

## (12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局(43) 国際公開日  
2014年12月18日(18.12.2014)

WIPO | PCT

(10) 国際公開番号

WO 2014/199596 A1

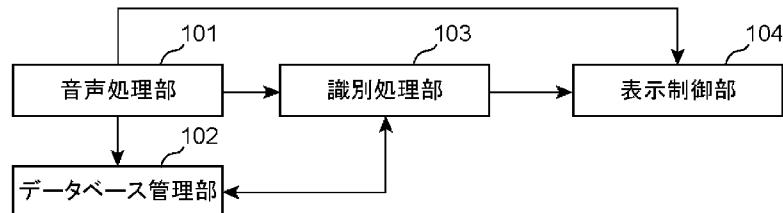
- (51) 国際特許分類:  
*G10L 15/22* (2006.01)      *G10L 15/00* (2013.01)  
*G06F 3/16* (2006.01)      *G10L 17/00* (2013.01)
- (21) 国際出願番号: PCT/JP2014/002970
- (22) 国際出願日: 2014年6月4日(04.06.2014)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
 特願 2013-121714 2013年6月10日(10.06.2013) JP
- (71) 出願人: パナソニック インテレクチュアル  
 プロパティ コーポレーション オブ アメリカ (PANASONIC INTELLECTUAL PROPERTY  
 CORPORATION OF AMERICA) [US/US]; 90503 カ  
 リフォルニア州、トーランス、スイート 20  
 0、マリナー アベニュー 20000 California (US).
- (72) 発明者: 松岡 智美(MATSUOKA, Tomomi).
- (74) 代理人: 小谷 悅司, 外(KOTANI, Etsushi et al.), 〒  
 5300005 大阪府大阪市北区中之島2丁目2番2  
 号大阪中之島ビル2階 Osaka (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

## 添付公開書類:

- 国際調査報告(条約第21条(3))
- 補正された請求の範囲及び説明書(条約第19条(1))

(54) Title: SPEAKER IDENTIFICATION METHOD, SPEAKER IDENTIFICATION DEVICE, AND SPEAKER IDENTIFICATION SYSTEM

(54) 発明の名称: 話者識別方法、話者識別装置及び話者識別システム

205

- 101 Speech processing unit  
 102 Database management unit  
 103 Identification processing unit  
 104 Display control unit

(57) Abstract: The present invention provides a speaker identification method for a speaker identification system. This system stores: a plurality of registered speech signals respectively generated on the basis of speech of a plurality of speakers; and a plurality of speaker images which are respectively associated with the plurality of registered speech signals, and which respectively represent the plurality of speakers. In this method, speech of a speaker located in the vicinity of a display is acquired, a speaker-speech signal is generated from the speech of the speaker, a registered speech signal corresponding to the speaker-speech signal is identified from among the plurality of stored registered speech signals, and a speaker image associated with the identified registered speech signal is displayed on the display for a time period in which at least the speech of the speaker serving as the source from which the speaker-speech signal was generated is being acquired.

(57) 要約:

[続葉有]



---

本開示は、話者識別システムにおける話者識別方法である。システムは、複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、複数の登録音声信号にそれぞれ対応付けられ複数の話者をそれぞれ表す複数の話者画像とを保存する。方法は、表示部の周辺に位置する話者の音声を取得し、話者の音声から話者音声信号を生成し、保存されている複数の登録音声信号のうち、話者音声信号に対応する登録音声信号を識別し、識別された登録音声信号に対応付けられている話者画像を、少なくとも話者音声信号を生成する元となった話者の音声が取得されている間、表示部に表示させる。

## 明細書

### 発明の名称：話者識別方法、話者識別装置及び話者識別システム 技術分野

[0001] 本開示は、話者を識別して表示部に識別した話者を表す話者画像を表示する話者識別方法、話者識別装置及び話者識別システムに関するものである。

#### 背景技術

[0002] 従来、話者識別および音声認識装置として、音声信号中に含まれる情報を用いて話者を識別する方法が提案されている。特許文献1では、会話内容を音声認識により文字データで記録する際に、一文字毎にタイムスタンプと音声から抽出する音声特徴を合わせて記録し、同一話者の発言内容を色や表示位置で区分して表示する方法が開示されている。これにより話者別に識別可能な会議システムを実現している。

[0003] また、特許文献2では、音声データを文字画像データに変換し、発声順に連なって移動する文字列からなる表示方法について開示されている。これにより画像と文字により情報が重層的に理解できる表示方法を実現している。

[0004] しかしながら、従来の構成において、更なる改善が必要であった。

#### 先行技術文献

##### 特許文献

[0005] 特許文献1：特開平10-198393号公報

特許文献2：特開2002-341890号公報

#### 発明の概要

[0006] 上記課題を解決するため、本開示の一態様は、表示部の周辺に位置する話者の音声を識別し識別結果を前記表示部に表示する話者識別システムにおける話者識別方法であって、

前記話者識別システムは、複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを保存するデータベースを含

み、

前記表示部の周辺に位置する話者の音声を取得し、

前記取得された話者の音声から話者音声信号を生成し、

前記データベースに保存されている複数の登録音声信号のうち、前記生成された話者音声信号に対応する登録音声信号を識別し、

前記識別された登録音声信号に対応付けて前記データベースに保存されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声が取得されている間、前記表示部に表示させるものである。

[0007] 本態様によれば、更なる改善を図ることができる。

### 図面の簡単な説明

[0008] [図1]実施の形態1における話者識別システムを構成する話者識別装置の構成例を示すブロック図である。

[図2]図1に示される話者識別装置の制御部の機能を示すブロック図である。

[図3]音声DBに蓄積されている音声情報の一例を示す図である。

[図4]音声DBに蓄積されている音声情報の別の例を示す図である。

[図5]実施の形態1の話者識別システムの図1に示される話者識別装置における処理を示すフローチャートである。

[図6]実施の形態1における話者識別システムの別の構成例を示すブロック図である。

[図7]図6の話者識別システムにおける動作の一例を示すシーケンス図である。

[図8A]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図8B]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図8C]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図8D]図1又は図6の話者識別システムにおいて、表示部に表示される登録

アイコンの具体的な表示例を示す図である。

[図8E]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図8F]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図8G]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図8H]図1又は図6の話者識別システムにおいて、表示部に表示される登録アイコンの具体的な表示例を示す図である。

[図9]実施の形態2における図1の話者識別装置の制御部の機能を示すブロック図である。

[図10]実施の形態2の図1の話者識別システムにおける処理を示すフローチャートである。

[図11A]ユーザが修正指示に用いる入力受付部の一例を示す図である。

[図11B]ユーザが修正指示に用いる入力受付部の一例を示す図である。

[図12]実施の形態2の図6の話者識別システムにおける動作の一例を示すシーケンス図である。

[図13A]実施の形態における話者識別システムの全体像を示す図である。

[図13B]データセンタ運営会社の一例を示す図である。

[図13C]データセンタ運営会社の一例を示す図である。

[図14]実施の形態におけるサービスの類型（自社データセンタ型）を示す図である。

[図15]実施の形態におけるサービスの類型（IaaS利用型）を示す図である。

[図16]実施の形態におけるサービスの類型（PaaS利用型）を示す図である。

[図17]実施の形態におけるサービスの類型（SaaS利用型）を示す図である。

## 発明を実施するための形態

[0009] (本開示の基礎となった知見)

家庭内の機器から使用状況、又はその機器を使用しているユーザの音声情報等を取得し、この取得した情報に基づいてユーザにサービスを提供するシステムが検討されている。しかし、ユーザにとって機器の使用状況又は音声情報は、個人情報に類する情報としての側面がある。よって、もし取得した機器の使用状況又は音声情報が可視化されないまま利用されると、どのように取得した情報が利用されているのかが不明確であり、ユーザにとって抵抗があると考えられる。よって、ユーザの抵抗を軽減するために、取得した情報を可視化して表示するシステムを開発する必要がある。

[0010] 更に、機器が取得した情報に誤検出がある場合は、誤検出の情報を可視化すると、ユーザに対してさらに不快感を与えててしまう。よって、取得した情報を可視化して表示しつつ、誤検出があった場合には、誤検出に基づき可視化された情報をユーザの操作で簡易に修正可能であることが好ましい。

[0011] また、ユーザから取得した情報を表示する装置として、わざわざ取得した情報を表示するためだけの専用表示機器を家庭内に配置することは、コストがかかる点及び設置場所を要してしまう点で、好ましくない。そこで例えば家庭内におけるテレビ受像機（以下、「テレビ」という）などの、本来、取得した情報の結果を表示することが目的ではない表示装置上に表示することが考えられる。テレビのような表示装置の場合、受信したテレビ放送の映像を表示画面に表示する必要がある。したがって、テレビ放送以外の取得した情報をテレビの表示画面上に表示する方法に関しては検討が必要である。一方で、上記したようにユーザにとっての抵抗感を減らすため、音声認識の結果は簡単に、かつ即座に確認できることが好ましい。

[0012] さらに、例えばテレビの表示画面上に取得した音声情報を表示する場合、テレビの周辺には不特定多数の人が存在する可能性が高い。それら複数の人の音声情報をテレビの表示画面上に、即座に明確かつ簡潔に表示し、修正まで可能なシステムに関しては、従来より検討がされていない。

- [0013] 特許文献1，2に記載の技術のように、話者識別および音声認識の結果を文字で表示する際には、複数の人が会話している場合又は話者が連續して複数回にわたり発声した場合において、文字列の表示画像が複雑になり、誰が識別されて表示されているのか明確にわかりづらい。また、まれに話者識別の結果が異なって表示された場合に、簡易な修正方法がないという問題点があった。
- [0014] また、特許文献1，2に記載の技術では、例えばテレビなど、本来、音声認識の結果を表示することが目的ではない表示装置上に、音声認識の結果を表示する際の、表示方法に関しては、十分な検討がされていない。
- [0015] 例えば特許文献1による技術では、単に、会議等の内容を記録する会話記録装置であって、音声認識結果を一文字毎にタイムスタンプおよび音声から抽出する特徴量を合わせて記録し、記録後にクラスタリング処理を行い、会話に参加していた人数と、各話者の音声特徴を求め、各話者の音声特徴と記録データを比較して話者を判別し、同一話者の発言内容を色や表示位置で区分して表示するものである。そのため、特許文献1に記載の技術では、複数話者が発声した場合、表示内容を簡易に且つ間違えなく確認して内容を修正することは困難であると考えられる。また、取得した音声情報を表示する例が示されているが、画面全体に音声情報を表示する例しか示されていない。よって、特許文献1に記載の技術には、テレビなど、本来音声認識の結果を表示することが目的ではない表示装置上に、音声情報を表示することに関しては課題の認識すらない。
- [0016] また、特許文献2による技術は、音声信号に含まれる言語情報と音声特性情報との双方を迅速かつ簡便に理解できる音声認識文字表示装置に関するものである。この技術では、単に文字画像データに変換し、発声順に連なって移動する文字列からなる表示方法について開示されている。特許文献2に記載の技術では、画像と文字により情報が重層的に理解できる表示方法を実現しているため、表示に誤りがあった際に、簡便に変更することは困難であると考えられる。

[0017] 本開示は、以上のような従来の話者識別および音声認識表示装置の課題を解決するものである。本開示の一態様により、複数の話者の音声情報を取得し、例えばテレビのような表示装置において、本来表示すべきコンテンツを表示しつつ、取得した音声情報を即座に明確かつ簡潔に表示することができる装置が提供される。さらには、本開示の一態様により、取得した情報に誤検出があった際等に、表示されている情報を簡易にユーザが修正可能な装置が提供される。

[0018] 本開示の一態様は、

表示部の周辺に位置する話者の音声を識別し識別結果を前記表示部に表示する話者識別システムにおける話者識別方法であって、

前記話者識別システムは、複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを保存するデータベースを含み、

前記表示部の周辺に位置する話者の音声を取得し、

前記取得された話者の音声から話者音声信号を生成し、

前記データベースに保存されている複数の登録音声信号のうち、前記生成された話者音声信号に対応する登録音声信号を識別し、

前記識別された登録音声信号に対応付けて前記データベースに保存されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声が取得されている間、前記表示部に表示させるものである。

[0019] 本態様によれば、表示部に話者を表す話者画像が表示されるため、ユーザに、話者の識別結果を明確に表示することができる。また、表示部に話者画像が表示されるのは、少なくとも話者音声信号を生成する元となった話者の音声が取得されている間である。このため、表示部により本来表示したいコンテンツ（例えば表示部がテレビ受像機の表示画面であればテレビ放送番組）の表示の過度の妨げになることが抑制される。

[0020] 上記態様において、例えば、

前記話者音声信号を生成する元となった前記話者の音声が取得されなくなった時点から所定時間が経過すると、前記表示されている前記話者画像を前記表示部から消去してもよい。

[0021] 本態様によれば、話者音声信号を生成する元となった話者の音声が取得されなくなった時点から所定時間が経過すると、表示されている話者画像が表示部から消去される。このため、表示部により本来表示したいコンテンツの表示の過度の妨げになるのが防止される。

[0022] 上記態様において、例えば、

前記データベースは、前記複数の音声登録信号として、第1話者の音声に基づき生成された第1登録音声信号と、第2話者の音声に基づき生成された第2登録音声信号とを保存し、かつ、前記第1登録音声信号に対応付けられ前記第1話者を表す第1話者画像と、前記第2登録音声信号に対応付けられ前記第2話者を表す第2話者画像とを保存し、

前記第1話者の音声が取得されると、第1話者音声信号が生成され、

前記生成された第1話者音声信号が前記第1登録音声信号に対応すると識別されると、少なくとも前記第1話者の音声が取得されている間、前記第1話者画像が前記表示部に表示され、

前記第1話者画像が前記表示部に表示されているときに、前記第2話者の音声が取得されると、第2話者音声信号が生成され、

前記生成された第2話者音声信号が前記第2登録音声信号に対応すると識別されると、少なくとも前記第2話者の音声が取得されている間、前記第1話者画像に加えて、前記第2話者画像が前記表示部に表示されてもよい。

[0023] 本態様によれば、少なくとも第1話者の音声が取得されている間、表示部には第1話者画像が表示され、少なくとも第2話者の音声が取得されている間、表示部には第2話者画像が表示される。このため、表示部に表示される話者画像によって、現在の話者を確認することができる。

[0024] 上記態様において、例えば、

前記第1話者の音声と前記第2話者の音声とが取得された順に、前記第1

話者画像と前記第2話者画像とが、前記表示部に並べられて表示されてもよい。

[0025] 本態様によれば、第1話者と第2話者との間で話者が交替する度に、表示部に表示される第1話者画像と第2話者画像との並び順が入れ替えられる。その結果、話者に対して発話が促されることとなる。

[0026] 上記態様において、例えば、

前記第1話者画像と前記第2話者画像とのうち、前記データベースに後で登録された方の話者画像は、前記データベースに先に登録された方の話者画像とは異なる態様で前記表示部に表示されてもよい。

[0027] 本態様によれば、第1話者画像と第2話者画像とのうち、データベースに後で登録された方の話者画像は、先に登録された方の話者画像とは異なる態様で表示部に表示される。このため、後から発話した話者を容易に確認することができる。

[0028] 上記態様において、例えば、

前記第1話者の発話回数と、前記第2話者の発話回数とがカウントされ、前記第1話者画像と前記第2話者画像とは、前記カウントされた発話回数の多い順に、前記表示部に並べられて表示されてもよい。

[0029] 本態様によれば、第1話者画像と第2話者画像とは、発話回数の多い順に、表示部に並べられて表示される。このため、第1話者と第2話者とに発話が促されることとなる。

[0030] 上記態様において、例えば、

前記話者画像についての話者からの修正指示が受け付けられると、前記修正を指示した話者の音声を新たに取得し、

前記新たに取得された話者の音声から新たに話者音声信号を生成し、前記データベースに保存されている、前記修正指示が行われた前記話者画像に対応付けられている前記登録音声信号を、前記新たに生成された話者音声信号に書き換えてよい。

[0031] 本態様によれば、話者画像についての話者からの修正指示が受け付けられ

ると、データベースに保存されている、修正指示が行われた話者画像に対応付けられている登録音声信号が、新たに生成された話者音声信号に書き換える。その結果、登録音声信号が誤っていたために、誤った話者画像が表示部に表示されても、修正が容易に行われる。

[0032] 上記態様において、例えば、

前記話者からの修正指示は、前記表示部に表示されている前記話者画像について受け付けられ、前記表示部に表示されていない前記話者画像については受け付けられないようにしてもよい。

[0033] 本態様によれば、話者からの修正指示は、表示部に表示されていない話者画像については受け付けられないため、例えば話者による誤った修正指示を受け付けるような事態が避けられる。

[0034] 上記態様において、例えば、

前記生成された話者音声信号から前記話者の属性を判別し、  
前記判別された前記話者の属性に基づき前記話者画像を作成し、  
前記生成された話者音声信号と前記判別された前記話者の属性と前記作成された前記話者画像とを、互いに対応付けて前記データベースに保存し、前記データベースには、前記生成された話者音声信号は、前記登録音声信号として保存されるようにしてもよい。

[0035] 本態様によれば、話者の音声が取得されると、登録音声信号と話者の属性と話者画像とが、互いに対応付けられてデータベースに保存される。したがって、ユーザによる登録のための操作数を低減することができる。話者の属性は、例えば話者の性別でもよい。話者の属性は、例えば話者の概略の年齢でもよい。

[0036] 本開示の他の態様は、

表示部と、

前記表示部の周辺に位置する話者の音声を取得する音声取得部と、  
前記取得された話者の音声から話者音声信号を生成する音声処理部と、  
複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前

記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを記憶するデータベースと、

前記データベースに記憶されている前記複数の登録音声信号のうち、前記生成された前記話者音声信号に対応する登録音声信号を識別する識別処理部と、

前記識別された前記登録音声信号に対応付けて前記データベースに記憶されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記表示部に表示する表示制御部とを備えるものである。

[0037] 本態様によれば、表示部に話者を表す話者画像が表示されるため、ユーザに、話者の識別結果を明確に表示することができる。また、表示部に話者画像が表示されるのは、少なくとも話者音声信号を生成する元となった話者の音声が取得されている間である。このため、表示部により本来表示したいコンテンツ（例えば表示部がテレビ受像機の表示画面であればテレビ放送番組）の表示の過度の妨げになることが抑制される。

