

(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
2016 年 10 月 13 日(13.10.2016)



(10) 国際公開番号
WO 2016/163028 A1

- (51) 国際特許分類:
G10L 15/22 (2006.01) *G10L 15/10* (2006.01)
G10L 15/00 (2013.01)
- (21) 国際出願番号: PCT/JP2015/061269
- (22) 国際出願日: 2015 年 4 月 10 日(10.04.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 株式会社東芝 (KABUSHIKI KAISHA TOSHIBA) [JP/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 Tokyo (JP). 東芝ソリューション株式会社 (TOSHIBA SOLUTIONS CORPORATION) [JP/JP]; 〒2128585 神奈川県川崎市幸区堀川町 7 2 番地 3 4 Kanagawa (JP).
- (72) 発明者: 長 健太 (CHIYO, Kenta); 〒2128585 神奈川県川崎市幸区堀川町 7 2 番地 3 4 東芝ソリューション株式会社内 Kanagawa (JP). 加納 敏行 (KANO, Toshiyuki); 〒2128585 神奈川県川崎市幸区堀川町 7 2 番地 3 4 東芝ソリューション株式会社内 Kanagawa (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒

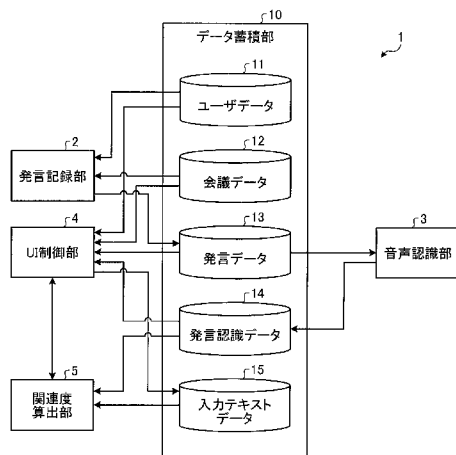
1000013 東京都千代田区霞が関 3 丁目 8 番 1 号
虎の門三井ビルディング Tokyo (JP).

- (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, 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, ST, 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).

[続葉有]

(54) Title: UTTERANCE PRESENTATION DEVICE, UTTERANCE PRESENTATION METHOD, AND PROGRAM

(54) 発明の名称: 発言提示装置、発言提示方法およびプログラム



- 2 Utterance recording unit
3 Speech recognition unit
4 UI control unit
5 Relevancy calculation unit
10 Data storage unit
11 User data
12 Conference data
13 Utterance data
14 Utterance recognition data
15 Input text data

(57) Abstract: The utterance presentation device (1) according to an embodiment is provided with an utterance recording unit (2), a speech recognition unit (3), a relevancy calculation unit (5), and a UI control unit (4). The utterance recording unit (2) records a vocalized utterance. The speech recognition unit (3) performs speech recognition on the recorded utterance. The relevancy calculation unit (5) calculates the degree of relevancy between each utterance on which speech recognition was performed and a designated character string from among character strings displayed in a second display region of a UI screen including a first display region and the second display region. The UI control unit (4) displays, in the first display region of the UI screen, the results of the speech recognition for the utterances selected on the basis of the degree of relevancy.

(57) 要約: 実施形態の発言提示装置 (1) は、発言記録部 (2) と、音声認識部 (3) と、関連度算出部 (5) と、UI 制御部 (4) と、を備える。発言記録部 (2) は、音声による発言を記録する。音声認識部 (3) は、記録された発言を音声認識する。関連度算出部 (5) は、音声認識された各発言に対し、第 1 表示領域と第 2 表示領域とを有する UI 画面の前記第 2 表示領域に表示されている文字列のうち指定された文字列との関連度を各々算出する。UI 制御部 (4) は、前記関連度の高さに基づいて選択された発言の音声認識結果を、前記 UI 画面の前記第 1 表示領域に表示させる。

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称： 発言提示装置、発言提示方法およびプログラム

技術分野

[0001] 本発明の実施形態は、発言提示装置、発言提示方法およびプログラムに関する。

背景技術

[0002] 会議中に記述したメモが会議中のどの発言に対応するかの対応付けを行うことは、例えば会議の議事録作成などの作業を効率化する上で有効である。このような対応付けを行う技術として、会議中の音声および映像の記録に合わせて、テキストとして入力したメモの入力時間を記録し、記録された音声や映像の再生時に対応するメモ部分を表示したり、メモに対応する音声や映像を再生したりする技術が知られている。

[0003] しかし、会議中の発言との対応付けが望まれる情報は、会議中にテキストとして入力されたメモに限らない。例えば、会議中に紙に書かれた手書きのメモや会議前に作成されたアジェンダなど、会議中に入力されない情報についても、会議中の発言との対応付けが望まれる場合もある。また、会議中の発言に限らず、音声による発言を記録する仕組みを持つ様々なシステムにおいて、任意の情報に対応する発言をユーザに分かり易く提示できるようにしたいというニーズがある。

先行技術文献

特許文献

[0004] 特許文献1：特開2008-172582号公報

発明の概要

発明が解決しようとする課題

[0005] 本発明が解決しようとする課題は、任意の情報に対応する発言をユーザに分かり易く提示できる発言提示装置、発言提示方法およびプログラムを提供することである。

課題を解決するための手段

[0006] 実施形態の発言提示装置は、発言記録部と、音声認識部と、関連度算出部と、UI制御部と、を備える。発言記録部は、音声による発言を記録する。音声認識部は、記録された発言を音声認識する。関連度算出部は、音声認識された各発言に対し、第1表示領域と第2表示領域とを有するUI画面の前記第2表示領域に表示されている文字列のうち指定された文字列との関連度を各々算出する。UI制御部は、前記関連度の高さに基づいて選択された発言の音声認識結果を、前記UI画面の前記第1表示領域に表示させる。

図面の簡単な説明

[0007] [図1]図1は、第1実施形態の発言提示装置の構成例を示すブロック図である。

[図2]図2は、実施形態の発言提示装置の動作概要を示すフローチャートである。

[図3]図3は、会議中の発言を収録するシーンの具体例を説明する図である。

[図4]図4は、ユーザデータの具体例を示す図である。

[図5]図5は、会議データの具体例を示す図である。

[図6]図6は、会議中の発言の具体例を示す図である。

[図7]図7は、発言データの具体例を示す図である。

[図8]図8は、発言認識データの具体例を示す図である。

[図9]図9は、UI画面の一例を示す図である。

[図10]図10は、「議事メモ」領域に議事メモが記入されたUI画面を示す図である。

[図11]図11は、入力テキストデータの具体例を示す図である。

[図12]図12は、第2実施形態の発言提示装置の構成例を示すブロック図である。

[図13]図13は、第3実施形態におけるUI画面の一例を示す図である。

[図14]図14は、第4実施形態の発言提示装置の構成例を示すブロック図である。

[図15]図 1 5 は、録音環境データ的具体例を示す図である。

[図16]図 1 6 は、会議設定画面の一例を示す図である。

[図17]図 1 7 は、発言提示装置のハードウェア構成の一例を概略的に示すブロック図である。

発明を実施するための形態

[0008] 以下、実施形態の発言提示装置、発言提示方法およびプログラムを、図面を参照して詳細に説明する。以下で示す実施形態では、会議中の音声による発言を記録し、会議後の議事メモ作成時に、指定されたメモ部分に関連する会議中の発言を提示する構成の発言提示装置を例示する。この発言提示装置は、例えば、ネットワークを利用したサーバ・クライアントシステムのサーバ装置として実現され、クライアント端末に後述のUI画面を表示させてこのUI画面を用いた操作に基づく処理を行うなどのサービスを提供する。なお、発言提示装置は、クラウドシステム上で動作する仮想マシンであってもよい。また、発言提示装置は、ユーザが直接利用する独立の装置として構成されていてもよい。

[0009] <第1実施形態>

図 1 は、本実施形態の発言提示装置 1 の構成例を示すブロック図である。この発言提示装置 1 は、図 1 に示すように、発言記録部 2、音声認識部 3、UI制御部 4、関連度算出部 5、およびデータ蓄積部 10を備える。

[0010] 発言記録部 2 は、会議中に発生する音声による発言を記録する。発言は、その発言を行った発言ユーザが装着するピンマイクもしくはヘッドセットのマイクなどの個別マイク、または集音マイクに入力される。個別マイクまたは集音マイクに入力された発言は、例えば、発言の発生日時、発言ユーザを識別するユーザID (identification) とともに、発言提示装置 1 に送信される。発言記録部 2 は、受信した音声を音声ファイルとして記録する。

[0011] 発言記録部 2 により記録された発言の音声ファイルは、発言を識別する発言ID、発言の発生日時、発言ユーザのユーザID、発言の収録に用いたマイク種別 (収録マイク種別)、発言が行われた会議を識別する会議IDなど

とともに、発言データ 13 としてデータ蓄積部 10 に蓄積される。収録マイク種別は、例えば、会議開催前のユーザによる登録操作に応じてデータ蓄積部 10 に格納されたユーザデータ 11 を参照することにより特定される。また、会議 ID は、例えば、会議開催前のユーザによる登録操作に応じてデータ蓄積部 10 に格納された会議データ 12 を参照することにより特定される。

[0012] 音声認識部 3 は、発言記録部 2 が記録した発言に対する音声認識を行う。音声認識の方法は公知の技術をそのまま利用できるように、ここでは詳細な説明を省略する。音声認識部 3 は、例えば、入力された発言の音声に対する認識結果の候補のうち、尤度が最大となる候補を音声認識結果として出力するとともに、各候補に含まれる単語すべてを認識キーワードとして出力する。

[0013] 音声認識部 3 が出力する発言の音声認識結果および認識キーワードは、発言を識別する発言 ID、想定される音声認識の精度を表す想定認識精度などとともに、発言認識データ 14 としてデータ蓄積部 10 に蓄積される。想定認識精度は、例えば、発言の音声の入力方式（具体的には収録マイク種別）などに応じて設定される。

[0014] UI 制御部 4 は、ユーザの議事メモ作成を支援する UI 画面を生成してクライアント端末に提供する。UI 画面は、会議中の発言の音声認識結果を表示する「発言一覧」領域（第 1 表示領域）と、議事メモの入力を受け付ける「議事メモ」領域（第 2 表示領域）とを有する。UI 画面の「発言一覧」領域には、会議中に収録された発言の音声認識結果が表示される。UI 画面の「議事メモ」領域は、ユーザが会議の議事メモを入力するために利用される。ユーザが入力した議事メモは「議事メモ」領域にテキストとして表示される。また、「議事メモ」領域は、議事メモのほかにも、会議前に登録されたアジェンダなどの会議に関連する他のテキストが表示される構成であってもよい。ユーザが「議事メモ」領域に入力した議事メモは、例えば行単位で管理され、各行のメモ部分を識別するメモ ID、入力された行、議事メモに対応する会議を識別する会議 ID などとともに、入力テキストデータ 15 とし

てデータ蓄積部 10 に蓄積される。

[0015] また、UI 制御部 4 は、ユーザが UI 画面の「議事メモ」領域に表示されているテキストから任意の文字列を指定する操作を行うと、対応する会議中に収録されて音声認識部 3 による音声認識が行われた発言のうち、指定された文字列との関連度の高さに基づいて選択された発言の音声認識結果を UI 画面の「発言一覧」領域に表示させる。「議事メモ」領域で指定された文字列との関連度は、後述の関連度算出部 5 により算出される。さらに UI 制御部 4 は、ユーザが UI 画面の「発言一覧」領域に音声認識結果が表示されている発言の中から任意の発言を指定する操作を行うと、その発言の音声を再生させる制御を行う。

[0016] UI 制御部 4 は、例えば、ウェブベースで UI 画面を生成してクライアント端末に提供するウェブサーバとして実装される。この場合、クライアント端末は、UI 制御部 4 が生成した UI 画面を、ウェブブラウザを用いてネットワーク越しに利用する。なお、UI 画面の具体的な構成例については詳細を後述する。

[0017] 関連度算出部 5 は、ユーザが UI 画面の「議事メモ」領域に表示されているテキストから任意の文字列を指定する操作を行うと、対応する会議中に収録され、音声認識部 3 により音声認識が行われた会議中の各発言に対し、指定された文字列との関連度を各々算出する。この関連度算出部 5 により算出された関連度の高さに基づいて、UI 画面の「議事メモ」領域で指定された文字列に対応する発言として、UI 画面の「発言一覧」領域に音声認識結果が表示される発言が選択される。なお、関連度の算出方法の具体例については詳細を後述する。

