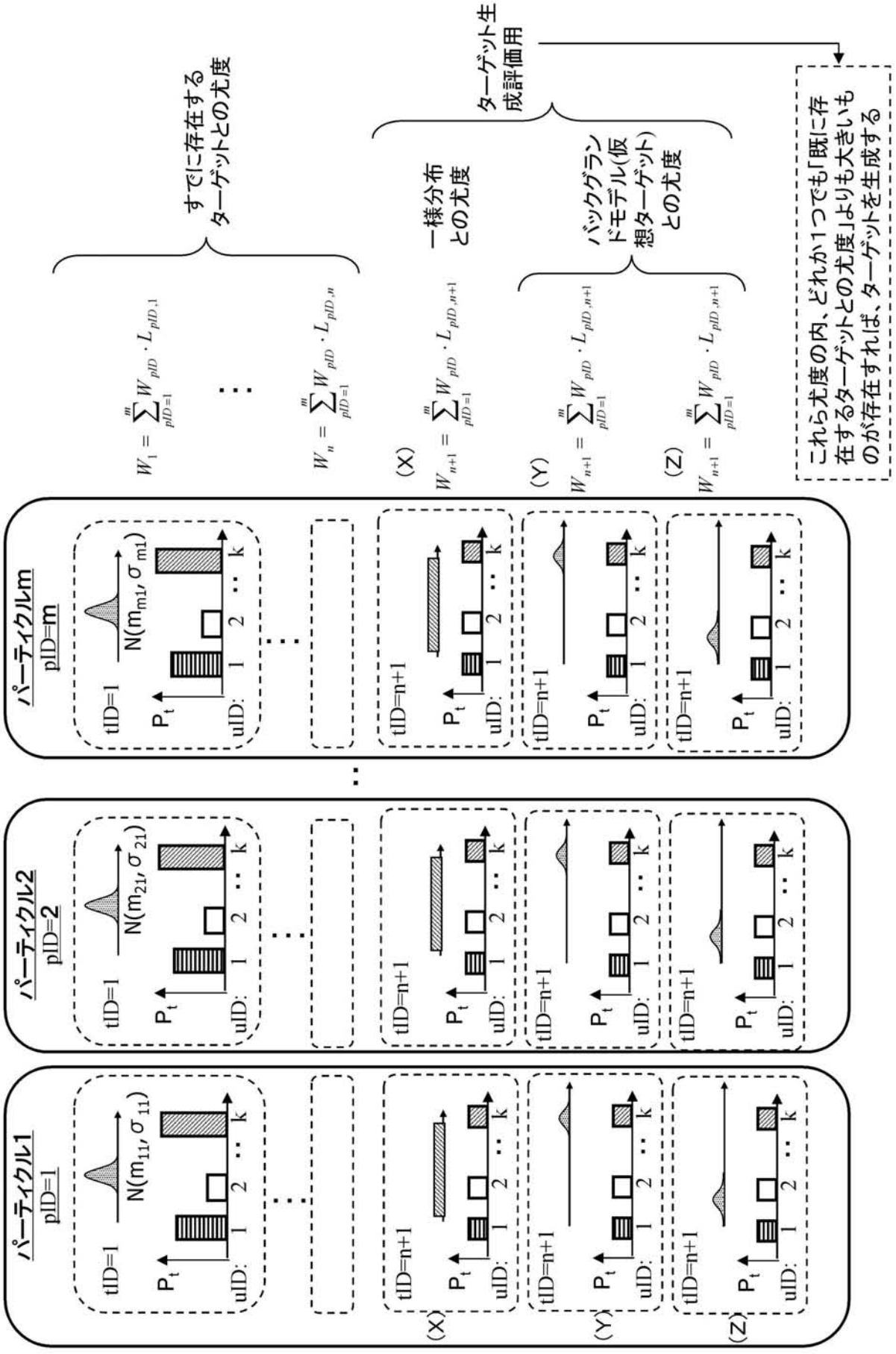
【図１２】


520
ターゲット情報
521
522
523
tID=1
tID=2
tID=n
閾値
531
311
パーティクル1
pID=1
パーティクル2
pID=2
パーティクルm
pID=m
W1=0.5
tID=2
W2=0.1
tID=n
Wm=0.1
tID=n

【図１３】


(a)実環境
画像フレーム
601
611
こんにちは
(b)音声イベント
解析情報
位置情報
音声認識結果：uID:
こんにちは
1 2 ‥ k
話者識別情報
(c)ターゲット
tID=1
tID=n
すでに存在する
ターゲット
一様分布
バックグラウンドモデ
ルの仮想ターゲット
(X) tID=n+1
(Y) tID=n+1
(Z) tID=n+1
ターゲット
生成評価
用
Pt
uID:

【図１４】


パーティクル1
pID=1
パーティクル2
pID=2
パーティクルm
pID=m
tID=1
tID=n+1
uID: 1 2 .. k
(X)
(Y)
(Z)
すでに存在するターゲットとの尤度
一様分布との尤度
バックグランドモデル(仮想ターゲット)との尤度
ターゲット生成評価用
これら尤度の内、どれか1つでも「既に存在するターゲットとの尤度」よりも大きいものが存在すれば、ターゲットを生成する

フロントページの続き

(51)Int.Cl.	F I	テーマコード（参考）
	Ｂ２５Ｊ 19/02	

Ｆターム(参考) 3C007 AS36 KS11 KS13 KS39 KT01 LW03 WB13 WB14 WB17
5B087 AA09 CC33
5D015 AA03
5E501 AA01 AC37 BA05 CB14 CB15