[0038] 本開示のさらに他の態様は、

表示部と、

前記表示部の周辺に位置する話者の音声を取得する音声取得部と、

前記取得された話者の音声から話者音声信号を生成する音声処理部と、

ネットワークを介して外部のサーバ装置と通信を行う通信部と、

前記表示部を制御する表示制御部と、

を備え、

前記通信部は、前記生成された話者音声信号を前記サーバ装置に送信し、かつ、前記サーバ装置から、前記話者音声信号を基に識別された前記話者を表す話者画像を受信し、

前記表示制御部は、前記受信された話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記表示部に表示させるものである。

[0039] 本態様によれば、サーバ装置において、話者音声信号を基に話者を表す話者画像が識別される。サーバ装置から話者画像が通信部により受信される。受信された話者画像が、表示部に表示される。このため、ユーザに、話者の識別結果を明確に表示することができる。また、表示部に話者画像が表示されるのは、少なくとも話者音声信号を生成する元となった話者の音声が取得されている間である。このため、表示部により本来表示したいコンテンツ（例えば表示部がテレビ受像機の表示画面であればテレビ放送番組）の表示の過度の妨げになることが抑制される。

[0040] 本開示のさらに他の態様は、  
表示部の周辺に位置する話者の音声を取得する音声取得部と、  
前記取得された話者の音声から話者音声信号を生成する音声処理部と、  
複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを記憶する記憶部と、  
前記複数の登録音声信号のうち前記生成された話者音声信号に対応する登録音声信号を識別する識別処理部と、  
前記識別された前記登録音声信号に対応付けて前記記憶部に記憶されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記表示部に表示させる表示制御部と、を備えるものである。

[0041] 本態様によれば、表示部に話者を表す話者画像が表示されるため、ユーザに、話者の識別結果を明確に表示することができる。また、表示部に話者画像が表示されるのは、少なくとも話者音声信号を生成する元となった話者の音声が取得されている間である。このため、表示部により本来表示したいコンテンツ（例えば表示部がテレビ受像機の表示画面であればテレビ放送番組）の表示の過度の妨げになることが抑制される。

[0042] 以下、実施の形態について、図面を参照しながら説明する。

[0043] なお、以下で説明する実施の形態は、いずれも本開示の一具体例を示すも

のである。以下の実施の形態で示される数値、形状、構成要素、ステップ、ステップの順序などは、一例であり、本開示を限定する主旨ではない。また、以下の実施の形態における構成要素のうち、最上位概念を示す独立請求項に記載されていない構成要素については、任意の構成要素として説明される。また全ての実施の形態において、各々の内容を組み合わせることもできる。

[0044] (実施の形態 1)

図 1 は、実施の形態 1 における話者識別システムを構成する話者識別装置 200 の構成例を示すブロック図である。図 2 は、図 1 に示される話者識別装置 200 の制御部 205 の機能を示すブロック図である。

[0045] 話者識別装置 200 は、図 1 に示されるように、音声取得部 201、音声データベース (DB) 203、表示部 204、及び制御部 205 を備える。また、話者識別装置 200 は、通信部 202、入力受付部 206 をさらに備えていてもよい。話者識別装置 200 の制御部 205 は、図 2 に示されるように、音声処理部 101、データベース管理部 102、識別処理部 103、及び表示制御部 104 を含む。

[0046] ここで、話者識別装置 200 は、例えば一般家庭用のテレビ、パーソナルコンピュータ (PC) におけるモニタなどでもよい。ここでは、話者識別装置 200 として、特に、上記の「本開示の基礎となった知見」にて説明した通り、話者識別結果を表示するためだけの専用表示装置というよりは、別のコンテンツ等を表示することのできる装置を想定している。ただし、表示機能を有する装置に上記の各構成が備えられていれば、いかなる装置を採用してもよい。

[0047] また、各構成は、必ずしも、話者識別装置 200 の筐体の内部に配置されていなくてもよい。例えば音声取得部 201 が、話者識別装置 200 の筐体の外部に接続されていても、その音声取得部 201 は、話者識別装置 200 に含まれる。話者識別装置 200 は、一の家庭に一台のみ配置されているとは限らず、一の家庭に複数台配置されている場合もある。話者識別装置 20

0は、この実施の形態1では、一般家庭用のテレビである。

- [0048] 音声取得部201は、例えばマイクロフォンなどである。音声取得部201は、話者識別装置200を視聴している視聴者が発話した音声を取得する。ここで、音声取得部201には、指向性を制御する機器が備わっていてよい。この場合には、視聴者の存在する方向に指向性を持たせることにより、視聴者が発話した音声を取得する精度を向上させることができる。また、話者が位置する方向を検出することができる。
- [0049] また、音声取得部201は、人の発話に関する音声以外の音を、取得しない機能（もしくは除去する機能）を有していてよい。この実施の形態1のように、例えば話者識別装置200がテレビの場合、音声取得部201は、取得した音声からテレビの音声信号を除去する機能を有していてよい。これによって、視聴者が発話した音声を取得する精度を向上させることができる。
- [0050] 音声DB203は、情報を蓄積（記録）することが可能な記録媒体等で構成される。音声DB203は、話者識別装置200の筐体の内部に設けられていなくてもよい。音声DB203が例えば外付けの記録媒体等で構成されて、話者識別装置200の筐体の外部に接続されていても、音声DB203は、話者識別装置200に含まれる。
- [0051] 音声DB203には、話者識別装置200を保持している家族の音声、家族の動作音または家族の音声以外の音声、さらに家族の構成員（ユーザ）の年齢及び性別等の情報が蓄積されて管理されている。音声DB203に蓄積されている情報の詳細については特に限定しないが、音声取得部201が取得した話者識別装置200周辺の音声から、ユーザを特定できる情報が蓄積されていればよい。
- [0052] この実施の形態1では、例えば、音声DB203には、登録音声信号（音声信号のスペクトル又は周波数等から生成された情報）と、ユーザ情報（年齢、性別又はニックネーム等の情報）とが、互いに対応付けられて蓄積されている。また、この実施の形態1では、音声DB203には、各ユーザに対

応する話者画像が対応付けられて蓄積されている。

- [0053] 図3は、音声DB203に蓄積されている音声情報800の一例を示す図である。音声情報800は、互いに対応付けられた、登録音声信号801と、ユーザ情報802と、登録アイコン803（話者画像の一例）とを含む。
- [0054] 図3では、登録音声信号801は、音声信号のスペクトル又は周波数等の情報を基に生成された予め設定された次元数の特徴ベクトルを表す信号である。登録音声信号801は、この実施の形態1では、「. wav」形式のファイルとして登録されている。なお、登録音声信号801は、「. wav」形式のファイルでなくてもよい。例えば、登録音声信号801は、MPEG-1 Audio Layer 3、Audio Interchange File Format等の音声圧縮されたデータで生成されてもよい。また、登録音声信号801は、例えば自動的に圧縮ファイルにエンコードされ、音声DB203に格納されてもよい。
- [0055] ユーザ情報802は、ユーザ（話者）の属性を表す情報である。この実施の形態1では、図3に示されるように、ユーザ情報802は、ユーザの属性として、「年齢」、「性別」、「ニックネーム」を含む。図3のユーザ情報802の例では、登録音声信号801が「0001.wav」のユーザに対応付けて、「年齢」が「40代」に、「性別」が「男性」に、「ニックネーム」が「パパ」に、それぞれ設定されている。「年齢」及び「性別」については、データベース管理部102等が自動的に登録してもよく、ユーザが入力受付部206を用いて登録してもよい。「ニックネーム」については、ユーザが入力受付部206を用いて登録すればよい。
- [0056] 登録アイコン803は、ユーザ（話者）を表す話者画像である。図3の登録アイコン803の例では、登録音声信号801が「0001.wav」のユーザに対応付けて、「アイコンA01」が設定され、登録音声信号801が「0003.wav」のユーザに対応付けて、「アイコンB05」が設定されている。登録アイコン803は、後述の図8Aに示されるように、○、□、△のような記号のアイコンでもよい。あるいは、登録アイコン803は

、後述の図8Bに示されるように、人の顔を模式的に表したアイコンでもよい。

[0057] 登録アイコン803に関しては、制御部205は、予め作成された複数のアイコンのなかからユーザが選択したアイコン、またはユーザが自ら作成した画像を、登録アイコン803として音声情報800に登録してもよい。また、ユーザによってアイコンが音声情報800に登録されていない場合でも、制御部205が、ユーザ情報802を基に、ユーザ情報802に適合するアイコンを選択して、又は作成して、音声情報800に登録するようにしてもよい。

[0058] 音声DB203に蓄積される音声情報800の構築方法についても特に限定しない。例えば予めユーザが初期登録することで音声情報800を構築できる。例えば、初期登録では、話者識別装置200の前にいるユーザが発話をするたびに、音声取得部201が音声を取得する。音声処理部101は、取得された話者の音声から特徴ベクトルを生成し、生成した特徴ベクトルを表す話者音声信号を生成する。データベース管理部102は、生成された話者音声信号を、登録音声信号801として自動的に音声DB203の音声情報800に登録する。このようにして、音声DB203が完成されてもよい。

[0059] また、その初期登録の際に、入力受付部206は、ユーザが発話したタイミングでユーザ情報802を入力させるようなユーザインターフェースを表示部204に表示させてもよい。データベース管理部102は、ユーザの入力受付部206に対するユーザ情報802の入力内容を用いて、音声DB203の音声情報800を更新してもよい。

[0060] なお、音声情報800に関して、上述のように初期登録によって音声DB203に予め登録しておかなくても、ある程度の話者の情報を識別することもできる。一般的に、年齢差や性別によって、話者の音声の基本周波数が異なることが知られている。例えば、男性が発話する音声の基本周波数の平均は150Hz～550Hz、女性が発話する音声の基本周波数の平均は40

0 Hz～700 Hzと言われている。よって、初期登録に代えて、話者識別装置200の識別処理部103は、音声処理部101で生成された音声を表す信号の周波数などの情報を基に、ある程度の年齢及び性別を判断してもよい。データベース管理部102は、識別処理部103の判断結果に基づき、自動で音声DB203の音声情報800に、登録音声信号801及びユーザ情報802を登録するようにしてもよい。

- [0061] また、ユーザ情報802に関しては、図3に記載のものに限られない。制御部205は、各ユーザごとに視聴履歴の多い番組などの嗜好情報を、ユーザ情報802として音声DB203に蓄積するようにしてもよい。また、ユーザ情報802の取得方法に関しても限定しない。ユーザが、話者識別装置200を最初に使用するときに入力受付部206を用いてユーザ情報802の初期設定をするようにしてもよい。あるいは、最初に、ユーザの音声が取得されたタイミングで、ユーザが入力受付部206を用いてユーザ情報802を登録するようにしてもよい。
- [0062] 図4は、音声DB203に蓄積される音声情報810の別の例を示す図である。図4に示される音声情報810は、互いに対応付けられた、登録音声信号801と音声データ804とを含む。音声データ804は、音声取得部201によって取得された話者の音声から、音声処理部101によって生成された発話内容を表すデータである。図4に示されるような音声情報810が、音声DB203に蓄積されていくようにしてもよい。
- [0063] この場合には、音声処理部101は、話者の音声の特徴ベクトルを表す話者音声信号に加えて、発話内容を表すデータを生成する。音声処理部101は、例えば音響モデル及び言語モデルを用いる音声認識技術によって、発話内容を表すデータを生成する。データベース管理部102は、音声処理部101によって生成された発話内容を表すデータを、音声DB203に音声データ804として蓄積する。
- [0064] 識別処理部103は、さらに、音声処理部101から出力された発話内容を表すデータと、音声DB203に蓄積されている音声データ804（発話

内容)とを比較する。これによって、話者の特定精度を向上させることができる。

[0065] 図4の例では、登録音声信号801が「0002. wav」のユーザが、あるタイミングで「料理番組を見ながら夕食を作りましょう」と発話していたことが記録されている。これにより、登録音声信号801が「0002. wav」に対応する話者が、別のタイミングで例えば「料理番組」等の同様の単語を発話した場合には、登録音声信号801が「0002. wav」に対応する話者の発話である可能性が高いと、識別処理部103が判断できる。

[0066] 図1に戻って、表示部204は、一般的なモニタなどであり、特に限定しない。この実施の形態1では、表示部204は、テレビの表示画面である。表示部204は、制御部205の表示制御部104により制御されて、画像または情報を表示する。本実施の形態1の話者識別システムでは、表示部204は、取得された話者の音声に対応付けられた登録アイコン803を表示する。これにより、ユーザは、話者識別表示システムにより、誰が、または複数の人数が識別されているのかが明確に分かるようになる。

[0067] また、後述の実施の形態2の話者識別システムでは、話者識別装置200の周辺に複数のユーザが存在する場合などに、話者識別が誤っていたことにより異なる登録アイコン803が表示された際には、簡易に修正が可能な構成にしている。表示部204に表示される登録アイコン803等の具体例は、図8A～8Fを参照して後述する。

[0068] 制御部205は、例えばCPU又はマイクロコンピュータ、及びメモリなどを含む。制御部205は、音声取得部201、音声DB203、及び表示部204などの各構成の動作を制御する。例えばメモリに格納されたプログラムにしたがってCPU又はマイクロコンピュータが動作することにより、制御部205は、図2に示される音声処理部101、データベース管理部102、識別処理部103、及び表示制御部104として機能する。図2に示される制御部205の各機能は、図5を参照して後述される。

[0069] ここで、上記したように、話者識別装置200は通信部202を備えていてもよい。通信部202は、インターネット等と接続することで他の機器やサーバ装置と通信し、情報のやりとりを行う。

[0070] また、話者識別装置200は、入力受付部206を備えていてもよい。入力受付部206は、ユーザからの入力を受け付ける。ユーザからの入力を受け付ける方法としては特に限定しない。入力受付部206をテレビのリモートコントローラで構成してもよい。あるいは、入力受付部206は、表示部204に操作のためのユーザインターフェースを表示してもよい。ユーザは、それらの入力受付部206によって、情報又は指示を入力することができる。

[0071] 図5は、本実施の形態1の話者識別システムの図1に示される話者識別装置200における処理を示すフローチャートである。

[0072] まず、ステップS301では、音声取得部201は、話者の発した音声を取得する。音声処理部101は、取得された話者の音声から、予め設定された次元数の特徴ベクトルを生成し、生成した特徴ベクトルを表す話者音声信号を生成する。

[0073] 続いて、ステップS302では、データベース管理部102は、音声DB203に蓄積されている音声情報800(図3)から、登録音声信号801を抽出して、識別処理部103に出力する。識別処理部103は、音声処理部101により生成された話者音声信号と、データベース管理部102から出力された登録音声信号801とを比較し、話者音声信号に対応する登録音声信号801を特定する。

[0074] 識別処理部103は、例えば話者音声信号と、音声DB203に蓄積されている各登録音声信号801との類似度をそれぞれ算出する。識別処理部103は、算出した類似度のうちで、最も高い類似度を抽出する。識別処理部103は、この最も高い類似度が予め設定された閾値以上のときに、その最も高い類似度に対応する登録音声信号801を、話者音声信号に対応すると判断する。具体的には例えば、識別処理部103は、話者音声信号の特徴ベ

クトルと登録音声信号 801 の特徴ベクトルとの距離をそれぞれ算出する。識別処理部 103 は、算出した距離が最も短い登録音声信号 801 と話者音声信号との類似度が最も高いと判断する。

- [0075] 続いて、ステップ S303 では、識別処理部 103 は、特定した登録音声信号 801 をデータベース管理部 102 に出力する。データベース管理部 102 は、音声 DB203 に蓄積されている音声情報 800 (図 3) を参照して、出力された登録音声信号 801 に対応付けられている登録アイコン 803 を抽出する。データベース管理部 102 は、抽出した登録アイコン 803 を識別処理部 103 に出力する。
- [0076] 識別処理部 103 は、出力された登録アイコン 803 を表示制御部 104 に出力する。音声処理部 101 は、音声取得部 201 により話者の音声が取得されている間だけ、その旨を表す取得信号を、話者ごとに表示制御部 104 に出力する。
- [0077] 表示制御部 104 は、識別処理部 103 から出力された登録アイコン 803 を、音声処理部 101 から取得信号が入力されている間、表示部 204 に表示する。表示制御部 104 は、音声取得部 201 が取得している話者の音声から、特定された話者を示す音声が所定時間途絶えたとき、つまり音声処理部 101 から特定された話者の取得信号が入力されなくなった時点から所定時間 (この実施の形態 1 では、例えば 10 秒) が経過したときに、表示制御部 104 は、表示部 204 に表示されているアイコンを消去する。この場合、表示制御部 104 は、表示されているアイコンの透明度を徐々に増大させて、アイコンを表示部 204 からフェードアウトさせるようにしてもよい。
- [0078] 図 6 は、本実施の形態 1 における話者識別システムの別の構成例を示すブロック図である。図 6 では、図 1 と同一要素については同一符号が付されている。以下、図 1 の話者識別システムとの相違点を中心に、図 6 の話者識別システムが説明される。
- [0079] 図 6 の話者識別システムは、話者識別装置 200 と、サーバ装置 210 と

を備える。図6の話者識別システムでは、音声DB203は、図1の話者識別システムと異なり、サーバ装置210に含まれている。すなわち、話者識別装置200は、音声取得部201、通信部202、表示部204、及び制御部205を備え、音声DBを備えない。図6の話者識別システムでも、話者識別装置200は、上記したように、一般家庭用のテレビ、またはパソコン用コンピュータ(PC)におけるモニタなどでもよい。図1と同様に、話者識別装置200は、一般家庭用のテレビである。

- [0080] また、サーバ装置210は、制御部211、通信部212、及び音声DB203を備える。サーバ装置210が置かれている場所に関しては、特に限定しない。ビッグデータを取り扱うデータセンターを管理又は運営する会社が保有するデータセンターに配置されていてもよく、各家庭に配置されていてもよい。
- [0081] 話者識別装置200の通信部202は、インターネットなどのネットワーク220を介して、サーバ装置210の通信部212と通信する。これにより、話者識別装置200の制御部205は、例えば生成した話者音声信号を、通信部202を介してサーバ装置210に送信することができる。サーバ装置210は、通信部212を介して、複数の話者識別装置200と接続されていてもよい。
- [0082] 図6の話者識別システムでは、図2に示される各機能は、サーバ装置210の制御部211と、話者識別装置200の制御部205との、いずれかに含まれていればよい。例えば音声処理部101は、音声取得部201が取得した話者の音声を処理するため、話者識別装置200の制御部205に含まれていてもよい。例えばデータベース管理部102は、音声DB203を管理するため、サーバ装置210の制御部211に含まれていてもよい。例えば表示制御部104は、表示部204を制御するため、話者識別装置200の制御部205に含まれていてもよい。
- [0083] 音声DB203は、サーバ装置210が複数の話者識別装置200と接続されている場合には、複数の話者識別装置200の各々に対応する音声情報