[0018] 次に、本実施形態の発言提示装置 1 による動作の流れを簡単に説明する。図 2 は、本実施形態の発言提示装置 1 の動作概要を示すフローチャートであり、(a) は会議が行われるたびに実施される発言提示装置 1 の動作を示し、(b) は会議後にクライアント端末において UI 画面が開かれたときの発言提示装置 1 の動作を示している。

- [0019] なお、会議中に個別マイクにより発言を行うユーザ（会議参加者）の情報や、開催される会議の情報は、会議の開始前にクライアント端末から発言提示装置１にアクセスして登録されるものとする。登録された会議参加者の情報は、ユーザデータ１１としてデータ蓄積部１０に格納され、登録された会議の情報は、会議データ１２としてデータ蓄積部１０に格納される。
- [0020] 会議が開始されると、会議中の音声による発言が、個別マイクまたは集音マイクに入力されてクライアント端末から発言提示装置１に送信される。発言提示装置１の発言記録部２は、個別マイクまたは集音マイクに入力された発言を、音声ファイルとして記録する（ステップＳ１０１）。発言記録部２により記録された発言の音声ファイルは、発言データ１３としてデータ蓄積部１０に格納される。
- [0021] 発言記録部２による発言の記録および発言データ１３の格納は、会議が終了するまで継続される。すなわち、会議終了を示すユーザの明示的な操作の有無などにより会議が終了したか否かが判定され（ステップＳ１０２）、会議が終了していなければ（ステップＳ１０２：Ｎｏ）、個別マイクまたは集音マイクに会議中の発言が入力されるたびに、発言記録部２によるステップＳ１０１の処理が繰り返される。そして、会議が終了すると（ステップＳ１０２：Ｙｅｓ）、音声認識部３が、発言データ１３としてデータ蓄積部１０に蓄積された会議中の各発言に対して音声認識を行う（ステップＳ１０３）。音声認識部３による音声認識によって得られる各発言の音声認識結果および認識キーワードは、発言認識データ１４としてデータ蓄積部１０に格納される。なお、音声認識部３による会議中の発言に対する音声認識は、会議中に行われてもよい。
- [0022] 会議の終了後、クライアント端末から議事メモ作成の要求があると、発言提示装置１のＵＩ制御部４が、ＵＩ画面をクライアント端末に表示させる。そして、ユーザがこのＵＩ画面の「議事メモ」領域に議事メモを記入する操作を行うと（ステップＳ２０１）、そのテキストが「議事メモ」領域に表示されるとともに、記入された議事メモが、入力テキストデータ１５としてデ

一タ蓄積部10に格納される。

[0023] その後、ユーザが「議事メモ」領域に表示されているテキストから任意の文字列を指定する操作を行うと（ステップS202）、関連度算出部5が、会議中に収録された各発言に対し、指定された文字列との関連度を算出する（ステップS203）。そして、UI制御部4が、関連度算出部5により算出された関連度が高い発言を表示対象の発言として選択し、選択した発言の音声認識結果を、UI画面の「発言一覧」領域に表示させる（ステップS204）。議事メモを作成するユーザは、「発言一覧」領域に表示された発言の音声認識結果を参照することで、「議事メモ」領域で指定した文字列に対応する会議中の発言を、視覚を通じて確認することができる。また、議事メモを作成するユーザは、必要に応じて「発言一覧」領域に音声認識結果が表示されたいずれかの発言を指定し、その発言の音声を再生することにより、「議事メモ」領域で指定した文字列に対応する会議中の発言を、聴覚を通じて確認することもできる。

[0024] その後、議事メモ作成の終了を示すユーザの明示的な操作の有無などにより議事メモ作成が終了したか否かが判定され（ステップS205）、議事メモ作成が終了していなければ（ステップS205：No）、ステップS201からステップS204までの処理が繰り返される。そして、議事メモ作成が終了すると（ステップS205：Yes）、発言提示装置1による一連の動作が終了する。

[0025] 次に、具体的な会議の事例を例示しながら、本実施形態の発言提示装置1の動作について、さらに詳しく説明する。

[0026] 図3は、会議中の発言を収録するシーンの具体例を説明する図である。図3では、会議室内で“池田”、“山本”、“田中”の3名により会議が行われている様子を例示している。会議室の卓上には、本実施形態の発言提示装置1とネットワークを介して接続されるクライアントPC（パーソナルコンピュータ）20が設置されている。会議参加者のうち、“池田”と“山本”はそれぞれヘッドセット30を装着しており、“池田”の発言と“山本”の

発言は、それぞれヘッドセット 30 の個別マイクに入力される。また、会議の卓上には集音マイク 40 が設置されており、ヘッドセット 30 を装着していない“田中”の発言は、この集音マイク 40 に入力される。なお、集音マイク 40 は、ヘッドセット 30 を装着していない“田中”の発言だけでなく、ヘッドセット 30 を装着している“池田”や“山本”の発言も含めて、会議中に発生した音声をすべて入力している。

[0027] “池田”と“山本”が装着しているヘッドセット 30 や卓上に設置された集音マイク 40 は、クライアント PC 20 に接続されている。これらヘッドセット 30 や集音マイク 40 に入力された会議中の発言は、クライアント PC 20 からネットワークを介して発言提示装置 1 に送信される。なお、ここではクライアント端末の一例としてクライアント PC 20 を例示しているが、これに限らず、例えばタブレット端末やテレビ会議用の端末などの他の端末をクライアント端末として用いてもよい。

[0028] また、ここではすべての会議参加者が 1 つの会議室に集まって会議を行うシーンを想定しているが、地理的に離れた拠点間で遠隔会議を行う場合にも、本実施形態の発言提示装置 1 は有効に動作する。この場合、遠隔会議を行う各拠点に、本実施形態の発言提示装置 1 とネットワークを介して接続されるクライアント PC 20 のような端末をそれぞれ配置し、各拠点の会議参加者が装着するヘッドセット 30 や集音マイク 40 を各拠点の端末に接続すればよい。

[0029] 本実施形態の発言提示装置 1 を用いて会議中の発言を記録する場合、会議参加者のうち、少なくとも個別マイクを用いて発言を収録するユーザの登録と、開催される会議の登録が会議の開催前に行われる。ユーザの登録は、例えば、ユーザがクライアント PC 20 を用いて発言提示装置 1 にアクセスし、発言提示装置 1 からクライアント PC 20 に提供されるユーザ登録画面に名前を入力するといった簡単な方法で実現できる。登録されたユーザには固有のユーザ ID が付与され、入力された名前とともに、ユーザデータ 11 としてデータ蓄積部 10 に格納される。

[0030] 図4は、データ蓄積部10に格納されたユーザデータ11の具体例を示す図である。ユーザデータ11は、例えば図4に示すように、登録されたユーザのユーザIDと名前とを対応付けた形式でデータ蓄積部10に格納される。また、ユーザデータ11には、集音マイク40を用いて収録された発言を区別するために設けた特殊なユーザとして、“集音マイク”ユーザが含まれる。図4に示したユーザデータ11の例では、“集音マイク”ユーザのユーザIDは“－1__u”である。なお、図4の形式は一例であり、ユーザデータ11として、各ユーザが発言提示装置1にログインする際に用いるアカウント名およびパスワード、メールアドレスなどといった他の情報を含んでもよい。

[0031] 会議の登録は、例えば、会議参加者のうちの1人がクライアントPC20を用いて発言提示装置1にアクセスし、発言提示装置1からクライアントPC20に提供される会議設定画面に会議参加者の名前と会議のタイトルを入力するといった簡単な方法で実現できる。ヘッドセット30を装着しない会議参加者（図3の例では“田中”）の名前は、“集音マイク”が入力される。会議設定画面に入力された会議参加者の名前は、上述のユーザデータ11を用いてユーザIDに変換される。登録された会議には固有の会議IDが付与され、会議参加者のユーザIDおよび入力された会議のタイトルとともに、会議データ12としてデータ蓄積部10に格納される。

[0032] 図5は、データ蓄積部10に格納された会議データ12の具体例を示す図である。この図5の会議データ12の例では、2つの会議が登録されていることが示されており、会議参加者はユーザデータ11内のユーザIDで管理されている。このうち、会議IDが“1__c”の会議が図3の例に対応しており、会議参加者のユーザIDが“1__u”、“3__u”、“－1__u”であるので、会議には“池田”と“山本”が参加するほか、集音マイク40を用いた音声の収録も行われることが示されている。なお、図5の形式は一例であり、会議データ12として、会議のアジェンダや関連キーワード、開催日時などといった他の情報を含んでもよい。

[0033] 会議の登録後、実際に会議が開始されると、会議中の音声による発言が個別マイクや集音マイクに入力される。図6は、会議中の発言の具体例を示す図であり、図3に例示した環境で収録される発言例を示している。“池田”の発言と“山本”の発言は、各々が装着しているヘッドセット30の個別マイクに入力される。クライアントPC20では、予めそれぞれのヘッドセット30をどのユーザが利用するかが登録されており、ヘッドセット30の個別マイクに入力された発言は、そのヘッドセット30を利用するユーザのユーザIDとともに発言提示装置1に送信されるものとする。また、“田中”を含む3人の発言は集音マイク40に入力され、“集音マイク”ユーザのユーザIDとともに発言提示装置1に送信される。発言提示装置1では、クライアントPC20から受信した発言が発言記録部2により音声ファイルとして記録され、発言データ13としてデータ蓄積部10に格納される。

[0034] 図7は、データ蓄積部10に蓄積される発言データ13の具体例を示す図であり、図6の発言例に対応する発言データ13を示している。発言データ13は、例えば図7に示すように、各発言に付与された固有の発言IDと、その発言の発生日時と、発言ユーザのユーザIDと、発言を記録した音声ファイルのファイル名と、収録マイク種別と、発言が行われた会議の会議IDとを対応付けた形式でデータ蓄積部10に格納される。

[0035] 発言の発生日時は、発言に付加されてクライアントPC20から送信される情報であってもよいし、発言提示装置1において発言を受信した際に付与する情報であってもよい。収録マイク種別は、発言に付加されてクライアントPC20から送信されるユーザIDをもとに、例えばユーザデータ11を参照することで取得できる。また、会議IDは、登録された会議データ12から取得できる。

[0036] なお、収録マイク種別が“個別マイク”の発言は、無音区間やユーザからの明示的な発言開始、終了の入力操作などを元に、一文の発言ごとに分けて記録される。一方、収録マイク種別が“集音マイク”の発言は、例えば1分間といった予め定めた記録単位ごとにまとめて記録される。例えば図7に示

す発言 I D “6__s” の発言は、10:05:00から10:06:00の間に集音マイク40を用いて収録された発言である。なお、図7の形式は一例であり、発言データ13として他の情報を含んでいてもよい。

[0037] 会議の終了後、例えば会議参加者の操作に応じてクライアントPC20から発言提示装置1に会議の終了が通知されると、発言提示装置1の音声認識部3により発言の音声認識が行われる。そして音声認識部3が出力する各発言の音声認識結果および認識キーワードが、発言認識データ14としてデータ蓄積部10に格納される。なお、音声認識部3による発言の音声認識は、発言記録部2による発言の記録と合せて会議中に行われてもよい。

[0038] 図8は、発言認識データ14の具体例を示す図であり、図6の発言例に対応する発言認識データ14を示している。発言認識データ14は、例えば図8に示すように、各発言の発言 I D と、その発言に対する音声認識結果のテキスト（認識結果）と、認識キーワードと、想定される音声認識の精度を表す想定認識精度とを対応付けた形式でデータ蓄積部10に格納される。

[0039] 認識結果は、認識結果の候補のうちで尤度が最大となる候補のテキストである。説明を簡単にするため、図8に例示する認識結果はすべて音声認識が正しく行われた例を示している。しかし実際には、発言を収録する環境やユーザの話し方の影響などを受けて、認識結果に誤りが含まれている場合もある。なお、後述の想定認識精度が50%を下回る発言については、認識結果は保存されず、認識キーワードのみが保存される。例えば図8に示す発言 I D “6__s” の発言と発言 I D “12__s” の発言は、想定認識精度が50%を下回る30%であるため、認識結果は保存されず、認識キーワードのみが保存されている。