800（図3）をそれぞれ保存して管理してもよい。

- [0084] 図7は、図6の話者識別システムにおける動作の一例を示すシーケンス図である。なお、図7では、図2に示される各機能のうち、データベース管理部102及び識別処理部103は、サーバ装置210の制御部211に含まれ、音声処理部101及び表示制御部104は、話者識別装置200の制御部205に含まれている。また、ここでは、図6に示したサーバ装置210と話者識別装置200とを含む話者識別システムの動作の例を示すが、あくまで一例であり、本実施の形態を限定するものではない。
- [0085] まず、ステップS401にて、話者識別装置200における音声取得部201が、話者の音声を取得する。音声処理部101は、取得された話者の音声から特徴量を抽出し、抽出した特徴量を表す話者音声信号を生成する。本ステップS401は、図5に示すステップS301に相当する。
- [0086] ステップS401において、音声取得部201により取得された話者の音声に対して音声処理部101が特徴量抽出などの処理を行うタイミングに関しては限定しない。話者識別装置200であるテレビの電源がONになっている間、常に、音声取得部201が音声を取得して、音声処理部101が特徴量抽出などの処理を行ってもよい。また、音声処理部101がマジックワード（所定の単語）を検出した時に、その検出時点以降に音声取得部201が取得した音声に対して音声処理部101が特徴量抽出などの処理を開始してもよい。また、人が発声した音声と、話者の音声以外の環境音とを音声処理部101が識別して、人が発声した音声のみに対して、音声処理部101が特徴量抽出などの処理を行うようにしてもよい。
- [0087] 続いて、ステップS402にて、話者識別装置200における通信部202は、ネットワーク220を介して、音声処理部101により生成された話者音声信号をサーバ装置210に送信する。この時、一つのサーバ装置210に対して複数の話者識別装置200が接続されている場合、話者音声信号とともに、話者識別装置200を特定する識別情報を送信するようにしてもよい。

- [0088] 続いて、ステップS403にて、サーバ装置210の制御部211の識別処理部103は、データベース管理部102を介して、音声DB203に蓄積されている登録音声信号801を取得する。そして、識別処理部103は、取得した登録音声信号801と、ステップS402にて通信部212を介して話者識別装置200から取得した話者音声信号とを比較して、話者音声信号と一致する登録音声信号801（話者）を特定する。本ステップS403は、図5に示すステップS302に相当する。
- [0089] 続いて、ステップS404にて、制御部211の識別処理部103は、データベース管理部102を介して、特定された登録音声信号801に対応する登録アイコン803を抽出する。例えば図3において、登録音声信号801が「0001.wav」及び「0003.wav」の話者に対しては、それぞれアイコンA01、B05が登録アイコン803として登録されている。そこで、これらの話者については、識別処理部103は、各々の登録アイコン803を抽出すればよい。
- [0090] また、図3の例では、登録音声信号801が「0002.wav」の話者には登録アイコン803が登録されていない。この場合は、制御部211の識別処理部103は、予め作成された複数のアイコンから自動的に抽出してもよい。また、話者識別装置200から取得した話者音声信号が、登録音声信号801のいずれとも対応しない場合にも、同様に、制御部211の識別処理部103は、取得した話者音声信号から類推される適切なアイコンを予め作成された複数のアイコンから抽出してもよい。あるいは、識別処理部103は、話者識別装置200から取得した話者音声信号に対応する登録アイコン803が音声情報800に登録されていない場合には、話者音声信号から類推される適切なアイコンを作成してもよい。この点については、図1に示される構成の話者識別システムの場合でも同様である。
- [0091] 続いて、ステップS405にて、サーバ装置210の通信部212は、ネットワーク220を介して、ステップS404にて識別処理部103により抽出されたアイコンを話者識別装置200へ送信する。

- [0092] 続いて、ステップS406にて、話者識別装置200の制御部205の表示制御部104は、ステップS405にて送信されたアイコンを表示部204に表示させる。本ステップS406は、図5に示すステップS303に相当する。
- [0093] このとき、上述のように、音声処理部101は、音声取得部201により話者の音声が取得されている間だけ、その旨を表す取得信号を、話者ごとに表示制御部104に出力する。表示制御部104は、音声処理部101から取得信号が入力されている間、つまり特定された話者の音声を認識している間、アイコンを表示部204に表示させる。
- [0094] 表示制御部104は、音声取得部201が取得している話者の音声から、特定された話者を示す音声が所定時間途絶えたとき、つまり音声処理部101から取得信号が入力されなくなった時点から所定時間（この実施の形態1では、例えば10秒）が経過したときに、表示制御部104は、表示部204に表示されているアイコンを消去する。この場合、表示制御部104は、表示されているアイコンの透明度を徐々に増大させて、アイコンを表示部204からフェードアウトさせるようにしてもよい。
- [0095] 図8A～8Hは、それぞれ、図1又は図6に示される話者識別システムにおいて、表示制御部104によって表示部204に表示される登録アイコン803の具体的な表示例を示す図である。なお、あくまで、図8A～8Hに記載の表示構成は一例であり、図8A～8Hに示されている表示構成以外の表示構成を備えていてもよく、また一部の表示構成が欠けていてもよい。
- [0096] 図8Aでは、ステップS403にて特定された話者に対応する記号をアイコンとして、それぞれ色分けして話者識別装置200の表示部204の右下隅に表示している。図8Aの例では、アイコン911は丸の記号であり、アイコン912は四角の記号であり、アイコン913は三角の記号である。上記したように、ステップS406にて、表示制御部104は、これらの記号で表示されたアイコンを、話者が発話している間及びその後の所定時間だけ表示部204に表示する。このように表示することで、ユーザは、テレビ放

送の表示が過度に邪魔されることなく、話者識別の結果を確認できる。

- [0097] ここで、図8Aにて表示されているタイミングでは、アイコン911、アイコン912、アイコン913の3人の話者が同時に発話していることになる。例えば、あるタイミングで、アイコン912に対応する話者の発話が止まった時点から所定時間（この実施の形態1では、例えば10秒）が経過すると、表示制御部104は、アイコン912だけを消去する。その結果、表示部204には、アイコン911及びアイコン913のみが表示されている状態になる。
- [0098] その際、表示制御部104は、アイコン911が表示される位置を右にスライドさせ、アイコン911をアイコン913の直ぐ隣に表示するようにしてもよい。これにより、常に表示部204の右下隅にアイコンが集まることになり、テレビ放送の表示が過度に妨げられるのを抑制できる。
- [0099] なお、発話が止まった際に、表示制御部104は、アイコンを消去するのに代えて、アイコンの色を半透明にしてもよい。あるいは、発話が止まった際に、表示制御部104は、アイコンの大きさを小さく変化させてもよい。これらによっても、同様の効果が得られる。
- [0100] また、認識された話者に対応する複数のアイコンを一定時間表示するようにして、右もしくは左から複数の話者が発声した順に並べて表示してもよい。図8Aの例では、アイコン911、912、913の順で、もしくはアイコン913、912、911の順で、対応する話者が発声していることが示される。もちろん、上もしくは下から順に表示してもよい。これにより、発話のたびにアイコンの表示の順序が入れ替わる。したがって、ユーザへ発話を促すことができる。
- [0101] また、図8Aに示されるように、表示制御部104は、認識されている話者のうち、発話をしている話者を示すアイコンに対して、発話をしている間だけ、補足アイコン914を表示してもよい。図8Aの例では、補足アイコン914として、発話中の話者を示すアイコンを丸の形状で囲むアイコンが採用されており、現在、アイコン911に対応する話者が発話をしているこ

とが示されている。

- [0102] この場合、表示制御部104は、音声処理部101から話者ごとに出力される取得信号に基づき、補足アイコン914を表示する対象のアイコンを決定する。これにより、話者識別装置200の付近にいることが認識されている話者を示すアイコン912、913と、現在発話をしている話者を示すアイコン911とを、明確に区別して表示できる。
- [0103] 図8Bに示されるように、表示制御部104は、表示部204に表示するアイコンを、図8Aのような記号ではなくて、人の形を概略的に表したアイコン915～918としてもよい。上記したように、このアイコン915～918を、ユーザが選択してもよく、ユーザが作成してもよく、サーバ装置210の制御部211もしくは話者識別装置200の制御部205が選択するようにしてもよい。この場合でも、図8Aと同様に、表示制御部104は、補足アイコン914を表示部204に表示してもよい。
- [0104] また、表示制御部104は、アイコン付近またはアイコンの上に話者が発話した内容を都度表示するようにしてもよい。この場合には、表示制御部104は、アイコンを例えば半透明にして常に表示しておき、発話内容を話者が発話している間だけ表示するようにしてもよい。
- [0105] なお、図8Bでは、音声取得部201または音声処理部101は、指向性を制御する機能を有している。これにより、制御部205は、表示部204の前に位置する話者の存在する方向に指向性を持たせ、話者が位置する方向を検出することができる。そこで、表示制御部104は、図8Bに示されるように、検出した発話者の位置する方向に合わせてアイコンを表示させる位置を変化させてもよい。図8Bの例では、アイコン915、916に対応する話者は、表示部204の中心線から左側に位置し、アイコン917、918に対応する話者は、表示部204の中心線から右側に位置することが分かる。このように表示することで、ユーザは、話者識別の結果を確認しやすくなる。
- [0106] 図8Cに示されるように、複数の話者が一斉に発声した場合に、表示制御

部104は、音声DB203に新規に登録された話者に対して、仮で設定したアイコン921を大きく表示してもよい。

[0107] ここで、「新規に登録」は以下のように行われる。この話者が発話した時点では、この話者の登録音声信号801は、音声情報800に登録されていない。そこで、識別処理部103は、音声処理部101により生成された話者音声信号を、データベース管理部102を介して、登録音声信号801として、音声情報800に登録する。識別処理部103は、話者音声信号から話者の属性を判断する。識別処理部103は、その判断結果に基づき、アイコンを仮に設定して、データベース管理部102を介して、音声情報800に登録アイコン803として登録する。このようにして、未登録の話者は、音声DB203に新規に登録される。

[0108] これにより、ユーザは、新規の話者を確認することができる。また、新規の話者に対して、アイコンの選択または作成によって、仮に設定されたアイコンを好みのアイコンに変更するのを促すことができる。

[0109] 図8Dに示されるように、複数の話者が発話した場合に、表示制御部104は、発話時間または発話回数が最も多い話者に対応するアイコンを、大きく表示してもよい。この場合には、識別処理部103は、話者ごとに発話時間または発話回数をカウントし、データベース管理部102を介して、カウント値を音声DB203に蓄積する。表示制御部104は、データベース管理部102を介して、蓄積されたカウント値を音声DB203から取得する。

[0110] 図8Dの例では、アイコン922に対応する話者の発話時間または発話回数が最も多いことが分かる。これにより、話者の発話を促すことができる。話者の発話を促すことで、音声DB203に蓄積される音声情報800の量を増加することができる。その結果、より精度の高い話者認識が可能となる。

[0111] 図8Dのようにアイコン922を拡大表示するだけではなくて、表示制御部104は、図8Eに示されるように、発話量表示部931、932を表示

部204に表示してもよい。発話量表示部931, 932は、発話時間または発話回数に基づく発話量をバーで表示したものである。発話量は、発話時間が長くなると、または発話回数が多くなると、増大する。

[0112] 発話量表示部931は、例えば話者識別装置200を所有する家族単位での発話量を表す。発話量表示部932は、例えばサーバ装置210に接続されている全ての話者識別装置200における発話量の平均値を表す。なお、発話量表示部932は、サーバ装置210に接続されている全ての話者識別装置200のうちで、同一のテレビ放送番組が視聴されている話者識別装置200における発話量の平均値を表すようにしてもよい。

図8Eの場合には、例えば、発話量表示部931のレベルが発話量表示部932のレベルに比べて低い場合には、話者の発話が促されることとなる。また、サーバ装置210の制御部211は、発話量表示部931のレベルによって、現在視聴しているテレビ放送番組またはコマーシャルをユーザが熱心に視聴しているか否かのデータを収集することができる。

[0113] なお、図1の話者識別システムの場合には、表示制御部104は、発話量表示部931の表示のみを行うことができる。表示制御部104による発話量表示部932の表示は、図6の話者識別システムによって実現される。

[0114] 図8Fに示されるように、表示制御部104は、アイコン911～914を表示部204に表示する際に、テレビ放送番組を表示するメイン表示領域941を表示部204の表示画面全体から縮小してもよい。表示制御部104は、メイン表示領域941の外側に、サブ表示領域942を設けて、このサブ表示領域942にアイコン911～914を表示してもよい。これによって、アイコン911～914の表示によりテレビ放送番組の視聴が過度に妨げられる事態を抑制することができる。

[0115] 図8A～8Fでは、複数のアイコンが表示部204に表示されているが、図8G, 8Hに示されるように、1個のアイコンが表示部204に表示される場合もある。例えば図8Aにおいて、アイコン913に対応する話者のみが発話を続け、アイコン911, 912に対応する話者の発話が停止した場

合、その停止時点から所定時間（この実施の形態1では、例えば10秒）が経過すると、表示制御部104は、図8Gに示されるように、アイコン913のみを表示部204に表示し、他のアイコンを消去する。

[0116] 例えば図8Bにおいて、アイコン915に対応する話者のみが発話を続け、アイコン916～918に対応する話者の発話が停止した場合、その停止

時点から所定時間（この実施の形態1では、例えば10秒）が経過すると、表示制御部104は、図8Hに示されるように、アイコン915のみを表示部204に表示し、他のアイコンを消去する。

[0117] 以上説明されたように、本実施の形態1における話者識別システムによれば、表示部204が本来表示したいコンテンツ（例えば表示部204がテレビの表示画面であれば、テレビ放送番組）の表示の妨げになることを抑制しつつも、ユーザへ話者の識別結果を明確に表示することができる。

[0118] なお、あくまで、図1、図6に記載の構成は、本実施の形態1の話者識別システムの一例であり、図1、図6に示されている構成以外の構成を備えていてもよく、また一部の構成が欠けていてもよい。また、図1、図6のいずれを採用してもよいし、図示した以外の装置を本実施の形態1の話者識別システムに採用することもできる。

[0119] （実施の形態2）

以下、実施の形態2における話者識別システムを説明する。なお、本実施の形態2では一部、実施の形態1と同様の説明は省略している。また、実施の形態2の技術を実施の形態1に記載の技術と組み合わせることも可能である。

[0120] 本実施の形態2における話者識別システムの構成は、図1または図6に示される実施の形態1の話者識別システムと同様であるので、詳細な説明を省略する。実施の形態2では、実施の形態1と同一構成については、同じ符号を用いて説明する。ただし、本実施の形態2においては、図1または図6に示される入力受付部206が必須の構成となる。

[0121] 図9は、実施の形態2における、図1に示される話者識別装置200の制

御部205の機能を示すブロック図である。図2に示される実施の形態1と異なるのは、修正制御部105を備えていることである。この修正制御部105により、識別処理部103が抽出したアイコンが誤っている場合に、ユーザが修正を行い、音声DB203の情報を更新することが可能となる。このような構成により、実施の形態2では、識別処理部103により識別された情報を修正することが簡易に行えるようになる。修正制御部105の具体的な動作については、図10を参照して、次に説明する。

- [0122] 図10は、実施の形態2の話者識別システムの図1に示される話者識別装置200における処理を示すフローチャートである。ステップS301～S303は、図5のステップS301～S303と同様である。
- [0123] ステップS303に続いて、ステップS304では、修正制御部105は、話者に対応するアイコンについて、ユーザからの修正指示を受け付ける。ユーザは、入力受付部206を用いて、修正指示を行う。修正制御部105は、ユーザの修正指示の内容にしたがって、データベース管理部102を介して、音声DB203の内容を更新する。
- [0124] ここで、ステップS304にて、修正制御部105は、ステップS303にてアイコンが表示されている間のみ、ユーザからの修正指示を受け付けるよう制御してもよい。これにより、修正を意図していないタイミングで誤って修正指示を受け付けてしまうことを低減できる。また、この場合、修正制御部105は、表示制御部104を通じて、アイコンが表示されている間に、ユーザからの修正指示を受け付けることが可能である旨の表示を表示部204に表示させてもよい。これにより、ユーザは修正機能があることを把握できる。
- [0125] 図11A, 11Bは、図10のステップS304における修正指示にユーザが用いる入力受付部206の一例を示す図である。図11A, 11Bを参照して、図10のステップS304において、ユーザが入力受付部206を用いて、アイコンについての修正指示を行う方法が説明される。図11Aは、入力受付部206の一例であるリモートコントローラ（以下、「リモコン

」という) 1001を示す。図11Bは、入力受付部206の別の例であるリモコン1002を示す。

- [0126] 図10のステップS303において、アイコンが誤って表示部204に表示された場合、ユーザは、例えばリモコン1001を使用して修正指示を送る(図10のステップS304)。アイコンが誤って表示部204に表示された場合とは、例えば、図8Bに示すような表示例では、アイコン915に対応する話者の発話中に、補足アイコン914が、誤って図8Bのように別の話者を示すアイコン916上に表示されていることなどである。
- [0127] ここで、図11Aのリモコン1001において、カラーボタン1003の各ボタンは、予め各アイコンと対応付けられている。例えば図8Bにおいて、アイコン915は「青」ボタンと、アイコン916は「赤」ボタンと、アイコン917は「緑」ボタンと、アイコン918は「黄」ボタンと、それぞれ対応付けられているとする。この場合において、各アイコン915～918には、それぞれ対応付けられている色が、ユーザに分かるように重畳されて表示されていることが望ましい。
- [0128] なお、あらかじめ話者とリモコンのカラーボタン1003の各ボタンとを対応付けておかなくてもよい。例えば、カラーボタン1003のうち任意のカラーボタンを押して修正することでもよい。また、アイコンが表示されている位置における左から順に、「青」、「赤」、「緑」、「黄」のボタンが対応付けられてもよい。
- [0129] 図10のステップS304における修正指示として、アイコン915に対応する話者が、リモコン1001の「青」ボタンを押しながら発話する。すると補足アイコン914がアイコン915上に移動し、登録されているアイコンに対して正しい話者画像を表示することができる。これにより、誤って識別結果が表示された場合であっても、話者に対応付けられたりモコン1001のカラーボタン1003を選択し、修正指示を送ることで簡易に修正可能となる。
- [0130] また、図11Aに示されるリモコン1001に代えて、図11Bに示され