[0040] 認識キーワードは、認識結果の候補に含まれる単語を抽出したものである。認識キーワードの抽出方式としては、認識結果の候補に含まれる形態素情報から名詞のみを抽出するなどの方法がある。また、頻出する一般的な名詞を認識キーワードに含めないなどの方法を用いてもよい。なお、認識結果の候補から抽出された各認識キーワードは、対応する発言の開始時刻から何秒

経過した後にその認識キーワードが発言されたかを表す発言中出現時間と併せて格納されることが望ましい。

[0041] 想定認識精度は、音声認識部 3 による音声認識の精度を表す想定値である。音声認識の精度は音声の収録環境に依存するため、例えば収録マイク種別を用いて、ユーザの口元から個別に音声を入力する個別マイクには 80% といった高い値を設定し、口元から離れた位置で複数のユーザの発言が同時に入力される可能性がある集音マイクには 30% といった低い値を設定することができる。なお、想定認識精度を設定する方法はこれに限らず、音声認識の精度に関わる他の情報も加味して想定認識精度を設定してもよい。また、図 8 の形式は一例であり、発言認識データ 14 として他の情報を含んでもよい。また、発言認識データ 14 を発言データ 13 と併せてデータ蓄積部 10 に蓄積する構成であってもよい。

[0042] 会議の終了後、会議の議事メモを作成するユーザがクライアント PC 20 を用いて発言提示装置 1 にアクセスし、会議を指定して議事メモ作成を要求すると、発言提示装置 1 の UI 制御部 4 が、指定された会議に関連するデータをデータ蓄積部 10 から収集し、UI 画面を生成してクライアント PC 20 に提供する。発言提示装置 1 の UI 制御部 4 が提供する UI 画面は、クライアント PC 20 に表示される。

[0043] 図 9 は、クライアント PC 20 が表示する UI 画面の一例を示す図である。この図 9 に示す UI 画面 100 は、画面左側に「発言一覧」領域 110 を有し、画面右側に「議事メモ」領域 120 を有する画面構成となっている。「発言一覧」領域 110 には、発言認識データ 14 の認識結果 111 が、発言の発生順に時系列で上から下方向に表示される。「発言一覧」領域 110 の左端に配置されたバー 112 は、集音マイク 40 を用いて収録された会議全体の音声を表し、その右側に配置された色分けされたバー 113 は、ヘッドセット 30 の個別マイクを用いて収録された各ユーザの発言を表している。これらのバー 112, 113 をクリックすることで、クリックした場所に対応する時間から音声再生される構成となっている。

[0044] また、「議事メモ」領域１２０は、議事メモを作成するユーザが任意のテキストを入力する領域である。「議事メモ」領域１２０には、テキスト入力用のカーソル１２１が配置されている。ユーザの要求に応じて最初に表示されるＵＩ画面１００では、図９に示すように、「議事メモ」領域１２０には何も表示されていない。ただし、会議データ１２として会議のアジェンダが登録されている場合には、そのアジェンダの内容が初期テキストとして「議事メモ」領域１２０に表示されていてもよい。議事メモを作成するユーザは、例えば会議中にメモ帳などに記載した手書きのメモなどを参照して、任意の文字列をこの「議事メモ」領域１２０に議事メモとして記入することができる。「議事メモ」領域１２０に記入された議事メモは、この「議事メモ」領域１２０にテキストとして表示される。なお、「議事メモ」領域１２０への議事メモの記入は、会議中に行われてもよい。すなわち、会議中にクライアントＰＣ２０にＵＩ画面１００を表示させ、会議を行いながらキーボードなどを用いて「議事メモ」領域１２０に議事メモを直接入力することもできる。

[0045] ＵＩ画面１００の「議事メモ」領域１２０に記入された議事メモは、例えば、行単位の入力テキストデータ１５として、データ蓄積部１０に格納される。図１０は、「議事メモ」領域１２０に議事メモが記入されたＵＩ画面１００を示す図である。また、図１１は、データ蓄積部１０に格納される入力テキストデータ１５の具体例を示す図であり、図１０の議事メモに対応する入力テキストデータ１５を示している。入力テキストデータ１５は、例えば図１１に示すように、固有のメモＩＤと、そのテキストが記入された行と、テキストの内容と、議事メモ作成の対象となる会議の会議ＩＤとを対応付けた形式でデータ蓄積部１０に格納される。なお、図１１の形式は一例であり、入力テキストデータ１５として他の情報を含んでいてもよい。

[0046] ＵＩ画面１００の「議事メモ」領域１２０に議事メモを記入した後、ユーザがカーソル１２１を移動させるなどの操作を行って「議事メモ」領域１２０に表示されている任意の文字列を指定する操作を行うと、発言提示装置１

の関連度算出部 5 が、会議中に記録された発言のうち、発言認識データ 14 に認識結果が含まれる各発言に対し、指定された文字列との関連度を各々算出する。そして、UI 制御部 4 が、例えば、関連度算出部 5 により算出された関連度が高い順に所定数の発言を表示対象の発言として選択し、選択した発言の音声認識結果を UI 画面 100 の「発言一覧」領域 110 に表示させる制御を行う。

[0047] 図 10 の UI 画面 100 の例では、「議事メモ」領域 120 に記入された議事メモのうち、“音声合成は？関連技術？”というメモ部分が記入された行にカーソル 121 が合っており、この行をテキスト解析することで得られる名詞である“音声合成”および“関連技術”が、指定された文字列となる。この場合、関連度算出部 5 は、発言認識データ 14 に認識結果が含まれる各発言に対し、“音声合成”や“関連技術”との関連度を各々算出する。そして、UI 制御部 4 は、図 10 に示すように、会議中の発言のうち、“音声合成”や“関連技術”との関連度が高い発言の音声認識結果を「発言一覧」領域 110 に表示させる。

[0048] なお、「議事メモ」領域 120 上で文字列を指定する操作方法は、カーソル 121 を合せる方法に限らない。例えば、マウスのドラッグ操作による範囲指定といった他の操作方法による文字列の指定を受け付ける構成であってもよい。

[0049] また、UI 制御部 4 は、想定認識精度が 50%を下回るために発言認識データ 14 に認識結果が含まれていない発言について、認識キーワードとして保存されている単語のうちで、指定された文字列の少なくとも一部を含む単語を、表示対象として選択した発言の音声認識結果とともに、「発言一覧」領域 110 に表示させる。この単語の表示位置は、会議中における音声の発生時刻に基づいて決定される。すなわち、UI 制御部 4 は、想定認識精度が 50%を下回る発言の発言認識データ 14 に含まれる認識キーワードのうち、指定された文字列の少なくとも一部を含む認識キーワードを、上述した発言中出现時間を用いて、その認識キーワードが発言された時間に相当する「

発言一覧」領域 110 上の位置に表示させる。ただし、その位置に関連度が高い発言の音声認識結果が表示される場合は、認識キーワードの表示は行われない。

[0050] 図 10 の UI 画面 100 の例は、図 3 に示した集音マイク 40 を用いて収録される“田中”の発言の音声認識結果は「発言一覧」領域 110 に表示されないが、“田中”の発言に含まれる“音声合成”や“関連技術”といった認識キーワード 114 が表示されていることを示している。これは、図 8 に示した発言認識データ 14 の例において、発言 ID “12__s”の認識キーワードのうち、ユーザが「議事メモ」領域 120 上で指定した“音声合成”や“関連技術”と一致するものを抽出し、その認識キーワードの発言中出现時間を元に「発言一覧」領域 110 に表示させたものである。なお、このような認識キーワードが存在しておらず、かつ、個別マイクによる発言がない時間については、図 10 のように「・・・」などを表示することで、発言は記録されているが音声認識結果は表示していないことを示すことが望ましい。

[0051] また、図 10 の UI 画面 100 の例では、発言認識データ 14 に認識結果が含まれる発言のうち、関連度算出部 5 により算出された関連度が低い発言の音声認識結果は表示されないようにしているが、関連度が低い発言についても、その発言の音声認識結果の先頭部分のみを「発言一覧」領域 110 に表示させるようにしてもよい。

[0052] ここで、関連度算出部 5 による関連度の算出方法の具体例について説明する。関連度算出部 5 は、例えば以下の手順で、指定された文字列に対する各発言の関連度を算出する。まず、発言認識データ 14 に含まれる各発言の認識結果のテキスト、および「議事メモ」領域 120 上で指定された文字列を、形態素解析を用いて単語に分割する。その後、分割された各単語に対して、発言認識データ 14 に含まれる各発言の認識結果のテキスト全体をコーパスとし、各発言の認識結果のテキストをドキュメントとした tf (Term Frequency) – idf (Inverse Document Frequency) を用いて重みを設定す

る。そして、各発言の認識結果のテキストと、「議事メモ」領域１２０上で指定された文字列それぞれに対して、 $t f - i d f$ の重みを付加した単語の出現ベクトルを生成し、各発言について生成した単語の出現ベクトルと、「議事メモ」領域１２０上で指定された文字列について生成した単語の出現ベクトルとのコサイン類似度を算出する。その後、各発言のコサイン類似度に対し、その発言の前後の一定数の発言のコサイン類似度を加算したものを、「議事メモ」領域１２０上で指定された文字列に対するその発言の関連度として算出する。なお、前後の発言のコサイン類似度を加算せずに、各発言のコサイン類似度を関連度として算出するようにしてもよい。また、各発言の単語の出現ベクトルは、その発言の認識結果に含まれる単語だけでなく、認識結果の候補に含まれる単語（認識キーワード）も含めて生成してもよい。

[0053] 関連度を以上の方法で算出する場合、UI制御部４は、発言認識データ１４に認識結果が含まれる各発言を、関連度算出部５により算出された関連度が高い順にソートして、上位の所定数の発言を表示対象として選択する。そして、UI制御部４は、表示対象として選択した発言の音声認識結果を、その発言の発生順に応じた時系列で、UI画面１００の「発言一覧」領域１１０に表示させる。

[0054] また、関連度算出部５は、以上のように単語に対する $t f - i d f$ の重み付けを行わず、単純に、「議事メモ」領域１２０上で指定された文字列が認識結果のテキストに含まれるか否かにより、各発言の関連度を算出するようにしてもよい。この場合、関連度算出部５により算出される関連度は、「議事メモ」領域１２０上で指定された文字列が認識結果のテキストに含まれることを示す“１”、含まれないことを示す“０”といった２値の値となる。UI制御部４は、関連度算出部５により算出される関連度が“１”となった発言を表示対象として選択し、その発言の音声認識結果を、その発言の発生順に応じた時系列で、UI画面１００の「発言一覧」領域１１０に表示させる。

[0055] 議事メモを作成するユーザは、UI画面１００の「発言一覧」領域１１０

に表示された発言の音声認識結果を参照し、必要に応じて、その音声認識結果に対応する発言の音声を再生させることにより、「議事メモ」領域 110 に記入した議事メモに関連する発言の内容を確認することができ、不足する情報を新たに追加するといった議事メモの拡充などを効率よく行うことができる。

[0056] 以上、具体的な例を挙げながら詳細に説明したように、本実施形態の発言提示装置 1 では、会議中に収録された発言を発言記録部 2 が記録し、音声認識部 3 がその発言の音声認識を行う。そして、UI 制御部 4 が「発言一覧」領域 110 と「議事メモ」領域 120 とを含む UI 画面 100 をクライアント端末に表示させ、「議事メモ」領域 120 上で文字列が指定されると、関連度算出部 5 が、音声認識された各発言に対して「議事メモ」領域 120 上で指定された文字列との関連度を算出する。そして、UI 制御部 4 が、関連度算出部 5 により算出された関連度の高い発言を表示対象として選択し、選択した発言の音声認識結果を UI 画面 100 の「発言一覧」領域 110 に表示させる。したがって、この発言提示装置 1 によれば、「議事メモ」領域 120 に入力された任意の情報に対応する発言をユーザに分かり易く提示して確認させることができ、議事メモ作成などのユーザの作業を適切に支援することができる。

[0057] <第 2 実施形態>

次に、第 2 実施形態について説明する。本実施形態は、関連度算出部 5 による関連度の算出方法が上述した第 1 実施形態と異なる。発言提示装置 1 の基本的な構成や動作は第 1 実施形態と同様であるため、以下では第 1 実施形態と共通部分については重複した説明を省略し、第 1 実施形態との相違点のみを説明する。

[0058] 本実施形態の関連度算出部 5 は、認識結果のテキストのみを用いて各発言の関連度を算出するのではなく、その会議に関連する様々な文書を用いてテキストのトピックを算出し、算出したトピックを用いて関連度を算出する。ここでトピックとは、そのテキストの大まかな意味合いを示し、例えば LD