るリモコン1002を用いてもよい。図11Bに示されるリモコン1002では、同様に、各アイコンと、リモコン1002の数字ボタンとが対応付けられていればよい。この場合には、ユーザは、リモコン1002の対応する数字ボタンを押しながら発話をすることで、修正指示を送ることができる。

- [0131] なお、ユーザが修正指示を送る方法は、上記に限られない。例えば、対応するリモコンのボタンを押したときに、表示制御部104は、修正可能な設定ページに表示部204の表示を切り替えるようにしてもよい。
- [0132] 図10に戻って、ステップS304で行われる音声DB203の内容の更新が説明される。アイコン915に対応する話者の発話中に、補足アイコン914が、誤って図8Bのように別の話者を示すアイコン916上に表示された原因是、アイコン915に対応する話者の登録音声信号801(図3)が、特徴ベクトルを正確に表していないことにある可能性が高い。
- [0133] そこで、アイコン915に対応する話者が、リモコン1001の「青」ボタンを押しながら発話をすると、音声処理部101は、音声取得部201により取得された音声から特徴ベクトルを生成し、生成した特徴ベクトルを表す話者音声信号を生成する。データベース管理部102は、識別処理部103を介して生成された話者音声信号を受け取り、音声DB203のアイコン915に対応する話者の登録音声信号801を、生成された話者音声信号に書き換える。
- [0134] ステップS304で行われる音声DB203の内容の更新の別の例が、図3、図8B及び図8Hを参照して説明される。
- [0135] リモコン1001のカラーボタン1003が、図3の3人の話者に対応付けられている。例えば、登録音声信号801が「0001. wav」の話者は「青」ボタンに対応付けられ、登録音声信号801が「0002. wav」の話者は「赤」ボタンに対応付けられ、登録音声信号801が「0003. wav」の話者は「緑」ボタンに対応付けられている。また、図3の登録アイコン「A01」は、図8Bのアイコン916である。また、図3の登録アイコン「B05」は、図8B, 8Hのアイコン915である。

- [0136] この場合において、登録音声信号801が「0001. w a v」の話者が発話中であるのに、図8Hに示されるように、アイコン915が表示部204に表示される。この原因は、図3の登録音声信号「0001. w a v」が、特徴ベクトルを正確に表していないことにある可能性が高い。
- [0137] そこで、登録音声信号801が「0001. w a v」の話者（つまりアイコン916に対応する話者）が、リモコン1001の「青」ボタンを押しながら発話する。音声処理部101は、音声取得部201により取得された音声から特徴ベクトルを生成し、生成した特徴ベクトルを表す話者音声信号を生成する。データベース管理部102は、識別処理部103を介して生成された話者音声信号を受け取り、音声DB203の登録音声信号「0001. w a v」を、生成された話者音声信号に書き換える。
- [0138] 図12は、実施の形態2の図6に示される話者識別システムにおける動作の一例を示すシーケンス図である。なお、図12では、図9に示される各機能のうち、データベース管理部102及び識別処理部103は、サーバ装置210の制御部211に含まれ、音声処理部101、表示制御部104、及び修正制御部105は、話者識別装置200の制御部205に含まれている。また、ここでは、図6に示したサーバ装置210と話者識別装置200とを含む話者識別システムの動作の例を示すが、あくまで一例であり、本実施の形態を限定するものではない。
- [0139] ステップS401～S406は、図7に示すステップS401～S406と同様なので、詳細な説明を省略する。
- [0140] ステップS406に続いて、ステップS407にて、修正制御部105は、入力受付部206を用いて行われるアイコンについてのユーザからの修正指示を、受け付ける。本ステップS407は、図10に示すステップS304の一部に相当する。すなわち、ユーザによる修正指示は、図10のステップS304と同様に行われる。
- [0141] ステップS407に続いて、ステップS408にて、話者識別装置200の通信部202は、修正制御部105が受け付けたユーザからの修正指示を

、サーバ装置210へ送信する。

[0142] 続いて、ステップS409において、サーバ装置210のデータベース管理部102は、ユーザの修正指示に基づき、音声DB203の内容を更新する。本ステップS409は、図10に示すステップS304の一部に相当する。すなわち、音声DB203の更新は、図10のステップS304と同様に行われる。

[0143] 以上説明されたように、本実施の形態2の話者識別システムによれば、話者の識別結果として表示部204に表示されたアイコンが、誤った識別により異なるアイコンであった際にも、ユーザが煩わしい操作をすることなく、修正を指示することができる。もし、話者識別の結果に誤検出があり、その結果がそのまま表示されていると、ユーザにとって不快感が生じると考えられる。しかし、本実施の形態2では、そのようなユーザの不快感を解消できる。さらに、ユーザによる音声DB203の修正を促進することにもなる。したがって、より高精度で、家族の音声DB203を構築することができる。

[0144] (その他)

(1) 上記実施の形態2において、図9の修正制御部105は、ユーザによる入力受付部206を用いた修正指示を、誤ったアイコンが表示部204に表示されている間のみ、受け付けるようにしてもよい。例えば図10のステップS304において図8Bを用いて説明した例では、補足アイコン914が誤って表示部204に表示されている間のみ、修正制御部105は、ユーザによるリモコン1001を用いた修正指示を受け付ける。例えば図10のステップS304において図3を用いて説明した例では、アイコン「B05」が誤って表示部204に表示されている間のみ、修正制御部105は、ユーザによるリモコン1001を用いた修正指示を受け付ける。

[0145] この場合、表示制御部104は、表示部204に表示中のアイコンに関する情報を、修正制御部105に出力するようにしてもよい。修正制御部105は、表示制御部104から入力される、表示部204に表示中のアイコン

に関する情報に基づき、ユーザによる入力受付部206を用いた修正指示が、表示部204に表示中のアイコンの修正指示であるか否かを判断してもよい。修正制御部105は、ユーザによる入力受付部206を用いた修正指示が、表示部204に表示中のアイコンの修正指示である場合にのみ、修正指示を受け付けるようにしてもよい。

[0146] このように、ユーザによる入力受付部206を用いた修正指示の受付可能な期間を限定することによって、表示部204に表示されていないアイコンに関する修正指示、又はユーザによる誤った修正指示を受け付けるような事態を避けることができる。

[0147] (2) 上記実施の形態1, 2では、図8A～図8Fに示されるように、アイコンを表示する表示部204は、話者識別装置200であるテレビの表示画面としている。しかし、本開示は、これに限られない。例えば、表示部204は、タブレット端末又はいわゆるスマートフォンなどの携帯装置の表示画面としてもよい。表示制御部104は、通信部202を通して、携帯装置の表示画面にアイコンを表示してもよい。

[0148] (3) 上記実施の形態1, 2において、連続して音声処理部101から入力される2つの話者音声信号が、図3の音声情報800の登録音声信号801が「0001. wav」及び「0003. wav」の話者に一致すると識別処理部103が判断すると、識別処理部103は、図3のユーザ情報802に基づき、父親と子供と一緒にテレビ放送番組を視聴していると判断できる。

[0149] あるいは、連続して音声処理部101から入力される2つの話者音声信号が、図3の音声情報800の登録音声信号801が「0001. wav」及び「0002. wav」の話者に一致すると識別処理部103が判断すると、識別処理部103は、図3のユーザ情報802に基づき、大人のみがテレビ放送番組を視聴していると判断できる。

[0150] そこで、表示制御部104は、表示部204を用いて、識別処理部103による視聴者の判断結果に基づき、視聴者に適合するコンテンツ（例えばテ

レビ放送番組）を、視聴者に対して推奨するようにしてよい。

[0151] (提供するサービスの全体像)

図13Aは、上記実施の形態1，2における図6に示される話者識別システムの全体像を示す図である。

[0152] グループ1100は、例えば企業、団体、家庭等であり、その規模を問わない。グループ1100には、複数の機器1101（例えば機器A、機器B）およびホームゲートウェイ1102が存在する。複数の機器1101は、インターネットに接続可能な機器（例えば、スマートフォン、パーソナルコンピュータ、テレビ等）を含む。また、複数の機器1101は、それ自身ではインターネットに接続不可能な機器（例えば、照明機器、洗濯機、冷蔵庫等）を含む。複数の機器1101は、それ自身ではインターネットに接続不可能な機器であっても、ホームゲートウェイ1102を介してインターネットに接続可能となる機器を含んでもよい。また、グループ1100には、複数の機器1101を使用するユーザ1010が存在する。

[0153] データセンタ運営会社1110には、クラウドサーバ1111が存在する。クラウドサーバ1111は、インターネットを介して様々な機器と連携する仮想化サーバである。クラウドサーバ1111は、主に通常のデータベース管理ツール等で扱うことが困難な巨大なデータ（ビッグデータ）等を管理する。データセンタ運営会社1110は、データ管理及びクラウドサーバ1111の管理を行うデータセンタの運営等を行っている。データセンタ運営会社1110が行っている役務については詳細を後述する。

[0154] ここで、データセンタ運営会社1110は、データ管理及びクラウドサーバ1111の管理を行うデータセンタの運営等のみを行っている会社に限らない。

[0155] 図13B、図13Cは、データセンタ運営会社1110の一例を示す図である。例えば複数の機器1101のうちの一つの機器を開発又は製造している機器メーカーが、併せてデータ管理及びクラウドサーバ1111の管理等を行っている場合は、機器メーカーがデータセンタ運営会社1110に該当する

(図13B)。また、データセンタ運営会社1110は、一つの会社に限らない。例えば機器メーカー及び他の管理会社が共同もしくは分担してデータ管理及びクラウドサーバ1111の管理等を行っている場合は、両者もしくはいずれか一方がデータセンタ運営会社1110に該当するものとする(図13C)。

- [0156] サービスプロバイダ1120は、サーバ1121を保有している。ここで言うサーバ1121とは、その規模は問わず例えば、個人用パーソナルコンピュータ内のメモリ等も含む。また、サービスプロバイダ1120がサーバ1121を保有していない場合もある。この場合、サービスプロバイダ1120は、サーバ1121の機能を果たす別の装置を保有する。
- [0157] なお、上記話者識別システムにおいて、ホームゲートウェイ1102は必須ではない。ホームゲートウェイ1102は、機器1101をインターネットに接続可能にするための装置である。そこで、例えば、グループ1100内のあらゆる機器1101がインターネットに接続されている場合のように、それ自身ではインターネットに接続不可能な機器が存在しない場合は、ホームゲートウェイ1102は不要となる。
- [0158] 次に、図13Aを用いて上記話者識別システムにおける情報の流れを説明する。
- [0159] まず、グループ1100の機器1101、例えば機器A又は機器Bは、各操作ログ情報をデータセンタ運営会社1110のクラウドサーバ1111に送信する。クラウドサーバ1111は、機器A又は機器Bの操作ログ情報を集積する(図13Aの矢印(a))。ここで、操作ログ情報とは複数の機器1101の、例えば運転状況や動作日時等を示す情報である。例えば、テレビの視聴履歴、レコーダの録画予約情報、洗濯機の運転日時及び洗濯物の量、冷蔵庫の開閉日時及び開閉回数などである。操作ログ情報は、これらのものに限らずあらゆる機器1101から取得が可能なすべての情報をいう。
- [0160] 操作ログ情報は、インターネットを介して複数の機器1101自体から直接クラウドサーバ1111に提供される場合もある。また、複数の機器11

01から一旦ホームゲートウェイ1102に操作ログ情報が集積され、ホームゲートウェイ1102からクラウドサーバ1111に提供されてもよい。

- [0161] 次に、データセンタ運営会社1110のクラウドサーバ1111は、集積した操作ログ情報を一定の単位でサービスプロバイダ1120に提供する。ここで、「一定の単位」は、データセンタ運営会社1110が集積した情報を整理してサービスプロバイダ1120に提供することのできる単位でもよく、サービスプロバイダ1120が要求した単位でもよい。「一定の単位」と記載したが、情報量は一定でなくてもよい。例えば、状況に応じて提供する情報量が変化する場合もある。前記操作ログ情報は、必要に応じてサービスプロバイダ1120が保有するサーバ1121に保存される（図13Aの矢印（b））。
- [0162] そして、サービスプロバイダ1120は、操作ログ情報をユーザに提供するサービスに適合する情報に整理し、ユーザに提供する。提供対象のユーザは、複数の機器1101を使用するユーザ1010でもよく、外部のユーザ1020でもよい。ユーザへのサービス提供方法は、例えば、サービスプロバイダ1120から直接ユーザ1010、1020へ提供されてもよい（図13Aの矢印（f）、（e））。また、ユーザへのサービス提供方法は、例えば、データセンタ運営会社1110のクラウドサーバ1111を再度経由して、ユーザ1010に提供されてもよい（図13Aの矢印（c）、（d））。また、データセンタ運営会社1110のクラウドサーバ1111が操作ログ情報をユーザに提供するサービスに適合する情報に整理し、サービスプロバイダ1120に提供してもよい。
- [0163] なお、ユーザ1010とユーザ1020とは、別でも同一でもよい。
- [0164] 上記態様において説明された技術は、例えば、以下のクラウドサービスの類型において実現されうる。しかし、上記態様において説明された技術が実現される類型はこれに限られるものでない。
- [0165] （サービスの類型1：自社データセンタ型）  
図14は、サービスの類型1（自社データセンタ型）を示す。本類型は、

サービスプロバイダ1120がグループ1100から情報を取得し、ユーザに対してサービスを提供する類型である。本類型では、サービスプロバイダ1120が、データセンタ運営会社の機能を有している。即ち、サービスプロバイダ1120が、ビッグデータの管理をするクラウドサーバ1111を保有している。従って、データセンタ運営会社は存在しない。

[0166] 本類型では、サービスプロバイダ1120は、データセンタ（クラウドサーバ1111）を運営、管理している（1203）。また、サービスプロバイダ1120は、OS（1202）及びアプリケーション（1201）を管理する。サービスプロバイダ1120は、サービスプロバイダ1120が管理するOS（1202）及びアプリケーション（1201）を用いてサービス提供を行う（1204）。

[0167] (サービスの類型2：IaaS利用型)

図15は、サービスの類型2（IaaS利用型）を示す。ここでIaaSとはインフラストラクチャー・アズ・ア・サービスの略であり、コンピュータシステムを構築および稼動させるための基盤そのものを、インターネット経由のサービスとして提供するクラウドサービス提供モデルである。

[0168] 本類型では、データセンタ運営会社1110がデータセンタ（クラウドサーバ1111）を運営、管理している（1203）。また、サービスプロバイダ1120は、OS（1202）及びアプリケーション（1201）を管理する。サービスプロバイダ1120は、サービスプロバイダ1120が管理するOS（1202）及びアプリケーション（1201）を用いてサービス提供を行う（1204）。

[0169] (サービスの類型3：PaaS利用型)

図16は、サービスの類型3（PaaS利用型）を示す。ここでPaaSとはプラットフォーム・アズ・ア・サービスの略であり、ソフトウェアを構築および稼動させるための土台となるプラットフォームを、インターネット経由のサービスとして提供するクラウドサービス提供モデルである。

[0170] 本類型では、データセンタ運営会社1110は、OS（1202）を管理

し、データセンタ（クラウドサーバ1111）を運営、管理している（1203）。また、サービスプロバイダ1120は、アプリケーション（1201）を管理する。サービスプロバイダ1120は、データセンタ運営会社1110が管理するOS（1202）及びサービスプロバイダ1120が管理するアプリケーション（1201）を用いてサービス提供を行う（1204）。

[0171] (サービスの類型4：SaaS利用型)

図17は、サービスの類型4（SaaS利用型）を示す。ここでSaaSとはソフトウェア・アズ・ア・サービスの略である。例えばデータセンタ（クラウドサーバ）を保有しているプラットフォーム提供者が提供するアプリケーションを、データセンタ（クラウドサーバ）を保有していない会社・個人（利用者）がインターネットなどのネットワーク経由で使用できる機能を有するクラウドサービス提供モデルである。

[0172] 本類型では、データセンタ運営会社1110は、アプリケーション（1201）を管理し、OS（1202）を管理し、データセンタ（クラウドサーバ1111）を運営、管理している（1203）。また、サービスプロバイダ1120は、データセンタ運営会社1110が管理するOS（1202）及びアプリケーション（1201）を用いてサービス提供を行う（1204）。

[0173] 以上いずれの類型においても、サービスプロバイダ1120がサービス提供行為を行ったものとする。また例えば、サービスプロバイダ1120若しくはデータセンタ運営会社1110は、OS、アプリケーション若しくはビックデータのデータベース等を自ら開発してもよいし、また、第三者に外注して開発させてもよい。

### 産業上の利用可能性

[0174] 本開示にかかる話者識別方法、話者識別装置及び話者識別システムは、不特定多数の発話者がいる環境において、話者識別を使用する際に、識別されている話者を表す話者画像を容易に表示する方法、装置及びシステムとして

有用である。

## 請求の範囲

- [請求項1] 表示部の周辺に位置する話者の音声を識別し識別結果を前記表示部に表示する話者識別システムにおける話者識別方法であって、  
前記話者識別システムは、複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを保存するデータベースを含み、  
前記表示部の周辺に位置する話者の音声を取得し、  
前記取得された話者の音声から話者音声信号を生成し、  
前記データベースに保存されている複数の登録音声信号のうち、前記生成された話者音声信号に対応する登録音声信号を識別し、  
前記識別された登録音声信号に対応付けて前記データベースに保存されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声が取得されている間、前記表示部に表示させる、  
話者識別方法。
- [請求項2] 前記話者音声信号を生成する元となった前記話者の音声が取得されなくなった時点から所定時間が経過すると、前記表示されている前記話者画像を前記表示部から消去する、  
請求項1に記載の話者識別方法。
- [請求項3] 前記データベースは、前記複数の音声登録信号として、第1話者の音声に基づき生成された第1登録音声信号と、第2話者の音声に基づき生成された第2登録音声信号とを保存し、かつ、前記第1登録音声信号に対応付けられ前記第1話者を表す第1話者画像と、前記第2登録音声信号に対応付けられ前記第2話者を表す第2話者画像とを保存し、  
前記第1話者の音声が取得されると、第1話者音声信号が生成され、