A (Latent Dirichlet Allocation) などのトピック解析手法を用いて算出される。

[0059] 図12は、本実施形態の発言提示装置1の構成例を示すブロック図である。図1に示した第1実施形態の発言提示装置1の構成との違いは、データ蓄積部10に蓄積されるデータとして会議関連文書データ16が追加され、関連度算出部5が、この会議関連文書データ16を用いて各発言の関連度を算出する点である。会議関連文書データ16は、例えば、ある会議について、データ蓄積部10に蓄積されている他の関連する会議の発言認識データ14や入力テキストデータ15を集約したデータである。なお、会議関連文書データ16として、例えば、インターネット上からクロールした、会議に関連する話題の文書を用いてもよい。

[0060] 本実施形態の関連度算出部5は、例えば以下の手順で、指定された文字列に対する各発言の関連度を算出する。まず、発言認識データ14に含まれる各発言の認識結果のテキスト、および「議事メモ」領域120上で指定された文字列を、形態素解析を用いて単語に分割する。その後、各発言の認識結果のテキストと、「議事メモ」領域120上で指定された文字列それぞれに対して、発言認識データ14に含まれる各発言の認識結果のテキスト全体と会議関連文書データ16とをコーパスとして、LDAなどを用いてトピックを表す単語とその重みの列からなるベクトルを生成し、各発言について生成したベクトルと、「議事メモ」領域120上で指定された文字列について生成したベクトルとのコサイン類似度を算出する。その後、各発言のコサイン類似度に対し、その発言の前後の一定数の発言のコサイン類似度を加算したものを、「議事メモ」領域120上で指定された文字列に対するその発言の関連度として算出する。なお、前後の発言のコサイン類似度を加算せずに、各発言のコサイン類似度を関連度として算出するようにしてもよい。また、トピックの算出には、LDA以外の手法、例えばLSI (Latent Semantic Indexing) などを用いてもよい。

[0061] 以上説明したように、本実施形態では、関連度算出部5が、各発言のトピ

ックと指定された文字列のトピックとの類似度を用いて、指定された文字列に対する各発言の関連度を算出する。このため、上述した第1実施形態と比べて、指定された文字列に対する各発言の関連度をより精度よく算出することができる。

[0062] <第3実施形態>

次に、第3実施形態について説明する。本実施形態は、U I画面100の「議事メモ」領域120上で指定された文字列に対応する発言の音声認識結果だけでなく、「議事メモ」領域120に表示されている文字列の構造に基づいて選択された文字列に対応する発言の音声認識結果も併せてU I画面100の「発言一覧」領域110に表示させる例である。発言提示装置1の基本的な構成や動作は第1実施形態と同様であるため、以下では第1実施形態と共通部分については重複した説明を省略し、第1実施形態との相違点のみを説明する。

[0063] 例えば、「議事メモ」領域120の任意の行にカーソル121を合せるといった方法で文字列を指定する場合、第1実施形態では、「議事メモ」領域120中に表示されている文字列のうち、カーソル121の合っている行の文字列に対応する発言の音声認識結果を「発言一覧」画面110に表示させるようにしている。これに対し、本実施形態では、「議事メモ」領域120のインデントを用いてテキスト構造を把握し、カーソル121の合っている行の話題の上位レベルの見出し語についても、対応する発言の音声認識結果を「発言一覧」領域110に表示させる。

[0064] 図13は、本実施形態においてクライアントPC20に表示されるU I画面100の一例を示す図である。図13のU I画面100の例では、「議事メモ」領域120に記入された議事メモのうち、“保守業務”というメモ部分が記入された行にカーソル121が合っており、この“保守業務”が指定された文字列となる。また、“保守業務”が記入された行は、先頭にスペース1文字分のインデントが設定されているのに対し、2行上の“展示会”というメモ部分が記入された行122は先頭にインデントが設定されておらず

、この行 122 の“展示会”という文字列が、指定された文字列である“保守業務”よりも上位の見出し語になっていると推定される。

[0065] この場合、関連度算出部 5 は、発言認識データ 14 に認識結果が含まれる各発言に対し、指定された文字列である“保守業務”との関連度に加えて、“展示会”との関連度も算出する。そして、UI 制御部 4 は、図 13 に示すように、会議中の発言のうち、“保守業務”との関連度が高い発言の音声認識結果と併せて、“展示会”との関連度が高い発言の音声認識結果についても、「発言一覧」領域 110 に時系列で表示させる。また、第 1 実施形態と同様に、想定認識精度が低いために認識結果が保存されていない発言の認識キーワードの中に“保守業務”や“展示会”が含まれている場合は、その認識キーワードが発言された時間に対応する位置に表示させる。

[0066] なお、「議事メモ」領域 120 上の文字列と「発言一覧」領域 110 に表示された発言の音声認識結果との対応関係を明確にするため、例えば、「議事メモ」領域 120 上の指定された文字列とそれに対応する「発言一覧」領域 110 上の発言の音声認識結果の背景を同色に色づけして表示するとともに、「議事メモ」領域 120 上のテキスト構造に基づいて選択された文字列とそれに対応する「発言一覧」領域 110 上の発言の音声認識結果の背景を同色に色づけして表示するといった方法を用いることが望ましい。図 13 の UI 画面 100 の例では、「議事メモ」領域 120 上の“保守業務”が記入された行と、「発言一覧」領域 110 の“保守業務”対応する発言の音声認識結果および認識キーワードが同色の背景上に表示され、「議事メモ」領域 120 上の“展示会”が記入された行と、「発言一覧」領域 110 の“展示会”対応する発言の音声認識結果が同色の背景上に表示されている。

[0067] 以上説明したように、本実施形態では、ユーザにより指定された文字列に対応する発言の音声認識結果だけでなく、その文字列の上位の見出し語などに対応する発言の音声認識結果も提示するので、例えば議事メモ作成などのユーザの作業をより適切に支援することができる。

[0068] <第 4 実施形態>

次に、第４実施形態について説明する。本実施形態は、想定認識精度の設定において、収録マイク種別だけでなく、いくつかの録音環境データを用意しておき、会議、ユーザごとに個別の設定を行う例である。発言提示装置１の基本的な構成や動作は第１実施形態と同様であるため、以下では第１実施形態と共通部分については重複した説明を省略し、第１実施形態との相違点のみを説明する。

[0069] 図１４は、本実施形態の発言提示装置１の構成例を示すブロック図である。図１に示した第１実施形態の発言提示装置１の構成との違いは、録音環境データ１７が追加され、音声認識部３が、この録音環境データ１７を参照して各発言の推定認識精度を設定している点である。録音環境データ１７では、収録マイク種別のほかに、特定のユーザの発言であるか、特定の場所で収録した発言であるか、収録した発言の音声に対して後処理を行ったかなどの条件ごとに、想定認識精度が定められている。

[0070] 図１５は、録音環境データ１７の具体例を示す図である。録音環境データ１７は、例えば図１５に示すように、個々のデータに付与された固有のデータＩＤと、収録マイク種別と、発言ユーザのユーザＩＤと、発言が収録された場所と、後処理の有無と、想定認識精度とを対応付けた形式とされる。図１５の録音環境データ１７の例において、内容が“*”となっている項目は、発言ユーザや発言が収録された場所を特定しない設定を示している。“話者照合”は、集音マイク４０を用いて収録された音声を、各話者の音声の音響的な特徴を用いてそれぞれの話者ごとに分離する後処理を示している。なお、図１５の形式は一例であり、録音環境データ１７として他の情報を含んでいてもよい。

[0071] 本実施形態の音声認識部３は、発言認識データ１４の想定認識精度を設定する際に、以上のような録音環境データ１７を利用する。各発言がどの条件に該当するかは、会議登録時に会議設定画面を用いて登録された会議に関する会議データ１２や、その会議中に収録された発言の発言データ１３などを用いて特定される。

[0072] 図16は、会議設定画面の一例を示す図である。この図16に示す会議設定画面200には、会議のタイトルを入力するためのテキストボックス201、会議が行われる場所（発言が収録される場所）を入力するためのテキストボックス202、会議の出席者（会議参加者）を入力するためのテキストボックス203、およびその出席者の発言の収録に用いるマイクの種別（収録マイク種別）を入力するためのテキストボックス204が設けられている。

[0073] 図16の会議設定画面200の例では、会議が行われる場所（発言が収録される場所）が“サーバ室”であることが示されている。このため、図15に例示した録音環境データ17のうち、データIDが“4__d”の条件に該当し、想定認識精度は“60%”に設定される。これは、サーバ室のように騒音の多い環境で収録された発言の音声認識精度は、騒音の少ない環境で収録された発言の音声認識精度よりも低くなることが想定されるため、個別マイクを用いて収録された発言の想定認識精度が80%から60%に下がることを示している。

[0074] なお、録音環境データ17に含まれる複数のデータの条件に合致する場合は、これら複数のデータで示される想定認識精度のうち、最も低い想定認識精度が設定される。例えば、図16の会議設定画面200の例では、ユーザIDが“2__u”の“大島”が会議に出席することが示されているため、この会議における“大島”の発言については、図15に例示した録音環境データ17のうち、データIDが“3__d”の条件と、データIDが“4__d”の条件との双方に合致する。この場合、データIDが“3__d”の想定認識精度である90%と、データIDが“4__d”の想定認識精度である60%とを比較し、低い方の60%が“大島”の発言の想定認識精度として設定される。

[0075] 以上説明したように、本実施形態では、収録マイク種別だけでなく、発言の収録に関わる様々な条件を考慮して想定認識精度を設定するので、想定認識精度をより精度よく設定することができる。

[0076] なお、以上のように設定される想定認識精度は、第1実施形態で説明したように、発言認識データ14として認識結果を保存するか否かの判定に用いることに加えて、UI制御部4がUI画面100の「発言一覧」領域110に認識結果を表示させる対象となる発言を選択するために用いることもできる。すなわち、UI制御部4は、関連度算出部5により算出された指定された文字列との関連度に加えて、音声認識部3により設定された想定認識精度を用いて、「発言一覧」領域110に認識結果を表示させる対象となる発言を選択するようにしてもよい。

[0077] 具体的には、UI制御部4は、例えば、第1実施形態や第2実施形態で説明した算出方法で関連度算出部5が算出した関連度に対し、音声認識部3が設定した想定認識精度を乗算した値を各発言のスコアとして求め、得られたスコアが大きい順に各発言をソートして、上位の所定数の発言を表示対象として選択する。そして、UI制御部4は、表示対象として選択した発言の音声認識結果を、その発言の発生順に応じた時系列で、UI画面100の「発言一覧」領域110に表示させる。これにより、指定された文字列との関連度が高い発言の中でも特に想定認識精度が高い発言を優先してユーザに提示することができる。なお、音声認識精度が極端に低い発言については、指定された文字列と一致する認識キーワードの表示を行わないようにしてもよい。

[0078] <補足説明>

以上、実施形態の発言提示装置として、会議中の発言を記録して、ユーザにより指定された任意の文字列に対応する発言を提示する構成の発言提示装置1を例示したが、実施形態の発言提示装置は会議中の発言に限らず、音声による様々な発言を記録して、ユーザにより指定された任意の文字列に対応する発言を提示する装置として構成することができる。

[0079] 以上説明した実施形態の発言提示装置1における各機能的な構成要素は、例えば、汎用のコンピュータシステムを基本ハードウェアとして用いて実行されるプログラム（ソフトウェア）により実現することができる。

- [0080] 図17は、発言提示装置1のハードウェア構成の一例を概略的に示すブロック図である。実施形態の発言提示装置1は、図17に示すように、CPUなどのプロセッサ51と、RAMなどの主記憶装置52と、各種の記憶装置を用いた補助記憶装置53と、通信インタフェース54と、これらの各部を接続するバス55とを含んだ汎用のコンピュータシステムとして構成される。なお、補助記憶装置53は、有線または無線によるLAN (Local Area Network) などで各部に接続されてもよい。
- [0081] 実施形態の発言提示装置1の各機能的な構成要素（発言記録部2、音声認識部3、UI制御部4および関連性算出部5）は、例えば、プロセッサ51が、主記憶装置52を利用して、補助記憶装置53に格納されたプログラムを実行することによって実現される。データ蓄積部10は、例えば、補助記憶装置53を用いて実現される。
- [0082] プロセッサ51により実行されるプログラムは、例えば、インストール可能な形式又は実行可能な形式のファイルでCD-ROM (Compact Disc Read Only Memory)、フレキシブルディスク (FD)、CD-R (Compact Disc Recordable)、DVD (Digital Versatile Disc) などのコンピュータで読み取り可能な記録媒体に記録されてコンピュータプログラムプロダクトとして提供される。
- [0083] また、このプログラムを、インターネットなどのネットワークに接続された他のコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、このプログラムをインターネットなどのネットワーク経由で提供または配布するように構成してもよい。また、このプログラムを、コンピュータ内部のROM（補助記憶装置53）などに予め組み込んで提供するように構成してもよい。
- [0084] このプログラムは、実施形態の発言提示装置1の機能的な構成要素を含むモジュール構成となっており、実際のハードウェアとしては、例えば、プロセッサ51が上記記録媒体からプログラムを読み出して実行することにより、上記の各構成要素が主記憶装置52上にロードされ、上記の各構成要素が

主記憶装置 52 上に生成されるようになっている。なお、実施形態の発言提示装置 1 の機能的な構成要素は、その一部または全部を、A S I C (Application Specific Integrated Circuit) や F P G A (Field-Programmable Gate Array) などの専用のハードウェアを用いて実現することも可能である。

[0085] 以上、本発明の実施形態を説明したが、この実施形態は例として提示したものであり、発明の範囲を限定することは意図していない。この新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、請求の範囲に記載された発明とその均等の範囲に含まれる。

請求の範囲

- [請求項1] 音声による発言を記録する発言記録部と、
記録された発言を音声認識する音声認識部と、
音声認識された各発言に対し、第1表示領域と第2表示領域とを有するUI画面の前記第2表示領域に表示されている文字列のうち指定された文字列との関連度を各々算出する関連度算出部と、
前記関連度の高さに基づいて選択された発言の音声認識結果を、前記UI画面の前記第1表示領域に表示させるUI制御部と、を備える発言提示装置。
- [請求項2] 前記UI制御部は、前記選択された発言の音声認識結果を、発言の発生順に応じた時系列で前記第1表示領域に表示させる、請求項1に記載の発言提示装置。
- [請求項3] 前記UI制御部は、音声の入力方式に基づいて想定される音声認識の精度が所定の基準を満たす発言であって、前記関連度の高さに基づいて選択された発言の音声認識結果を、前記第1表示領域に表示させる、請求項1または2に記載の発言提示装置。
- [請求項4] 前記UI制御部は、前記精度が前記基準を満たさない発言の音声認識結果の候補に含まれる単語のうち、前記指定された文字列の少なくとも一部を含む単語を、前記選択された発言の音声認識結果とともに前記第1表示領域に表示させる、請求項3に記載の発言提示装置。
- [請求項5] 前記UI制御部は、前記第1表示領域における前記単語の表示位置を、該単語に対応する音声の発生時刻に基づいて決定する、請求項4に記載の発言提示装置。
- [請求項6] 前記精度は、音声の入力方式に加えてさらに、音声の入力環境および音声の後処理の有無の少なくとも一方に基づいて想定される、請求項3乃至5のいずれか一項に記載の発言提示装置。
- [請求項7] 前記UI制御部は、前記関連度の高さと前記精度とに基づいて選択された発言の音声認識結果を、前記第1表示領域に表示させる、請求

項3乃至6のいずれか一項に記載の発言提示装置。

[請求項8] 前記指定された文字列は、前記第2表示領域に対するユーザの操作に基づいて指定された文字列である、請求項1乃至7のいずれか一項に記載の発言提示装置。

[請求項9] 前記関連度算出部は、音声認識された各発言に対し、前記指定された文字列との関連度と、前記第2表示領域に表示されている文字列の構造に基づいて選択された文字列との関連度とを各々算出し、

前記UI制御部は、前記指定された文字列との関連度の高さに基づいて選択された発言の音声認識結果と、前記選択された文字列との関連度の高さに基づいて選択された発言の音声認識結果とを、前記第1表示領域に表示させる、請求項1乃至8のいずれか一項に記載の発言提示装置。

[請求項10] 前記UI制御部は、前記第1表示領域に表示された音声認識結果を指定する操作に応じて、該音声認識結果に対応する発言の音声を再生させる、請求項1乃至9のいずれか一項に記載の発言提示装置。

[請求項11] 前記関連度算出部は、前記指定された文字列の少なくとも一部が、発言の音声認識結果または該音声認識結果の候補に含まれるか否かに基づいて、前記指定された文字列に対する該発言の前記関連度を算出する、請求項1乃至10のいずれか一項に記載の発言提示装置。

[請求項12] 前記関連度算出部は、前記指定された文字列について、該文字列に含まれる各単語に $t f - i d f$ を用いた重みを付加した単語の出現ベクトルを生成するとともに、音声認識された各発言について、発言の音声認識結果に含まれる各単語に $t f - i d f$ を用いた重みを付加した単語の出現ベクトルを生成し、各発言について生成した単語の出現ベクトルと、前記指定された文字列について生成した単語の出現ベクトルとのコサイン類似度に基づいて、前記指定された文字列に対する各発言の前記関連度を算出する、請求項1乃至10のいずれか一項に記載の発言提示装置。

- [請求項13] 前記関連度を算出する対象となる発言を対象発言とし、該対象発言に対して発生時刻が近い所定数の発言をそれぞれ近傍発言としたときに、前記関連度算出部は、前記対象発言について生成した単語の出現ベクトルと、前記指定された文字列について生成した単語の出現ベクトルとのコサイン類似度に対し、前記近傍発言について生成した単語の出現ベクトルと、前記指定された文字列について生成した単語の出現ベクトルとのコサイン類似度を加算して、前記関連度を算出する、請求項12に記載の発言提示装置。
- [請求項14] 前記関連度算出部は、前記指定された文字列について、該文字列のトピックを表す単語と該単語の重みの列からなるベクトルを生成するとともに、音声認識された各発言について、該発言のトピック表す単語と該単語の重みの列からなるベクトルを生成し、各発言について生成したベクトルと、前記指定された文字列について生成したベクトルとのコサイン類似度に基づいて、前記指定された文字列に対する各発言の前記関連度を算出する、請求項1乃至10のいずれか一項に記載の発言提示装置。
- [請求項15] 前記関連度を算出する対象となる発言を対象発言とし、該対象発言に対して発生時刻が近い所定数の発言をそれぞれ近傍発言としたときに、前記関連度算出部は、前記対象発言について生成したベクトルと、前記指定された文字列について生成したベクトルとのコサイン類似度に対し、前記近傍発言について生成したベクトルと、前記指定された文字列について生成したベクトルとのコサイン類似度を加算して、前記関連度を算出する、請求項14に記載の発言提示装置。
- [請求項16] 発言提示装置により実行される発言提示方法であって、
音声による発言を記録するステップと、
記録された発言を音声認識するステップと、
音声認識された各発言に対し、第1表示領域と第2表示領域とを有するUI画面の前記第2表示領域に表示されている文字列のうち指定

された文字列との関連度を各々算出するステップと、

前記関連度の高さに基づいて選択された発言の音声認識結果を、前記U I画面の前記第1表示領域に表示させるステップと、を含む発言提示方法。

[請求項17]