前記生成された第1話者音声信号が前記第1登録音声信号に対応すると識別されると、少なくとも前記第1話者の音声が取得されている間、前記第1話者画像が前記表示部に表示され、

前記第1話者画像が前記表示部に表示されているときに、前記第2話者の音声が取得されると、第2話者音声信号が生成され、

前記生成された第2話者音声信号が前記第2登録音声信号に対応すると識別されると、少なくとも前記第2話者の音声が取得されている間、前記第1話者画像に加えて、前記第2話者画像が前記表示部に表示される、

請求項1または2に記載の話者識別方法。

[請求項4] 前記第1話者の音声と前記第2話者の音声とが取得された順に、前記第1話者画像と前記第2話者画像とが、前記表示部に並べられて表示される、

請求項3に記載の話者識別方法。

[請求項5] 前記第1話者画像と前記第2話者画像とのうち、前記データベースに後で登録された方の話者画像は、前記データベースに先に登録された方の話者画像とは異なる態様で前記表示部に表示される、

請求項3に記載の話者識別方法。

[請求項6] 前記第1話者の発話回数と、前記第2話者の発話回数とがカウントされ、

前記第1話者画像と前記第2話者画像とは、前記カウントされた発話回数の多い順に、前記表示部に並べられて表示される、

請求項3に記載の話者識別方法。

[請求項7] 前記話者画像についての話者からの修正指示が受け付けられると、前記修正を指示した話者の音声を新たに取得し、

前記新たに取得された話者の音声から新たに話者音声信号を生成し、

前記データベースに保存されている、前記修正指示が行われた前記

話者画像に対応付けられている前記登録音声信号を、前記新たに生成された話者音声信号に書き換える、

請求項 1～6 のいずれか 1 項に記載の話者識別方法。

[請求項8] 前記話者からの修正指示は、前記表示部に表示されている前記話者画像について受け付けられ、前記表示部に表示されていない前記話者画像については受け付けられない、

請求項 7 に記載の話者識別方法。

[請求項9] 前記生成された話者音声信号から前記話者の属性を判別し、  
前記判別された前記話者の属性に基づき前記話者画像を作成し、  
前記生成された話者音声信号と前記判別された前記話者の属性と前記作成された前記話者画像とを、互いに対応付けて前記データベースに保存し、前記データベースには、前記生成された話者音声信号は、前記登録音声信号として保存される、

請求項 1～8 のいずれか 1 項に記載の話者識別方法。

[請求項10] 表示部と、  
前記表示部の周辺に位置する話者の音声を取得する音声取得部と、  
前記取得された話者の音声から話者音声信号を生成する音声処理部と、

複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを記憶するデータベースと、

前記データベースに記憶されている前記複数の登録音声信号のうち、前記生成された前記話者音声信号に対応する登録音声信号を識別する識別処理部と、

前記識別された前記登録音声信号に対応付けて前記データベースに記憶されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記表示部に表示する表示制御部と、

を備える話者識別装置。

[請求項11] 表示部と、

前記表示部の周辺に位置する話者の音声を取得する音声取得部と、

前記取得された話者の音声から話者音声信号を生成する音声処理部

と、

ネットワークを介して外部のサーバ装置と通信を行う通信部と、

前記表示部を制御する表示制御部と、

を備え、

前記通信部は、前記生成された話者音声信号を前記サーバ装置に送信し、かつ、前記サーバ装置から、前記話者音声信号を基に識別された前記話者を表す話者画像を受信し、

前記表示制御部は、前記受信された話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記表示部に表示させる、

話者識別装置。

[請求項12] 表示部の周辺に位置する話者の音声を取得する音声取得部と、

前記取得された話者の音声から話者音声信号を生成する音声処理部

と、

複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを記憶する記憶部と、

前記複数の登録音声信号のうち前記生成された話者音声信号に対応する登録音声信号を識別する識別処理部と、

前記識別された前記登録音声信号に対応付けて前記記憶部に記憶されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記表示部に表示させる表示制御部と、

を備える話者識別システム。

補正された請求の範囲  
[2014年10月31日(31.10.2014)国際事務局受理]

[請求項1] (補正後)

表示部の周辺に位置する話者の音声を識別し識別結果を前記表示部に表示する話者識別システムにおける話者識別方法であって、

前記話者識別システムは、複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを保存するデータベースを含み、

前記表示部の周辺に位置する話者の音声を取得し、

前記取得された話者の音声から前記話者の位置を表す位置情報を検出し、

前記取得された話者の音声から話者音声信号を生成し、

前記データベースに保存されている複数の登録音声信号のうち、前記生成された話者音声信号に対応する登録音声信号を識別し、

前記識別された登録音声信号に対応付けて前記データベースに保存されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声が取得されている間、前記検出された位置情報により表される前記話者の位置に対応した前記表示部上の位置に表示させる、

話者識別方法。

[請求項2] (補正後)

前記位置情報は、前記表示部を基準にしたときの水平方向の前記話者の位置を表し、

前記話者画像を、前記表示部上の前記位置情報により表される水平方向の前記話者の位置に表示する、

請求項1に記載の話者識別方法。

[請求項3] (補正後)

前記話者画像を、

前記話者の音声が取得されている間は第1の表示形式で表示し、  
前記話者の音声が取得された後に、前記話者の音声が取得されていない間は、第2の表示形式で表示する、  
請求項1に記載の話者識別方法。

[請求項4] (補正後)

前記生成された話者音声信号から、前記話者の発話内容を検出し、  
前記検出した発話内容を、前記話者音声信号を生成する元となった  
前記話者の音声が取得されている間、前記話者画像の近傍に表示する

、  
請求項1に記載の話者識別方法。

[請求項5] (補正後)

前記話者音声信号を生成する元となった前記話者の音声が取得されなくなった時点から所定時間が経過すると、前記表示されている前記話者画像を前記表示部から消去する、  
請求項1～4のいずれか1項に記載の話者識別方法。

[請求項6] (補正後)

前記データベースは、前記複数の音声登録信号として、第1話者の音声に基づき生成された第1登録音声信号と、第2話者の音声に基づき生成された第2登録音声信号とを保存し、かつ、前記第1登録音声信号に対応付けられ前記第1話者を表す第1話者画像と、前記第2登録音声信号に対応付けられ前記第2話者を表す第2話者画像とを保存し、

前記第1話者の音声が取得されると、第1話者音声信号が生成され

、  
前記生成された第1話者音声信号が前記第1登録音声信号に対応すると識別されると、少なくとも前記第1話者の音声が取得されている間、前記第1話者画像が前記表示部に表示され、

前記第1話者画像が前記表示部に表示されているときに、前記第2

話者の音声が取得されると、第2話者音声信号が生成され、

前記生成された第2話者音声信号が前記第2登録音声信号に対応すると識別されると、少なくとも前記第2話者の音声が取得されている間、前記第1話者画像に加えて、前記第2話者画像が前記表示部に表示される。

請求項1～5のいずれか1項に記載の話者識別方法。

[請求項7] (補正後)

前記第1話者の音声と前記第2話者の音声とが取得された順に、前記第1話者画像と前記第2話者画像とが、前記表示部に並べられて表示される。

請求項6に記載の話者識別方法。

[請求項8] (補正後)

前記第1話者画像と前記第2話者画像とのうち、前記データベースに後で登録された方の話者画像は、前記データベースに先に登録された方の話者画像とは異なる態様で前記表示部に表示される。

請求項6に記載の話者識別方法。

[請求項9] (補正後)

前記第1話者の発話回数と、前記第2話者の発話回数とがカウントされ、

前記第1話者画像と前記第2話者画像とは、前記カウントされた発話回数の多い順に、前記表示部に並べられて表示される。

請求項6に記載の話者識別方法。

[請求項10] (補正後)

前記話者画像についての話者からの修正指示が受け付けられると、前記修正を指示した話者の音声を新たに取得し、

前記新たに取得された話者の音声から新たに話者音声信号を生成し、

前記データベースに保存されている、前記修正指示が行われた前記

話者画像に対応付けられている前記登録音声信号を、前記新たに生成された話者音声信号に書き換える。

請求項 1～9 のいずれか 1 項に記載の話者識別方法。

[請求項11] (補正後)

前記話者からの修正指示は、前記表示部に表示されている前記話者画像について受け付けられ、前記表示部に表示されていない前記話者画像については受け付けられない、

請求項 10 に記載の話者識別方法。

[請求項12] (補正後)

前記生成された話者音声信号から前記話者の属性を判別し、前記判別された前記話者の属性に基づき前記話者画像を作成し、前記生成された話者音声信号と前記判別された前記話者の属性と前記作成された前記話者画像とを、互いに対応付けて前記データベースに保存し、前記データベースには、前記生成された話者音声信号は、前記登録音声信号として保存される、

請求項 1～11 のいずれか 1 項に記載の話者識別方法。

[請求項13] (追加)

表示部と、

前記表示部の周辺に位置する話者の音声を取得する音声取得部と、前記取得された話者の音声から話者音声信号を生成する音声処理部と、

複数の話者の音声に基づきそれぞれ生成された複数の登録音声信号と、前記複数の登録音声信号にそれぞれ対応付けられ前記複数の話者をそれぞれ表す複数の話者画像とを記憶するデータベースと、

前記データベースに記憶されている前記複数の登録音声信号のうち、前記生成された前記話者音声信号に対応する登録音声信号を識別する識別処理部と、

前記取得された話者の音声から、前記話者の位置を表す位置情報を

検出する位置情報検出部と、

前記識別された前記登録音声信号に対応付けて前記データベースに記憶されている前記話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記検出された位置情報により表される前記話者の位置に対応した前記表示部上の位置に表示する表示制御部と、

を備える話者識別装置。

[請求項14] (追加)

前記位置情報は、前記表示部を基準にしたときの水平方向の前記話者の位置を表し、

前記表示制御部は、前記話者画像を、前記表示部上の前記位置情報により表される水平方向の前記話者の位置に表示する、

請求項13に記載の話者識別装置。

[請求項15] (追加)

前記音声取得部と前記音声処理部とのいずれかは、取得する音声の指向性を制御する機能を有し、

前記位置情報検出部は、前記音声取得部もしくは前記音声処理部が、取得する音声の指向性を制御することで、前記位置情報を検出する、

請求項13に記載の話者識別装置。

[請求項16] (追加)

表示部と、

前記表示部の周辺に位置する話者の音声を取得する音声取得部と、前記取得された話者の音声から話者音声信号を生成する音声処理部と、

前記取得された話者の音声から、前記話者の位置を表す位置情報を検出する位置情報検出部と、

ネットワークを介して外部のサーバ装置と通信を行う通信部と、

前記表示部を制御する表示制御部と、

を備え、

前記通信部は、前記生成された話者音声信号を前記サーバ装置に送信し、かつ、前記サーバ装置から、前記話者音声信号を基に識別された前記話者を表す話者画像を受信し、

前記表示制御部は、前記受信された話者画像を、少なくとも前記話者音声信号を生成する元となった前記話者の音声を前記音声取得部が取得している間、前記検出された位置情報により表される前記話者の位置に対応した前記表示部上の位置に表示させる、

話者識別装置。

[請求項17] (追加)

前記位置情報は、前記表示部を基準にしたときの水平方向の前記話者の位置を表し、

前記表示制御部は、前記話者画像を、前記表示部上の前記位置情報により表される水平方向の前記話者の位置に表示する、

請求項16に記載の話者識別装置。

[請求項18] (追加)

前記音声取得部と前記音声処理部とのいずれかは、取得する音声の指向性を制御する機能を有し、

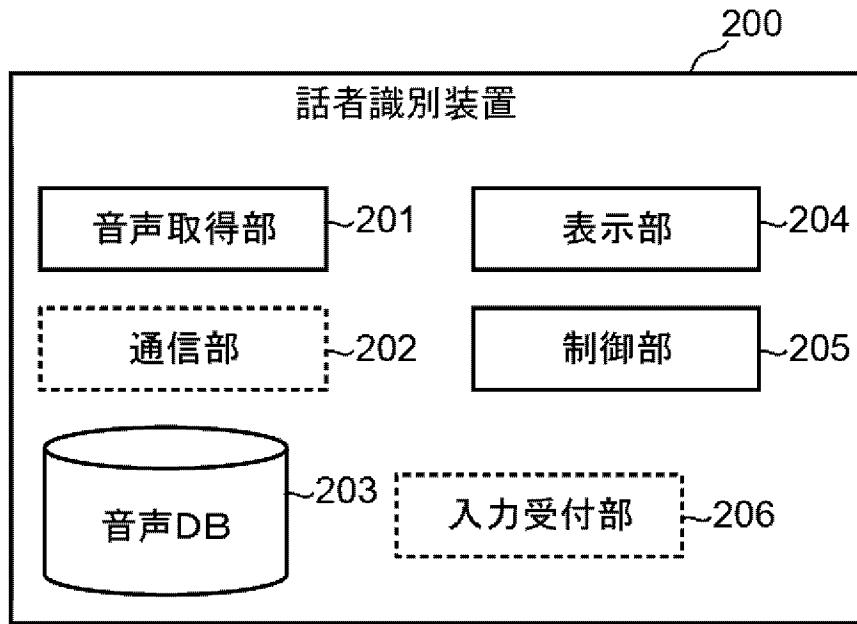
前記位置情報検出部は、前記音声取得部もしくは前記音声処理部が、取得する音声の指向性を制御することで、前記位置情報を検出する、

請求項16に記載の話者識別装置。

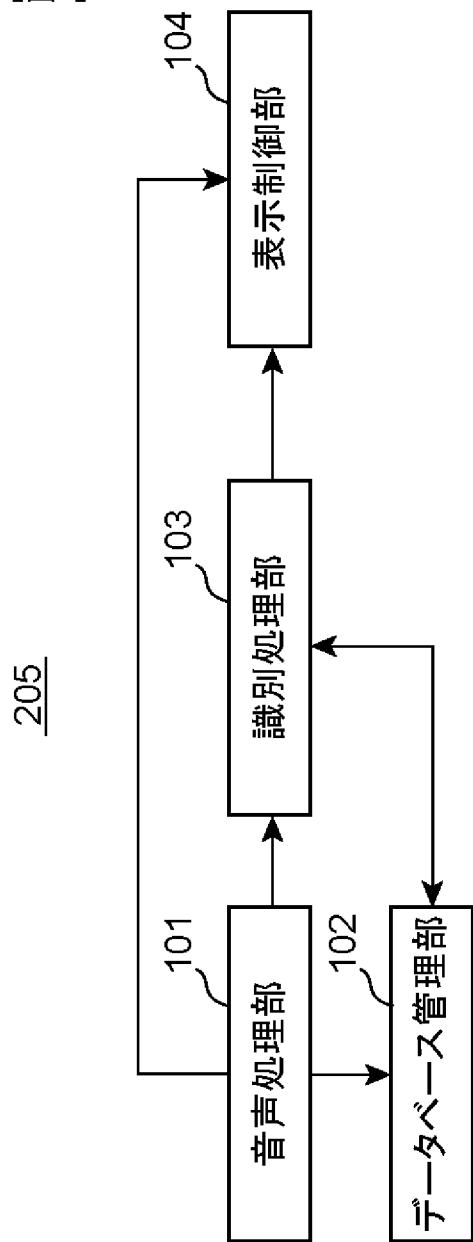
## 条約第19条(1)に基づく説明書

請求項1-12は、補正されました。請求項13-18は、追加されました。なお、補正後の請求項5-12は、それぞれ出願時の請求項2-9に対応しています。

[図1]



[図2]



[図3]

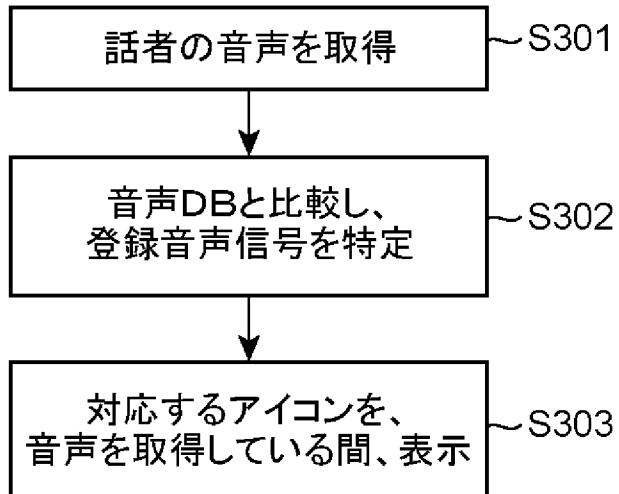
登録音声信号	ユーザ情報			登録アイコン
	年齢	性別	ニックネーム	
0001.wav	40代	男性	/アス	A01
0002.wav	30代	女性	ママ	—
0003.wav	10代	女性	あーちゃん	B05

[図4]

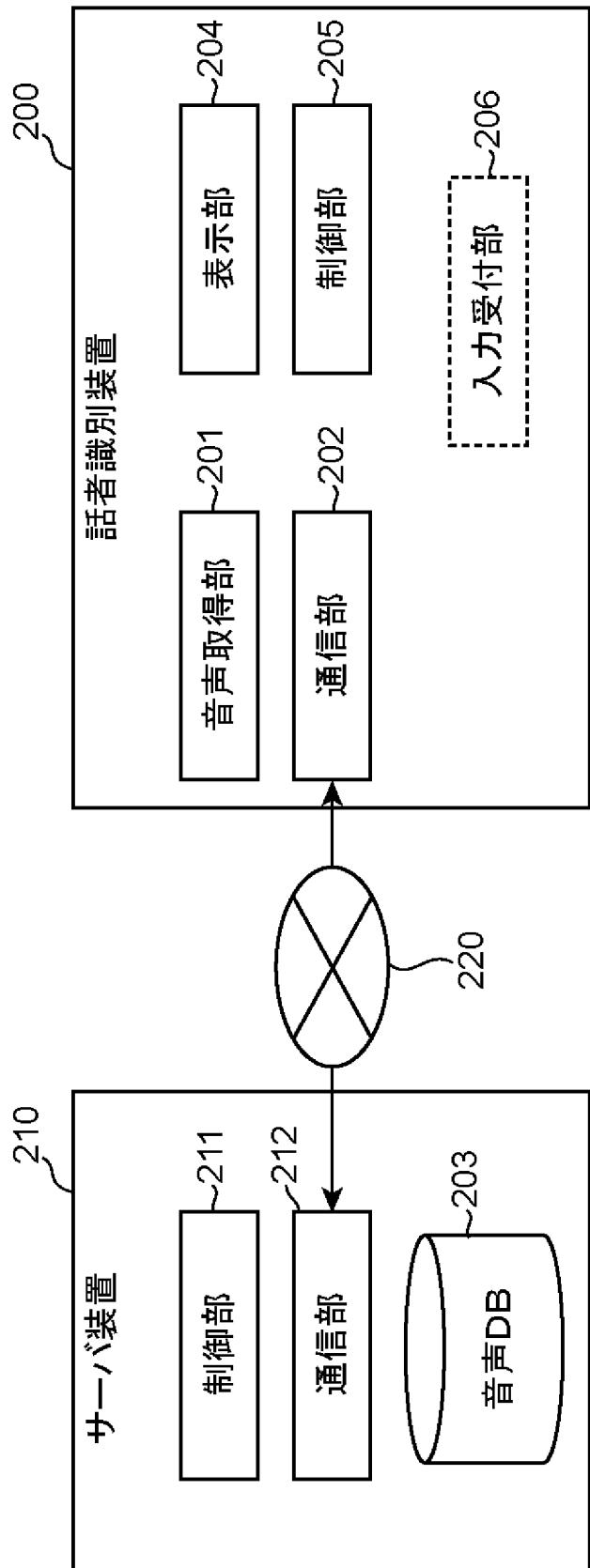
810

登録音声信号	音声データ
0001.wav	番組表、8チャンネル
0002.wav	料理番組を見ながら 夕食を作りましょう
0003.wav	スポーツ番組の 野球の結果は？

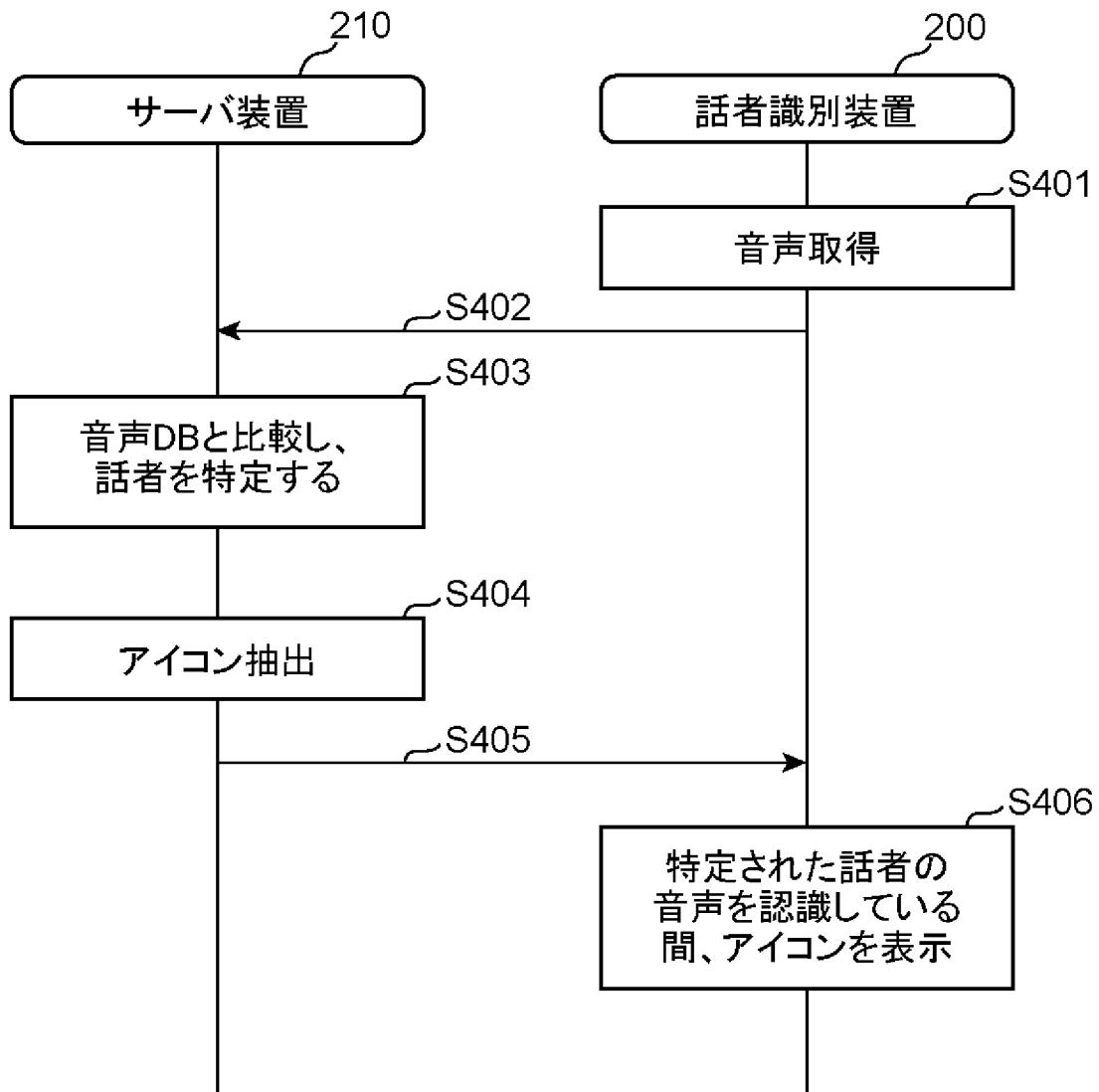
[図5]



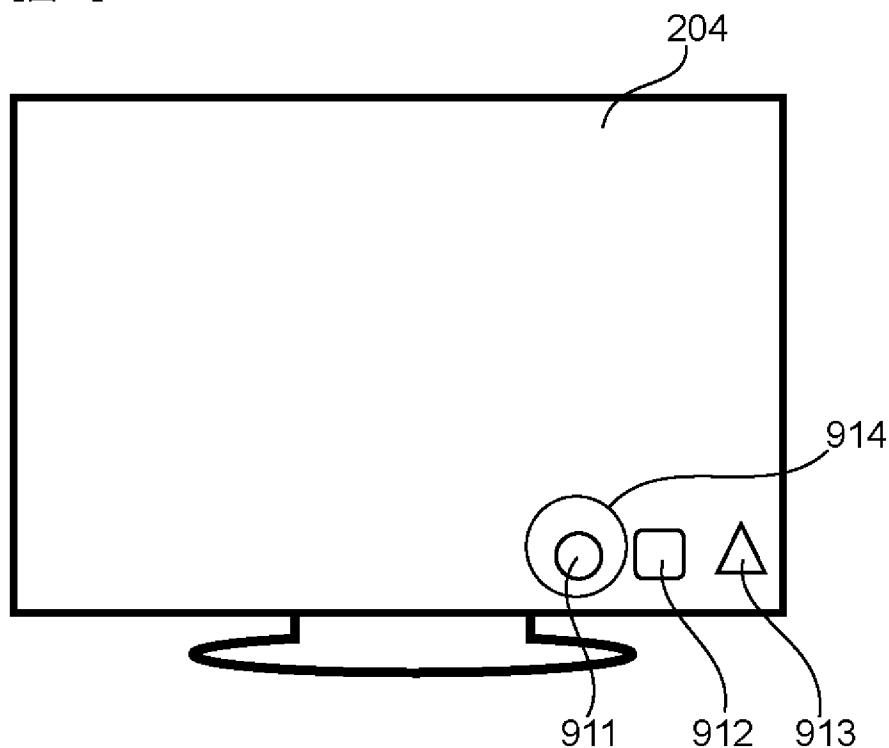
[図6]



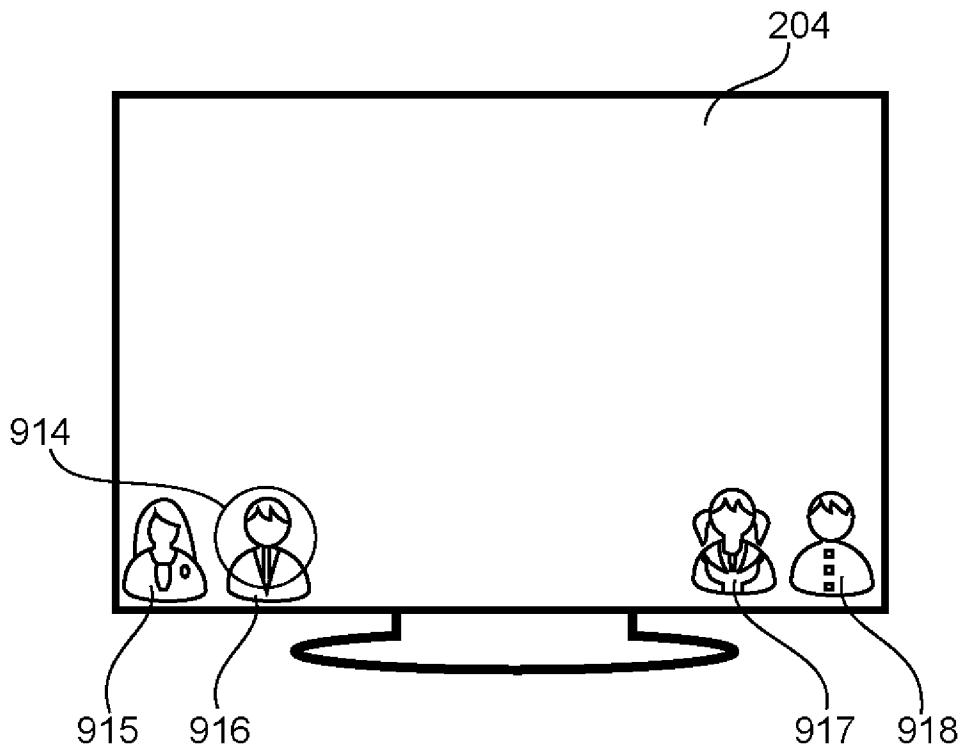
[図7]



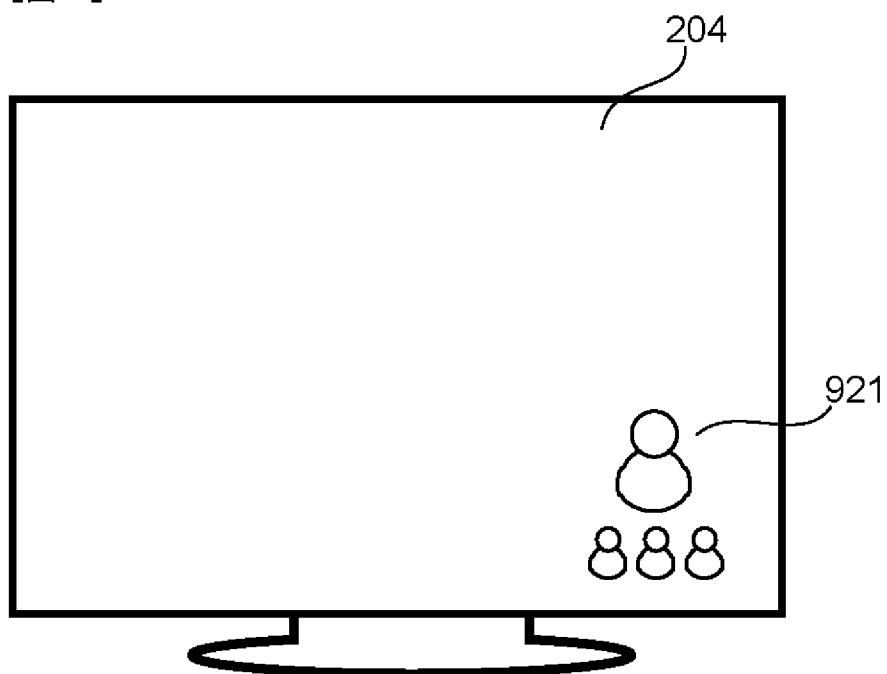
[図8A]



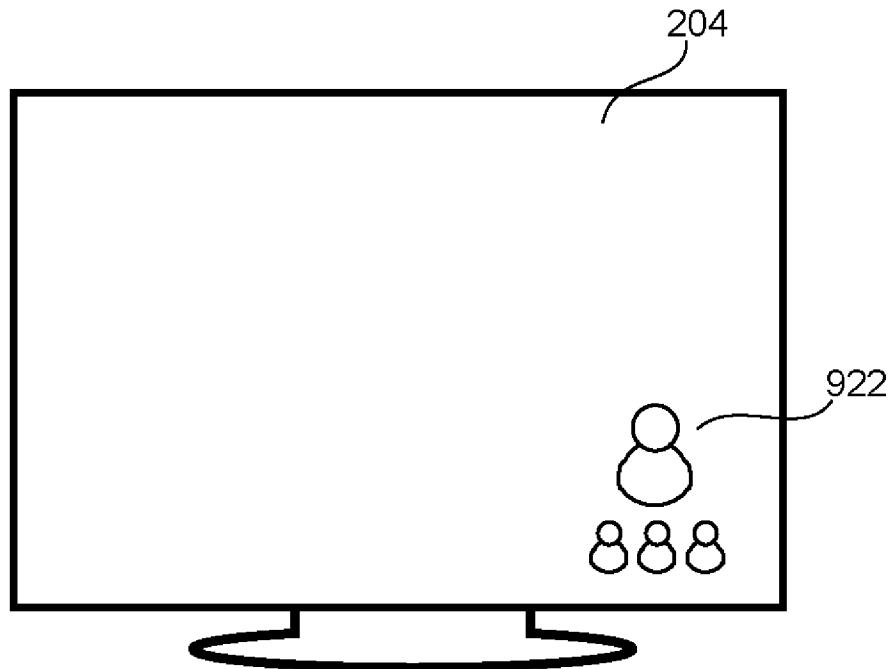
[図8B]



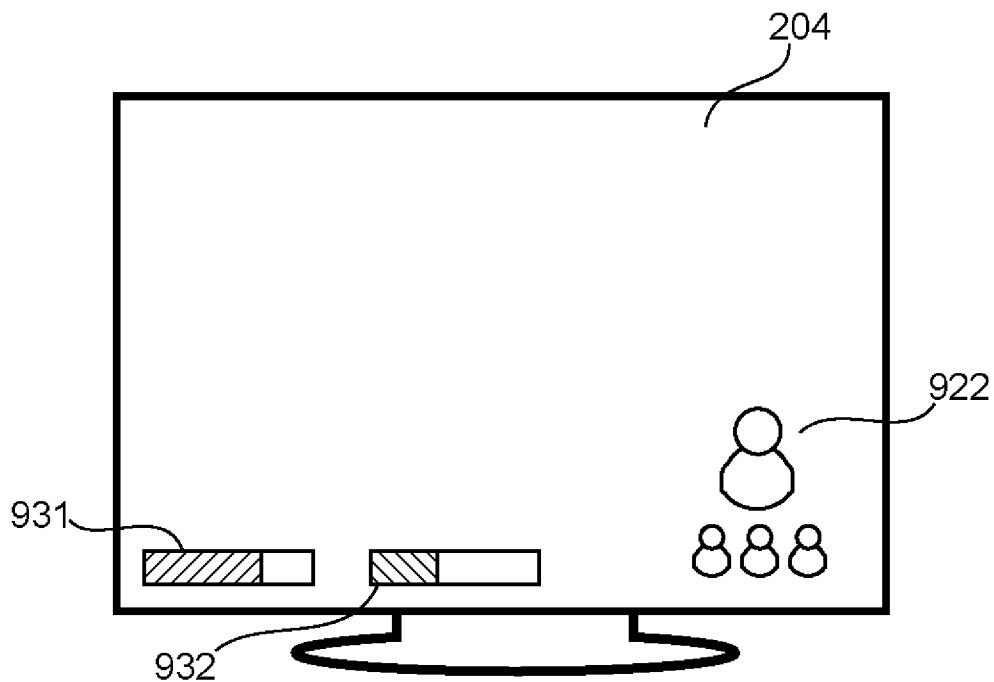
[図8C]



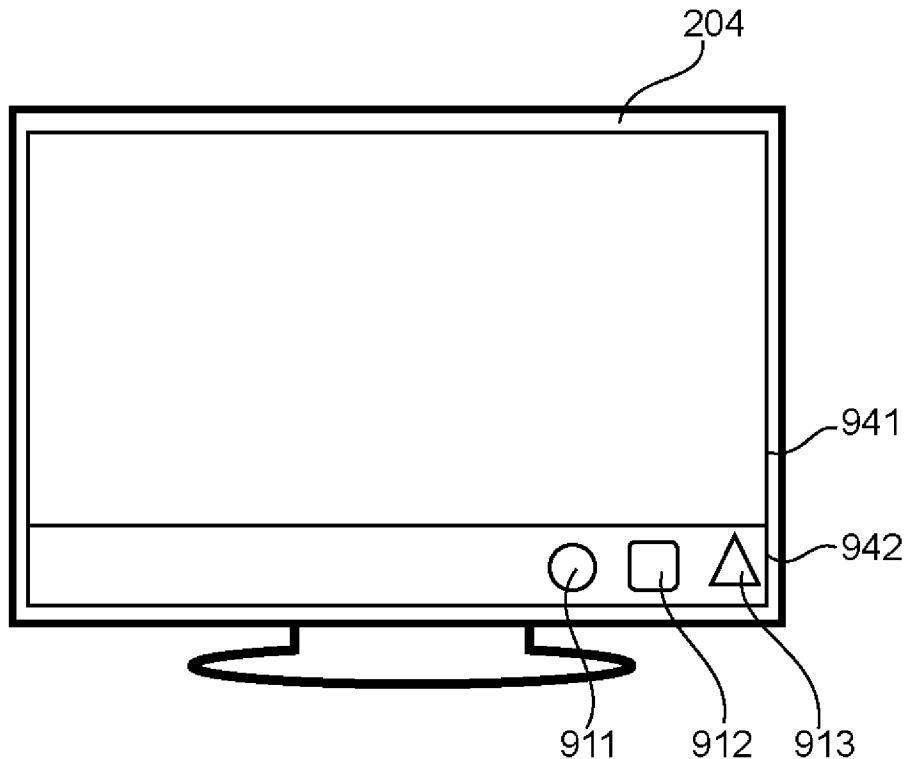
[図8D]