コンピュータに、

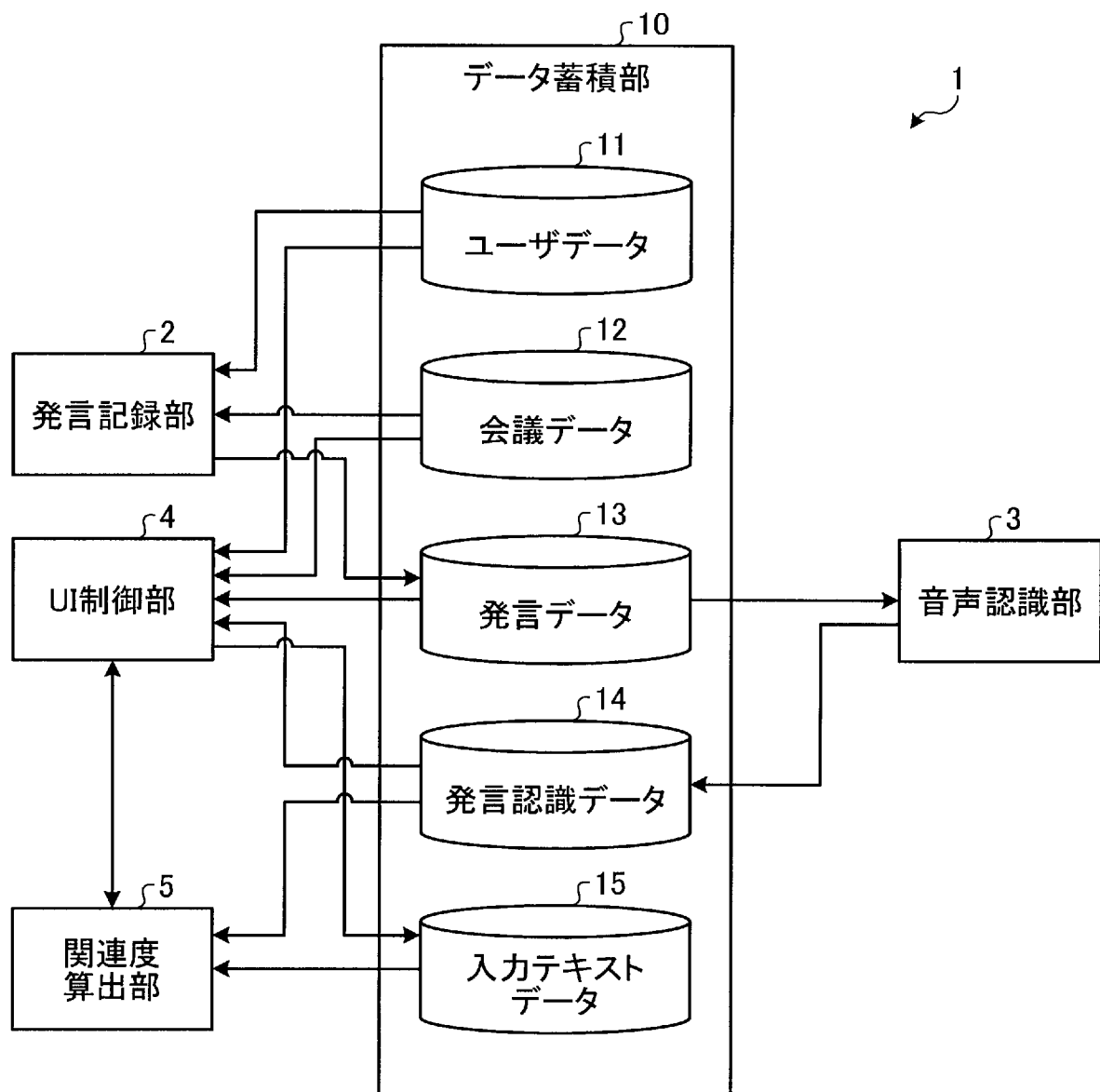
音声による発言を記録する機能と、

記録された発言を音声認識する機能と、

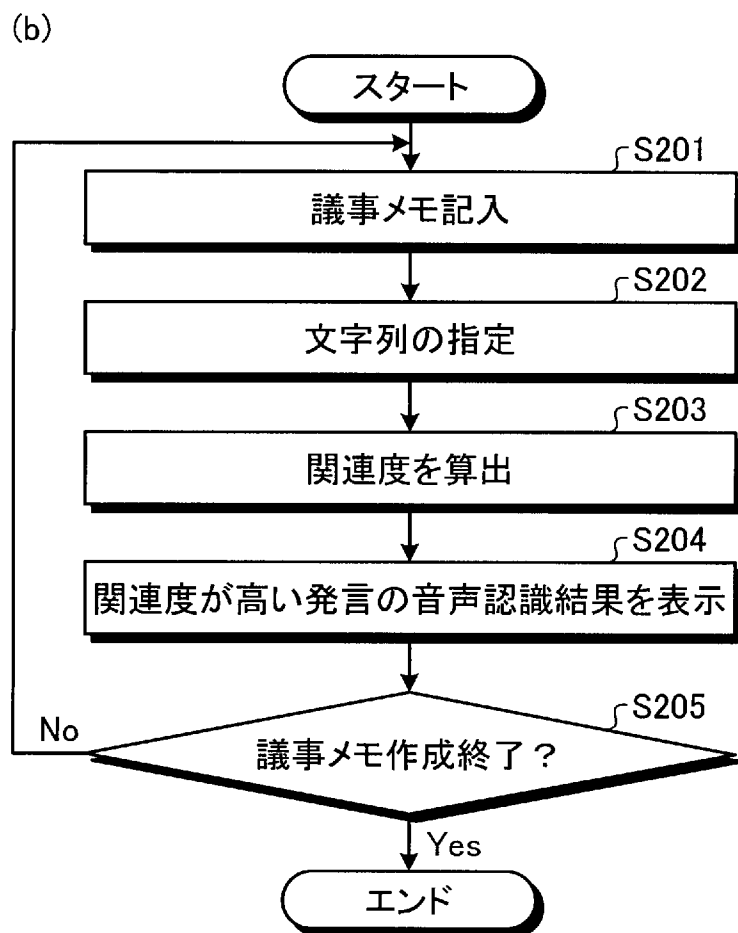
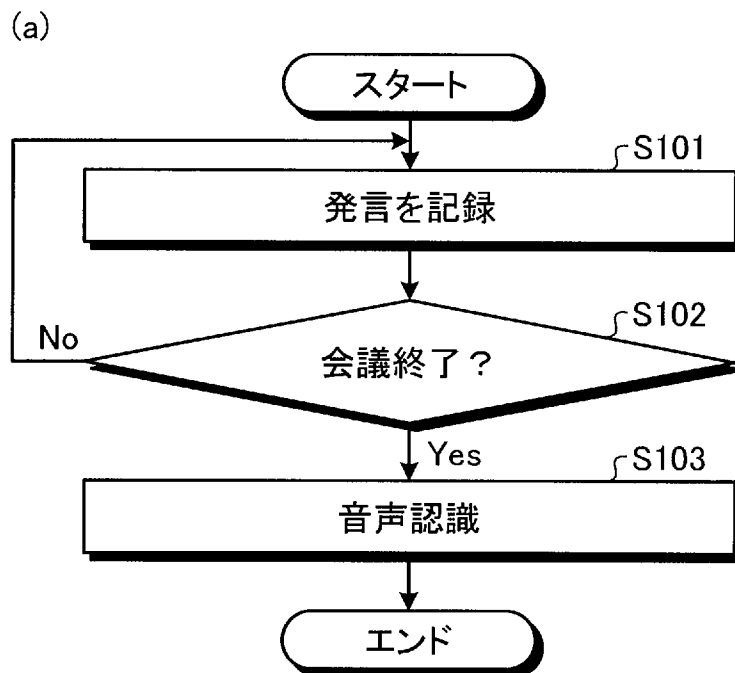
音声認識された各発言に対し、第1表示領域と第2表示領域とを有するU I画面の前記第2表示領域に表示されている文字列のうち指定された文字列との関連度を各々算出する機能と、

前記関連度の高さに基づいて選択された発言の音声認識結果を、前記U I画面の前記第1表示領域に表示させる機能と、を実現させるためのプログラム。

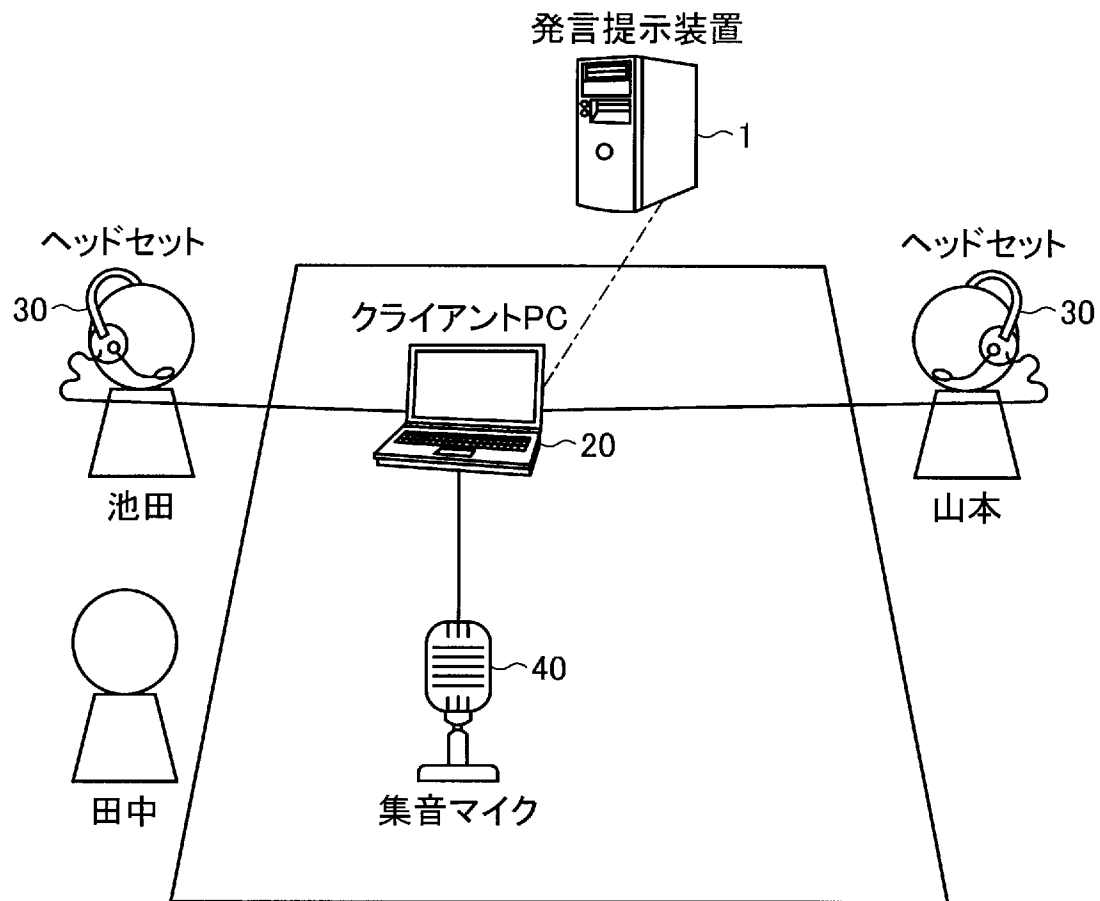
[図1]



[図2]



[図3]



[図4]

11
↙

ユーザID	名前
-1_u	集音マイク
1_u	池田
2_u	田中
3_u	山本
⋮	⋮

[図5]

12
↙

会議ID	参加者	会議ID
1_c	1_u, 3_u, -1_u	音声認識技術のデモ実施について
2_c	2_u, 3_u	会議向け音声認識技術開発
⋮	⋮	⋮

[図6]

池田:

来週の展示会で音声認識技術のデモを実施します。

詳細を田中さんから説明してもらえますか。

田中:

保守業務での点検項目を口頭で記録して、それが音声認識される様子をデモしようと思います。

ブースにPCを一台置いて、そこでデモします。

池田:

PCでの保守業務向けのデモの実施ですね。

デモの内容について何か質問はありますか。

山本:

今回は音声合成のデモは行わないのですか。

池田:

音声合成のデモはどうでしょう。

田中:

今回は行いません。

でも音声合成についても、関連技術ということで説明したいですね。

山本:

音声合成の手持ち資料を用意しましょうか。

池田:

そうですね、音声合成は手持ち資料で説明することにしましょう。

山本:

音声の入力はヘッドセットからですか。

田中:

確実に音声を入力するには、そうした方が良さそうですね。

池田:

音声入力用のヘッドセットの準備も必要ですね。

池田:

ではアクションアイテムをまとめましょう。

山本さんは音声合成の手持ち資料の印刷を。

田中さんはヘッドセットの準備をお願いします。

[図7]

13

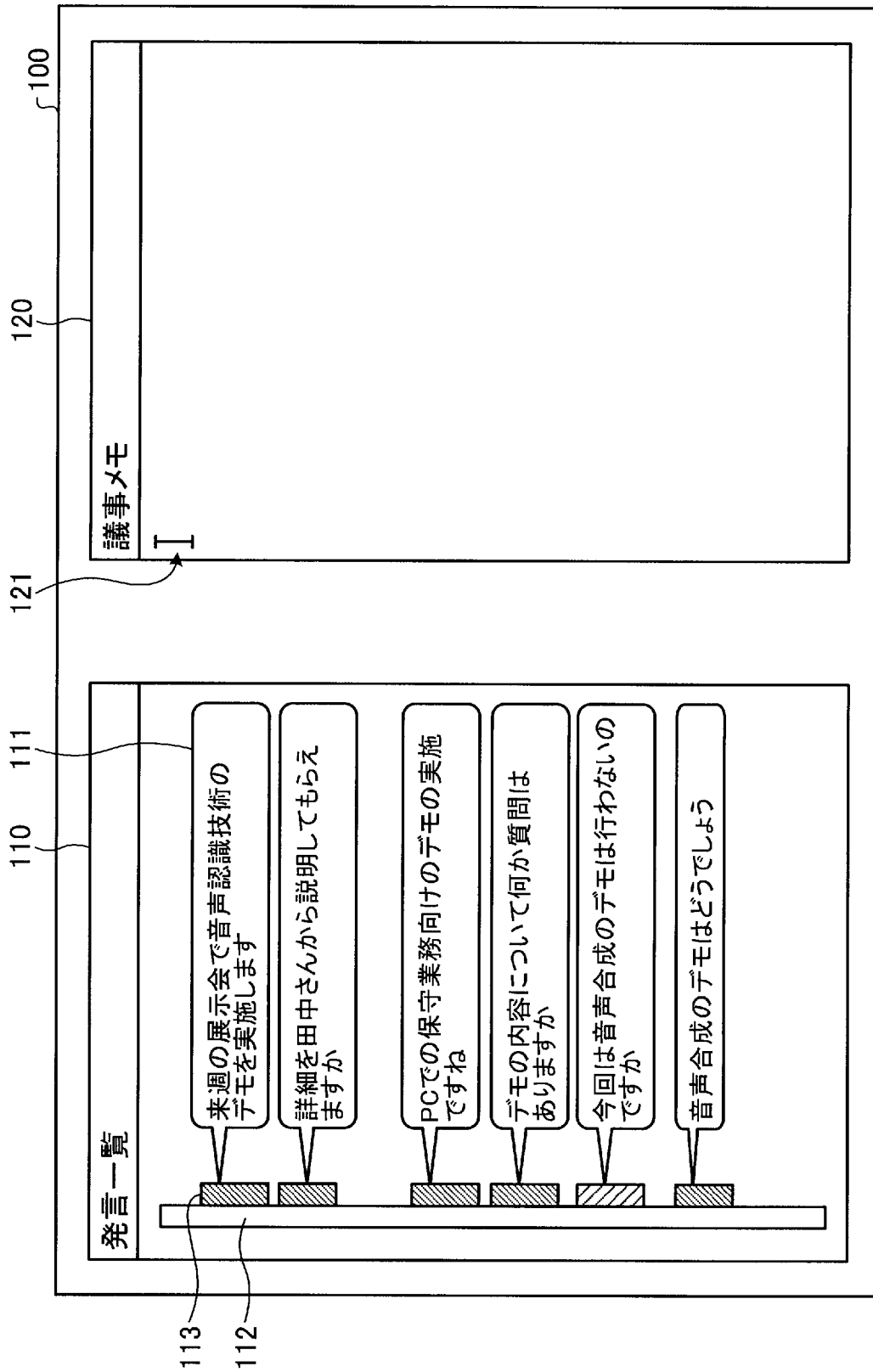
発信ID	日時	発言ユーザ	音声ファイル名	収録マイク種別	会議ID
1_s	10:05:10	1_u	1.wav	個別マイク	1_c
2_s	10:05:15	1_u	2.wav	個別マイク	1_c
3_s	10:05:40	1_u	3.wav	個別マイク	1_c
4_s	10:05:45	1_u	4.wav	個別マイク	1_c
5_s	10:05:55	3_u	5.wav	個別マイク	1_c
6_s	10:06:00	-1_u	6.wav	集音マイク	1_c
7_s	10:06:05	1_u	7.wav	個別マイク	1_c
8_s	10:06:25	3_u	8.wav	個別マイク	1_c
9_s	10:06:35	1_u	9.wav	個別マイク	1_c
10_s	10:06:45	3_u	10.wav	個別マイク	1_c
11_s	10:06:55	1_u	11.wav	個別マイク	1_c
12_s	10:07:00	-1_u	12.wav	集音マイク	1_c
13_s	10:07:05	1_u	13.wav	個別マイク	1_c
14_s	10:07:10	1_u	14.wav	個別マイク	1_c
15_s	10:07:15	1_u	15.wav	個別マイク	1_c

[図8]

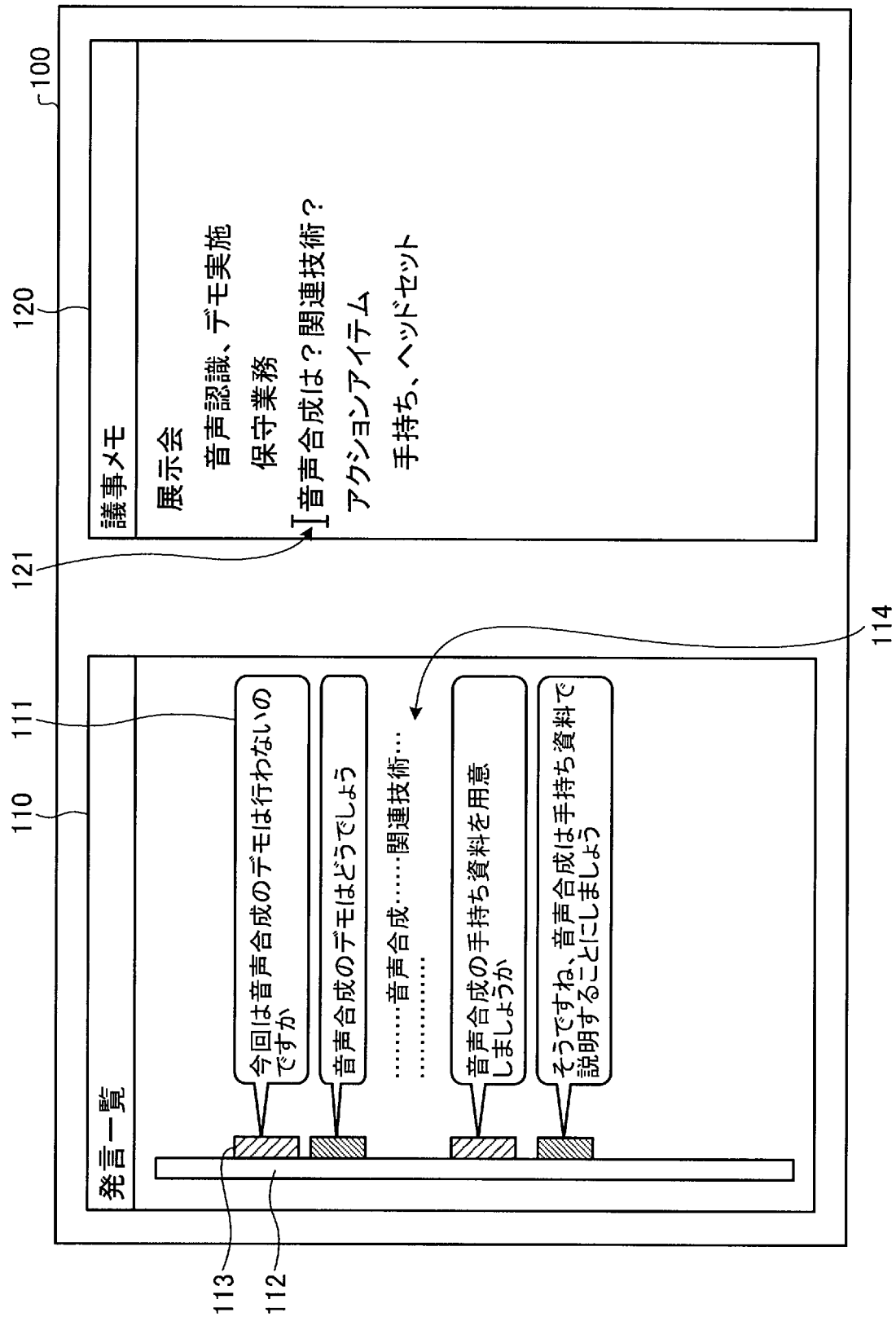
14

発言ID	認識結果	認識キーワード	想定認識精度
1_s	来週の展示会で	来週、展示会、	80%
2_s	音声認識技術の	音声認識、技術、	80%
3_s	デモを実施します	でも、デモ、実施	80%
4_s	詳細を田中さんから	詳細、田中	80%
5_s	説明してもらえますか	説明	80%
6_s	PCでの保守業務向けの	PC、残し、保守業務	80%
7_s	デモの実施ですね	デモ	80%
8_s	デモの内容について	デモ、内容	80%
9_s	何か質問はありますか	質問	80%
10_s	今回は音声合成の	音声合成	80%
11_s	デモは行わないのですか	デモ	30%
12_s		音声認識、事実 田中、説明、エルゴ 保守業務、岐路、サラダ 音声認識、デモ、負い目 残し、保守業務、デモ でも、デモ、質問 音声合成、デモ	
13_s	音声合成のデモはどうでしょう	音声合成、デモ	80%
14_s	音声合成の手持ち資料を	音声合成、手持ち資料	80%
15_s	用意しましょうか	用意	80%
16_s	そうですね、音声合成は	音声合成	80%
17_s	手持ち資料で説明する	手持ち資料、説明	80%
18_s	ことにしましょう		
19_s	音声の入力は	音声、入力	80%
20_s	ヘッドセットからですか	ヘッドセット	80%
21_s	音声入力用のヘッドセットの	音声、入力、ヘッドセット	80%
22_s	準備も必要ですね	準備、必要	30%
23_s		音声合成、デモ、銅版 岩手、音声合成、関連技術、 音声合成、手持ち資料、余市 音声合成、手持ち資料、するめ 音声、入力、ヘッドセット、殻 確実、音声、入力 音声、入力、ヘッドセット、準備	
24_s	ではアクションアイテムを	アクションアイテム	80%
25_s	をまとめましょう	まとめ	80%
26_s	山本さんは音声合成の	山本、音声合成	80%
27_s	手持ち資料の印刷を	手持ち資料、印刷	80%
28_s			
29_s	田中さんはヘッドセットの	田中、ヘッドセット	80%
30_s	準備をお願いします	準備	

[図9]



[図10]

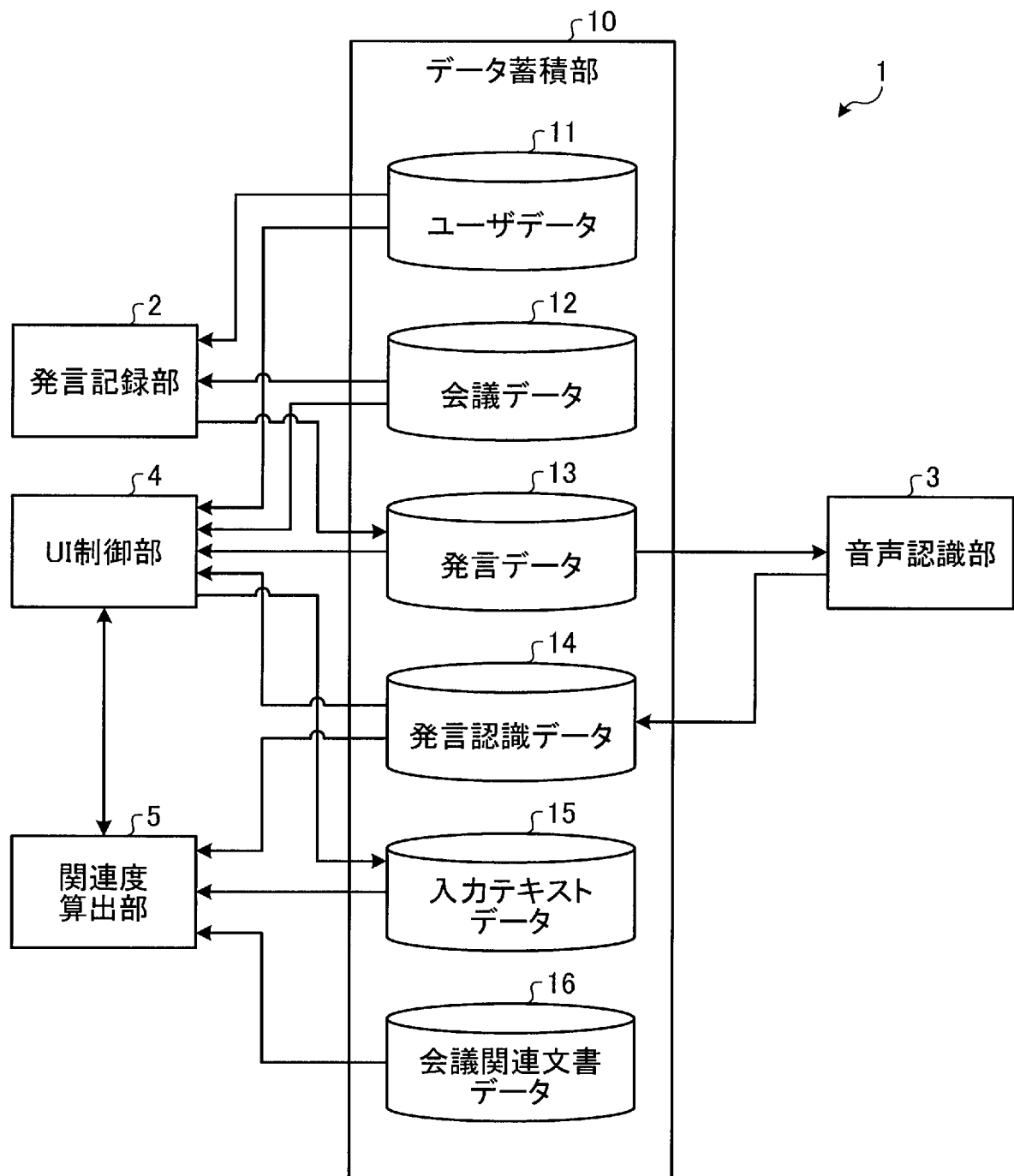


[図11]