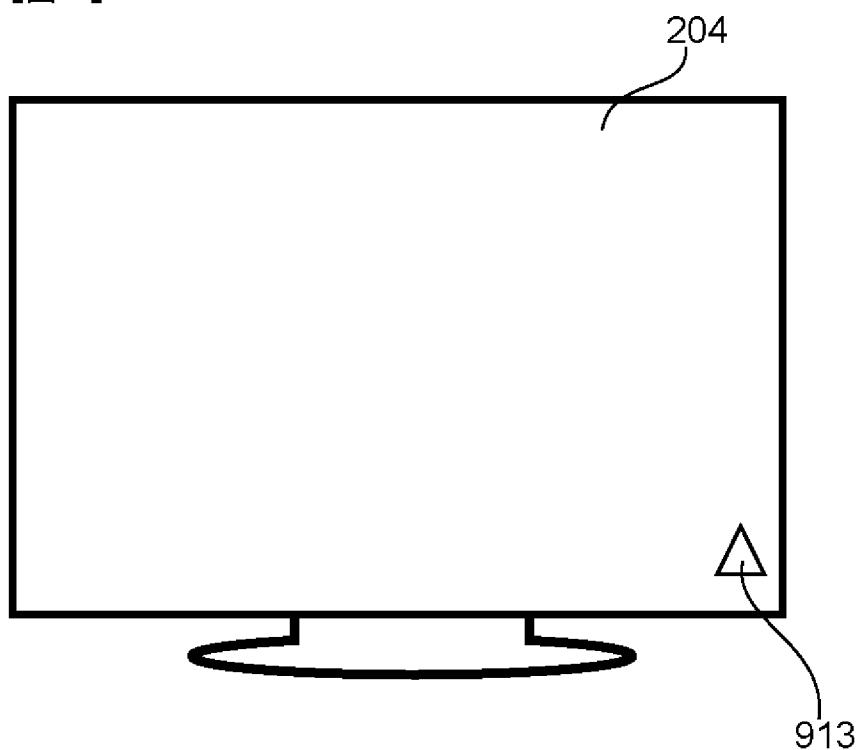
[図8E]



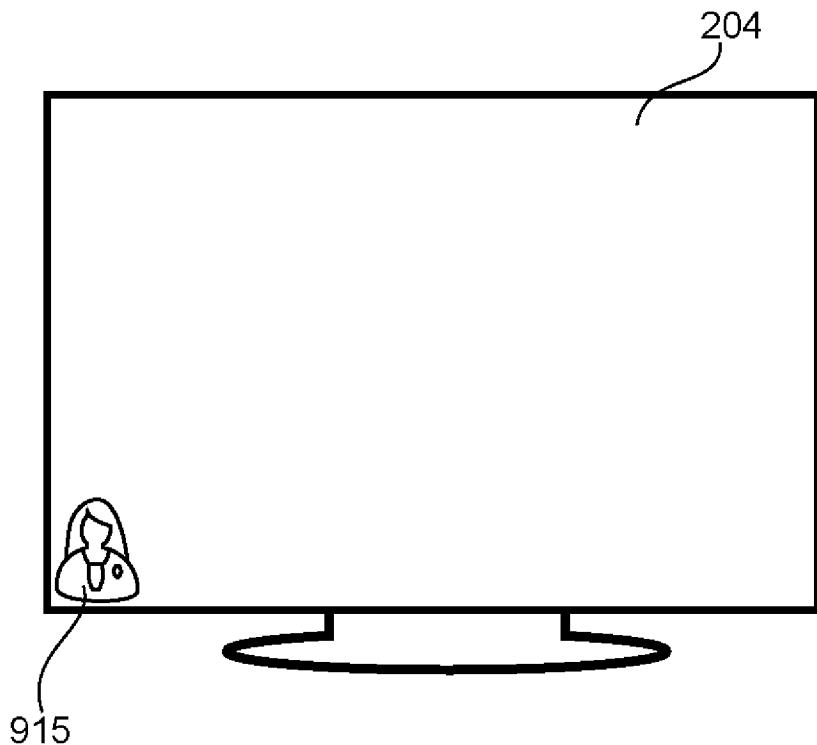
[図8F]



[図8G]

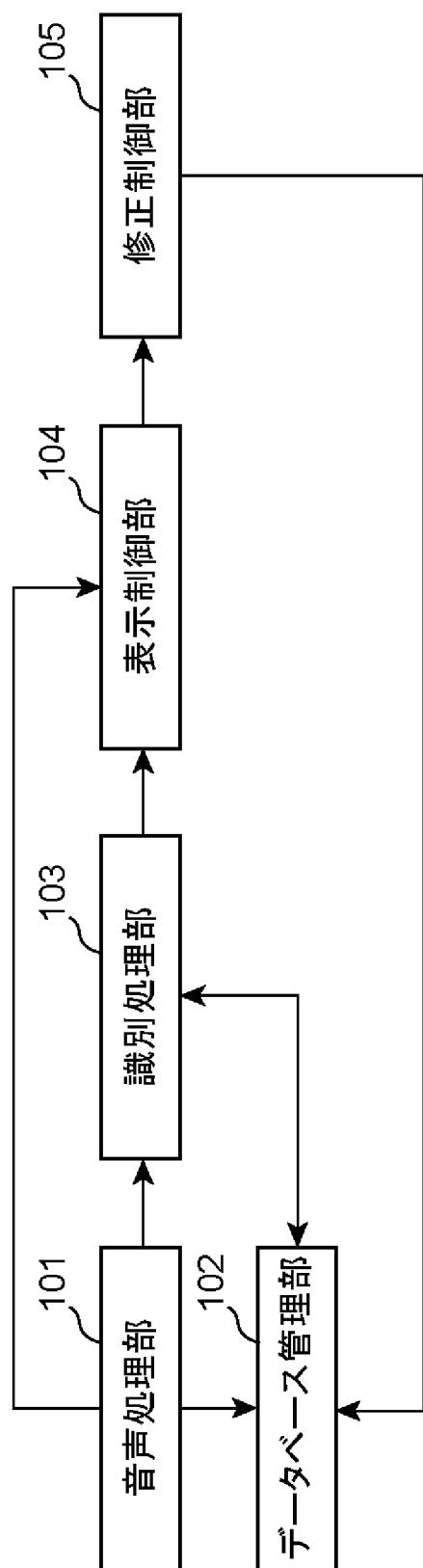


[図8H]

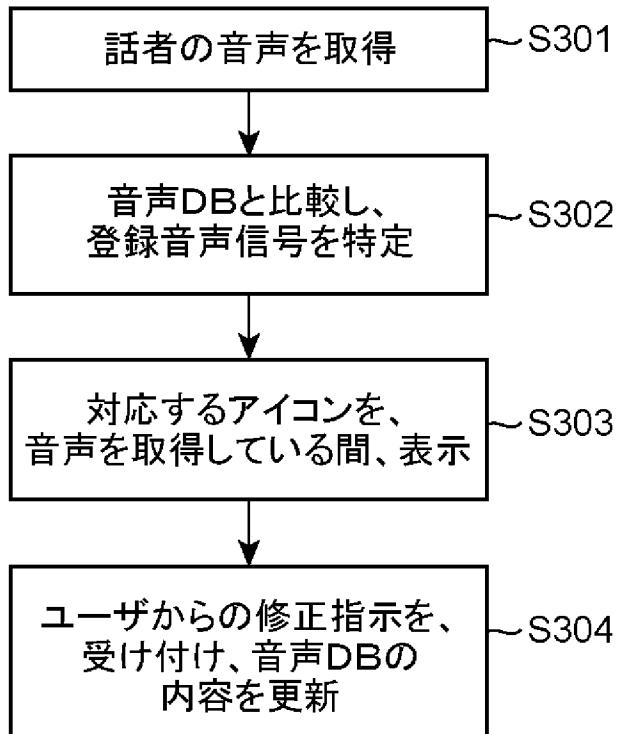


[図9]

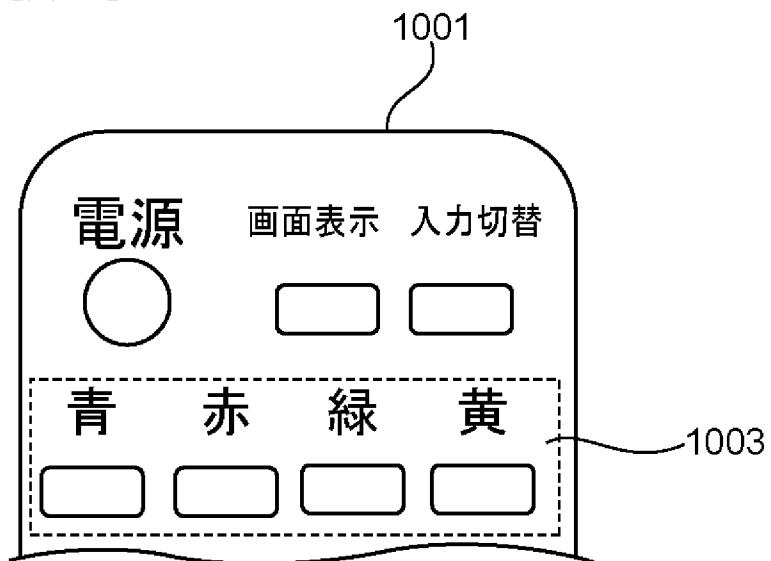
205



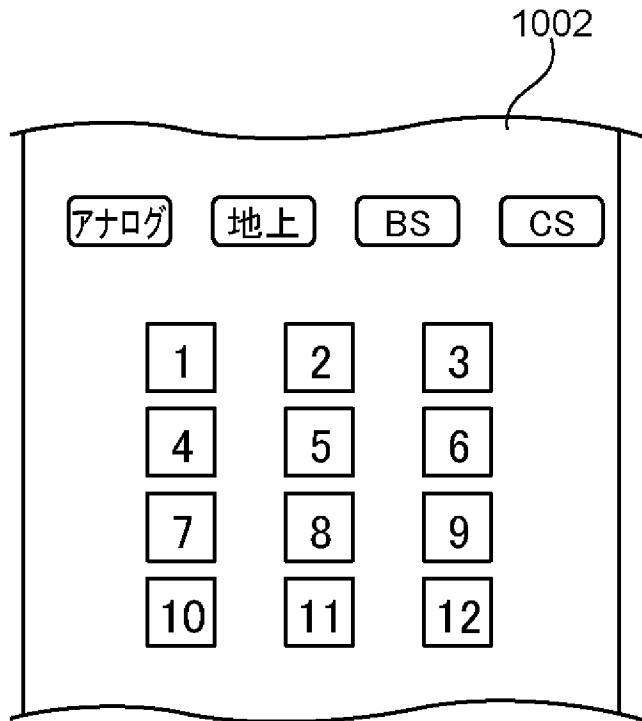
[図10]



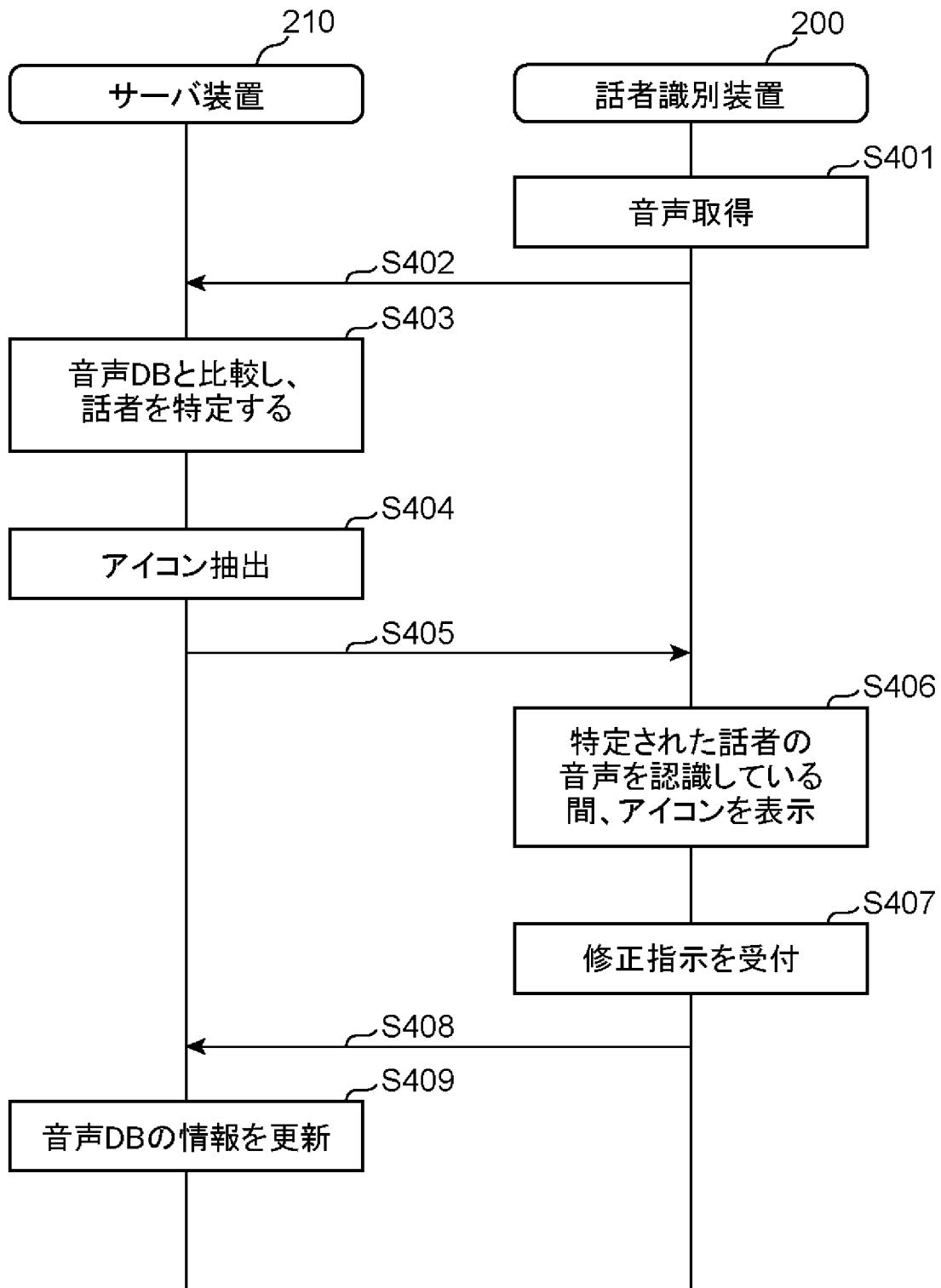
[図11A]



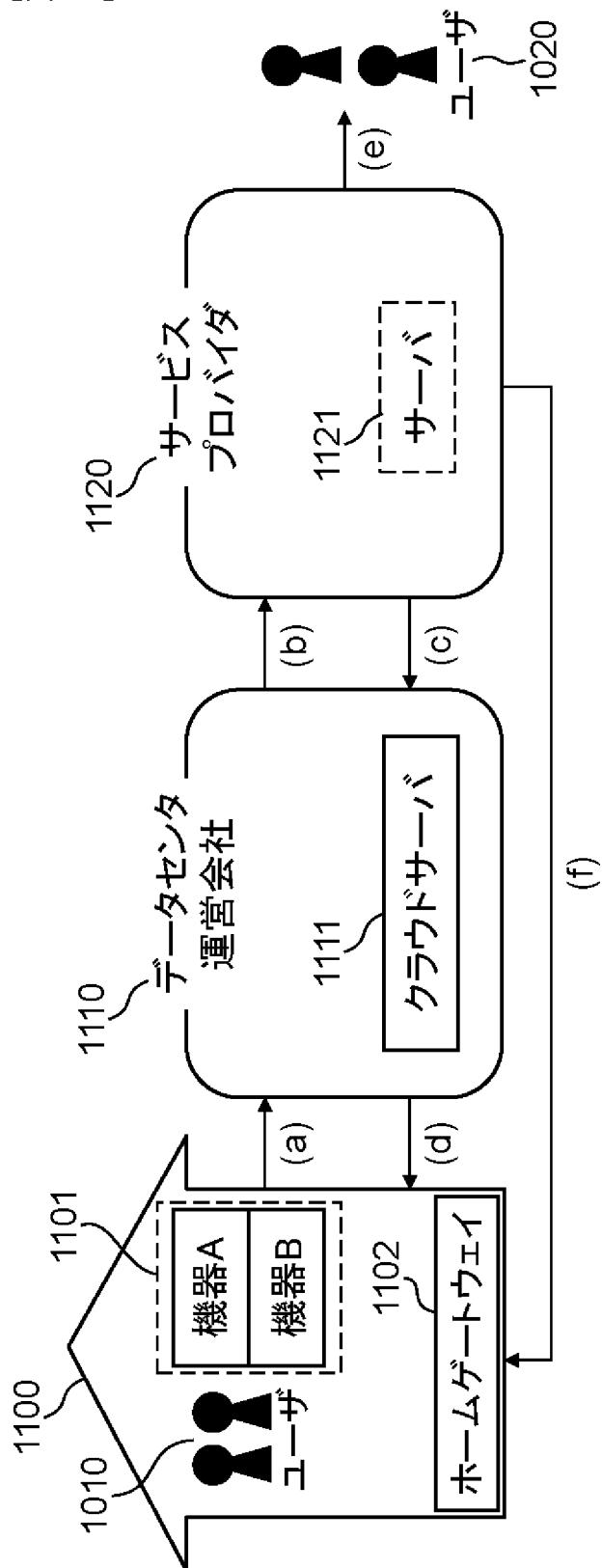
[図11B]



[図12]



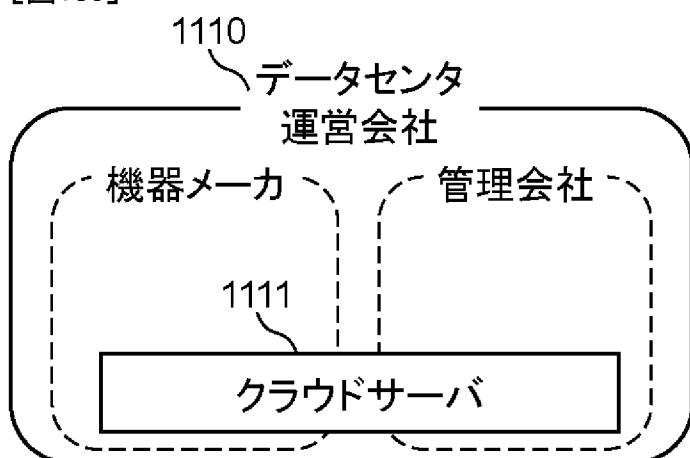
[図13A]