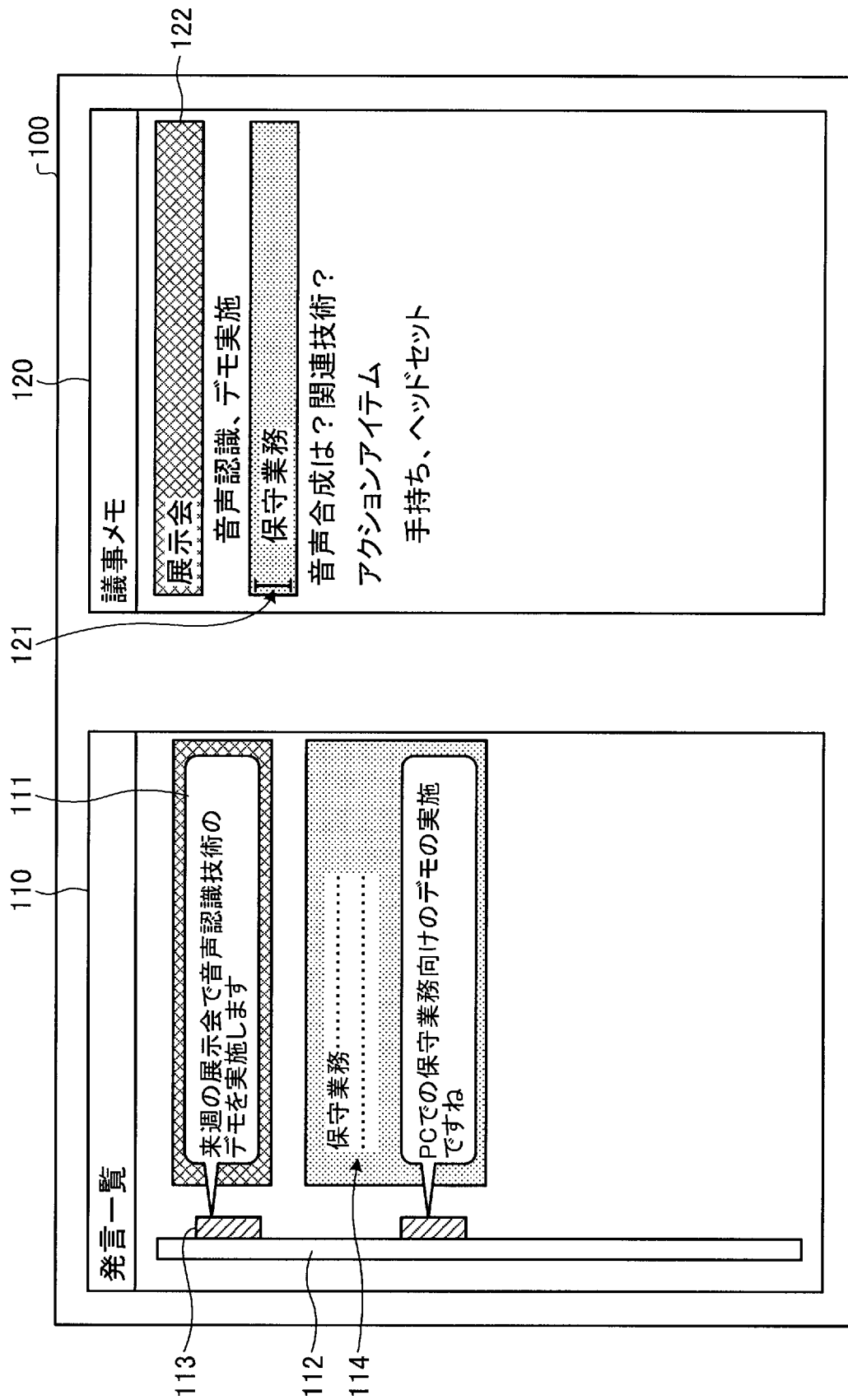
15
↙

メモID	行	テキスト	会議ID
1_m	1	展示会	1_c
2_m	2	音声認識、デモ実施	1_c
3_m	3	保守業務	1_c
4_m	4	音声合成は？関連技術？	1_c
5_m	5	アクションアイテム	1_c
6_m	6	手持ち、ヘッドセット	1_c
⋮	⋮	⋮	⋮

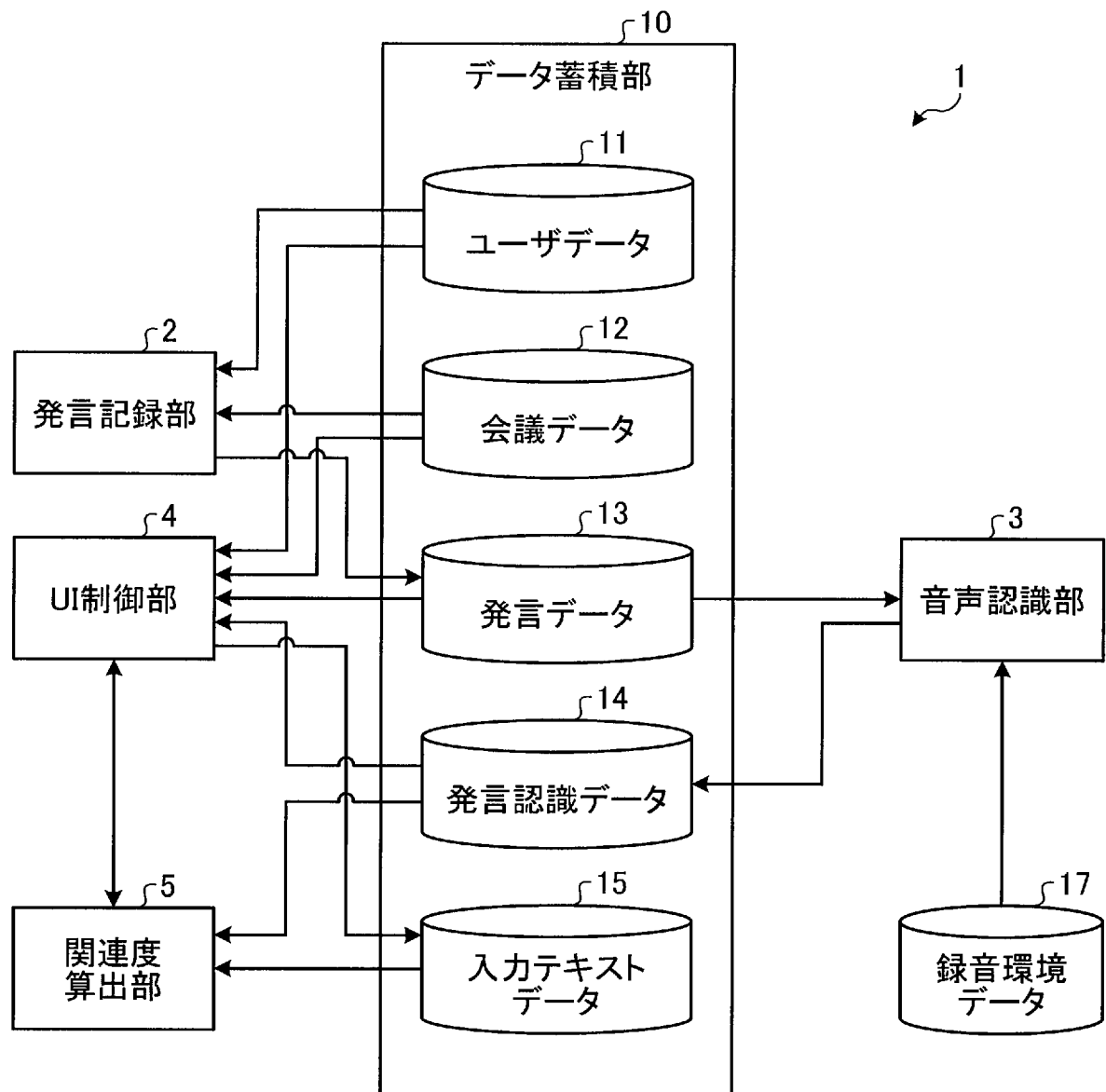
[図12]



[図13]



[図14]

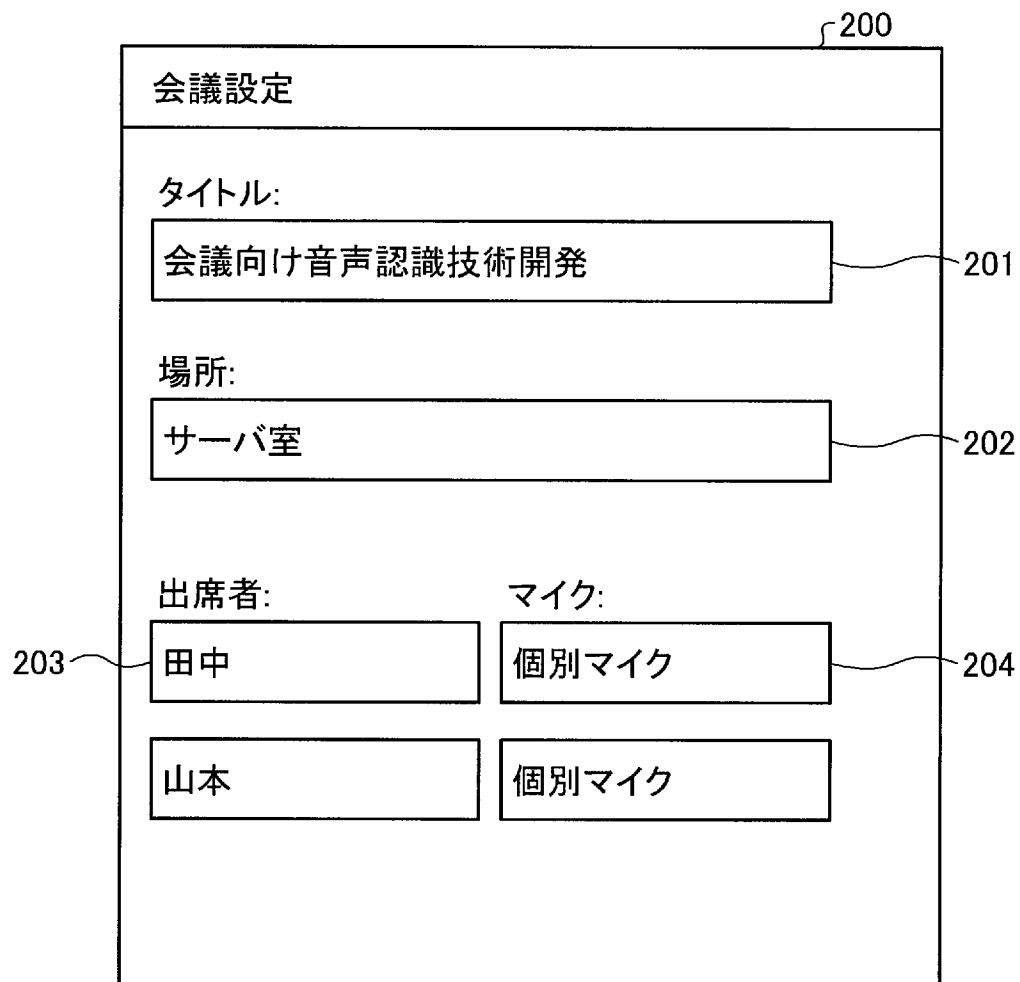


[図15]

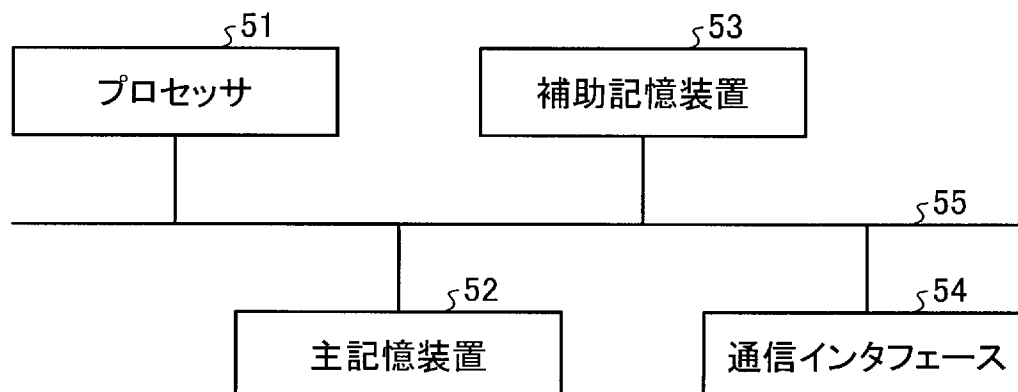
17
↘

データID	収録マイク種別	発言ユーザ	場所	後処理	想定認識精度
1_d	個別マイク	*	*	無し	80%
2_d	集音マイク	*	*	無し	30%
3_d	個別マイク	2_u	*	無し	90%
4_d	個別マイク	*	サーバ室	無し	60%
5_d	集音マイク	*	*	話者照合	50%

[図16]



[図17]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/061269

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/22(2006.01)i, G10L15/00(2013.01)i, G10L15/10(2006.01)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, G10L15/00, G10L15/10

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	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

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	JP 2010-238050 A (NEC Corp.), 21 October 2010 (21.10.2010), paragraphs [0010], [0076] to [0087]; fig. 10 to 12 (Family: none)	1-3, 6-8, 10-12, 14, 16-17 4-5, 9, 13, 15
A		
X	JP 2014-174465 A (Ricoh Co., Ltd.), 22 September 2014 (22.09.2014), paragraphs [0049] to [0065], [0130] to [0135], [0149], [0157]; fig. 12 (Family: none)	1, 16-17

☐ 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
20 May 2015 (20.05.15)

Date of mailing of the international search report
02 June 2015 (02.06.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G10L15/22(2006.01)i, G10L15/00(2013.01)i, G10L15/10(2006.01)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（I P C））

Int.Cl. G10L15/22, G10L15/00, G10L15/10

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	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 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 A	JP 2010-238050 A（日本電気株式会社）2010.10.21, [0010]、 [0076] - [0087]、第10 - 12図（ファミリーなし）	1-3, 6-8, 10-1 2, 14, 16-17 4-5, 9, 13, 15
X	JP 2014-174465 A（株式会社リコー）2014.09.22, [0049] - [0065]、[0130] - [0135]、[0149]、[0157]、 第12図（ファミリーなし）	1, 16-17

☐ 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」国際出願日前で、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

2 0 . 0 5 . 2 0 1 5

国際調査報告の発送日

0 2 . 0 6 . 2 0 1 5

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
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）

大野 弘

5 Z

9 1 7 5

電話番号 03-3581-1101 内線 3591