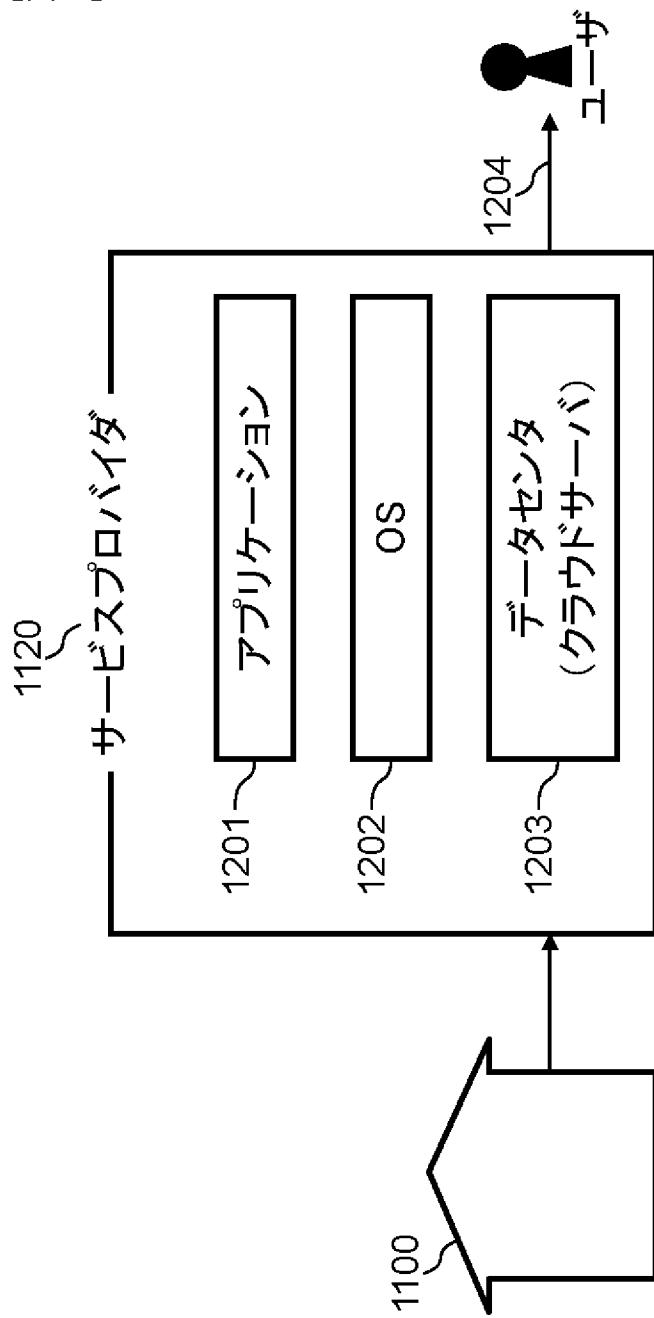
[図13B]



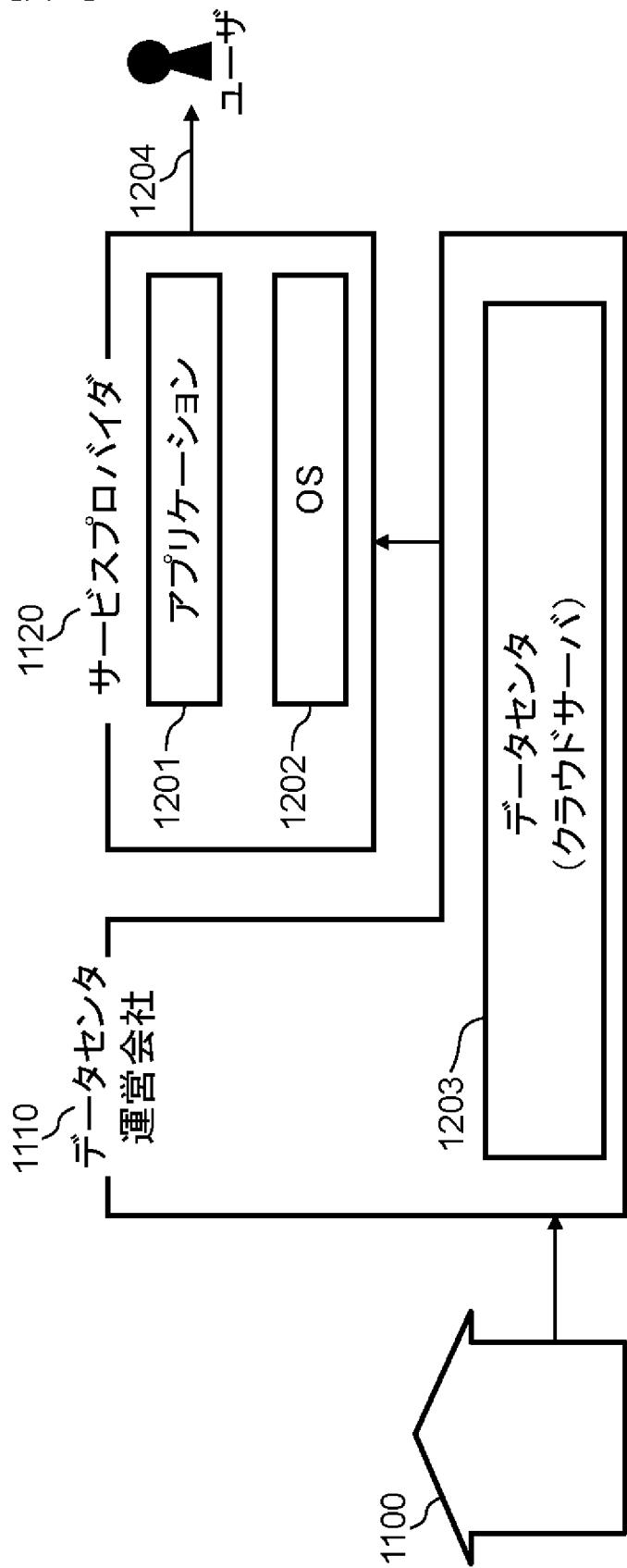
[図13C]



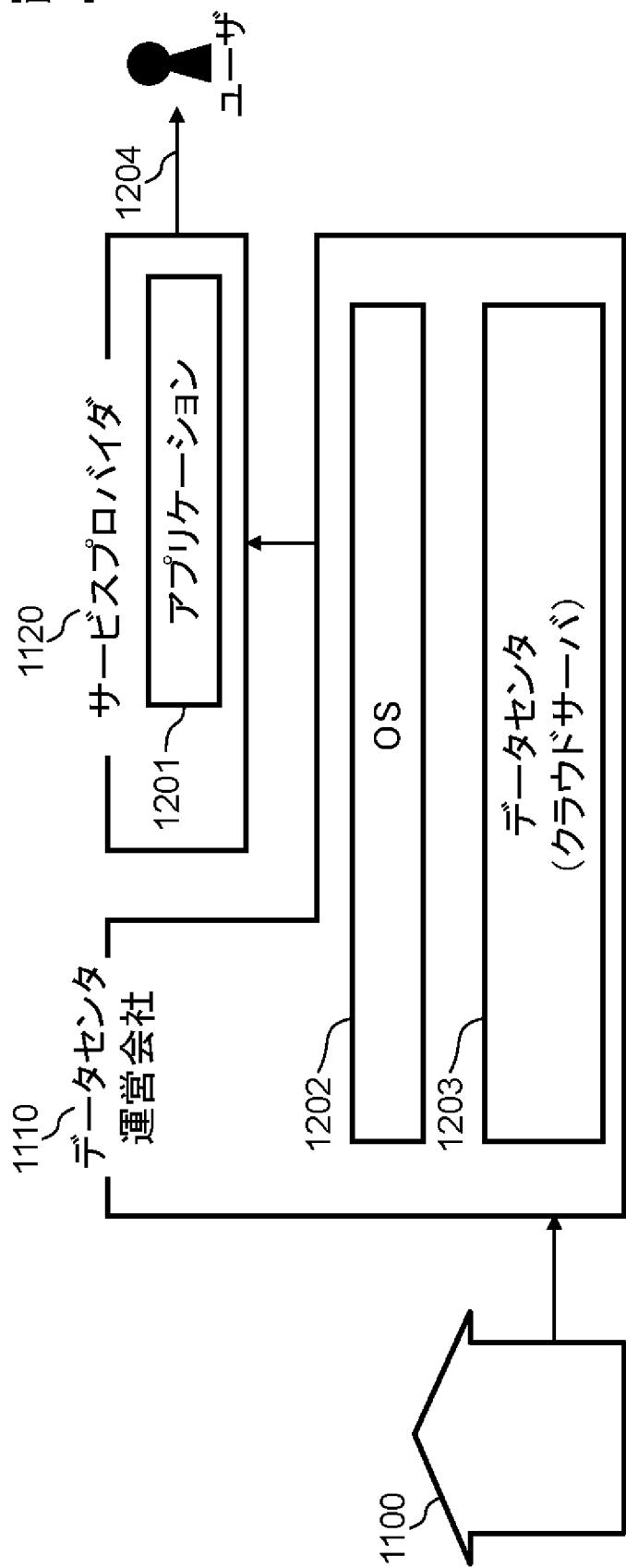
[図14]



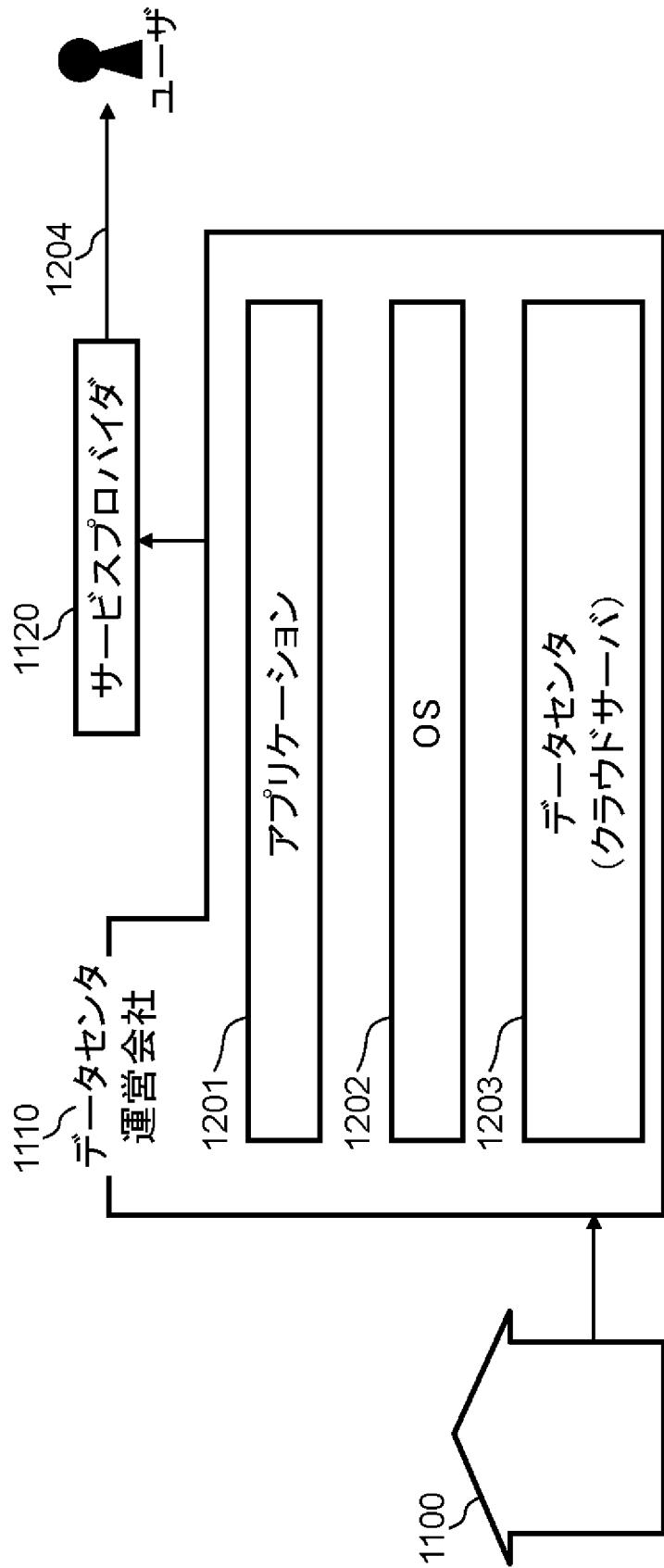
[図15]



[図16]



[図17]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2014/002970

### A. CLASSIFICATION OF SUBJECT MATTER

G10L15/22(2006.01)i, G06F3/16(2006.01)i, G10L15/00(2013.01)i, G10L17/00 (2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

### B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L15/22, G06F3/16, G10L15/00, G10L17/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2014
Kokai Jitsuyo Shinan Koho	1971-2014	Toroku Jitsuyo Shinan Koho	1994-2014

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

### C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 2011/116309 A1 (DIGIMARC CORP.), 22 September 2011 (22.09.2011), column 76, lines 9 to 21; fig. 19 & JP 2013-527947 A & US 2011/0161076 A1 & US 2012/0154633 A1 & US 2011/0244919 A1 & EP 2559030 A4 & CA 2792336 A1 & CN 102893327 A & KR 10-2013-0027081 A	1, 3, 4, 9-12 2-4, 6-9
Y	JP 2010-219703 A (Mitsumi Electric Co., Ltd.), 30 September 2010 (30.09.2010), paragraph [0032] (Family: none)	2-4, 6-9

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier application or patent but published on or after the international filing date
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search  
02 September, 2014 (02.09.14)

Date of mailing of the international search report  
16 September, 2014 (16.09.14)

Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2014/002970

**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2007/058135 A1 (NEC Corp.), 24 May 2007 (24.05.2007), entire text; all drawings & US 2009/0264085 A1 & EP 1954014 A1 & CN 101310505 A	6-9
Y	JP 2001-314649 A (Seta Corp.), 13 November 2001 (13.11.2001), paragraphs [0021] to [0024] (Family: none)	7-9
A	JP 2001-339529 A (Sharp Corp.), 07 December 2001 (07.12.2001), entire text; all drawings (Family: none)	1-4, 6-12
A	JP 2011-34382 A (Ricoh Co., Ltd.), 17 February 2011 (17.02.2011), entire text; all drawings (Family: none)	1-4, 6-12
A	JP 2004-56286 A (Fuji Photo Film Co., Ltd.), 19 February 2004 (19.02.2004), entire text; all drawings (Family: none)	1-4, 6-12

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2014/002970

**Box No. II      Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.: 5  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
(See extra sheet)
  
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III      Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2014/002970

Continuation of Box No.II-2 of continuation of first sheet(2)

Claim 5 states that the form of display is varied according to which of a first speaker image and a second speaker image is registered earlier in a database, but it is not certain to which statement in the description this statement corresponds, and in the meaning of PCT Article 6, the claim lacks sufficient support by the description.

## A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G10L15/22(2006.01)i, G06F3/16(2006.01)i, G10L15/00(2013.01)i, G10L17/00(2013.01)i

## B. 調査を行った分野

## 調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G10L15/22, G06F3/16, G10L15/00, G10L17/00

## 最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2014年
日本国実用新案登録公報	1996-2014年
日本国登録実用新案公報	1994-2014年

## 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X	WO 2011/116309 A1 (DIGIMARC CORPORATION) 2011.09.22, 第76欄 第9-21行, 第19図 & JP 2013-527947 A & US 2011/0161076 A1 & US 2012/0154633 A1 & US 2011/0244919 A1 & EP 2559030 A4 & CA 2792336 A1 & CN 102893327 A & KR 10-2013-0027081 A	1, 3, 4, 9-12
Y	JP 2010-219703 A (ミツミ電機株式会社) 2010.09.30, 段落【0032】 (ファミリーなし)	2-4, 6-9

 C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

- 「A」特に関連のある文献ではなく、一般的技術水準を示すもの
- 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
- 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
- 「O」口頭による開示、使用、展示等に言及する文献
- 「P」国際出願目前で、かつ優先権の主張の基礎となる出願

## の日の後に公表された文献

- 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
- 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
- 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
- 「&」同一パテントファミリー文献

国際調査を完了した日  02.09.2014	国際調査報告の発送日  16.09.2014
国際調査機関の名称及びあて先  日本国特許庁（ISA/JP） 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員）  間宮 嘉誉 電話番号 03-3581-1101 内線 3591 5Z 3454

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	WO 2007/058135 A1 (日本電気株式会社) 2007.05.24, 全文, 全図 & US 2009/0264085 A1 & EP 1954014 A1 & CN 101310505 A	6-9
Y	JP 2001-314649 A (株式会社セタ) 2001.11.13, 段落【0021】-【0024】(ファミリーなし)	7-9
A	JP 2001-339529 A (シャープ株式会社) 2001.12.07, 全文, 全図 (ファミリーなし)	1-4, 6-12
A	JP 2011-34382 A (株式会社リコー) 2011.02.17, 全文, 全図 (ファミリーなし)	1-4, 6-12
A	JP 2004-56286 A (富士写真フィルム株式会社) 2004.02.19, 全文, 全図 (ファミリーなし)	1-4, 6-12

## 第II欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1.  請求項 \_\_\_\_\_ は、この国際調査機関が調査をすることを要しない対象に係るものである。  
つまり、

2.  請求項 \_\_\_\_\_ 5 \_\_\_\_\_ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、

請求項5には、第1話者画像と第2話者画像とのうち、データベースにいずれの画像が先に登録されたかに応じて表示の態様を異ならせることが記載されているが、そのようなことは明細書のいずれの記載に対応したものか定かではなく、PCT第6条の意味における明細書による十分な裏付けを欠いている。

3.  請求項 \_\_\_\_\_ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

## 第III欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

1.  出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求項について作成した。
2.  追加調査手数料を要求するまでもなく、すべての調査可能な請求項について調査することができたので、追加調査手数料の納付を求めなかった。
3.  出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求項のみについて作成した。
4.  出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求項について作成した。

## 追加調査手数料の異議の申立てに関する注意

- 追加調査手数料及び、該当する場合には、異議申立手数料の納付と共に、出願人から異議申立てがあった。
- 追加調査手数料の納付と共に出願人から異議申立てがあったが、異議申立手数料が納付命令書に示した期間内に支払われなかった。
- 追加調査手数料の納付はあったが、異議申立てはなかった。