(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
2024年12月19日(19.12.2024)

WIPO | PCT

(10) 国際公開番号
WO 2024/257308 A1

(51) 国際特許分類:
G10L 15/04 (2013.01) *G10L 17/02* (2013.01)

(21) 国際出願番号： PCT/JP2023/022279

(22) 国際出願日： 2023年6月15日(15.06.2023)

(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: 日本電気株式会社 (**NEC CORPORATION**) [JP/JP]; 〒1088001 東京都港区芝五丁目７番１号 Tokyo (JP).

(72) 発明者: カク レイ (**GUO, Ling**); 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP). 山本 仁 (**YAMAMOTO, Hitoshi**); 〒1088001 東京都港区芝五丁目７番１号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 江上 達夫, 外 (**EGAMI, TATSUO** et al.); 〒1040031 東京都中央区京橋一丁目１６番１０号 ＶＰＯ京橋３階 東京セントラル特許事務所内 Tokyo (JP).

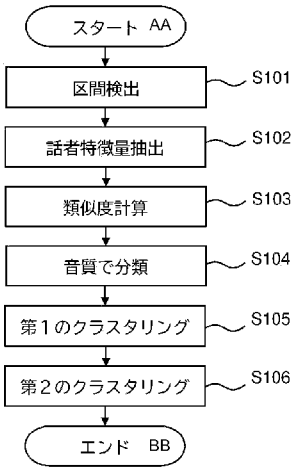
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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,

(54) **Title:** CLASSIFICATION DEVICE, CLASSIFICATION METHOD, RECORDING MEDIUM, AND INFORMATION DISPLAY DEVICE

(54) 発明の名称： 分類装置、分類方法、記録媒体及び情報表示装置

[図3]



S101 Section detection
S102 Speaker feature amount extraction
S103 Similarity calculation
S104 Classification by sound quality
S105 First clustering
S106 Second clustering
AA Start
BB End

(57) **Abstract:** This classification device comprises: a first classification means that classifies a plurality of section voices generated by dividing an input voice into a plurality of sections into a plurality of first section voices having a first sound quality and a plurality of second section voices having a second sound quality different from the first sound quality; and a second classification means that classifies the plurality of second section voices as any of a plurality of groups based on the plurality of first section voices or a new group different from the plurality of groups.

(57) 要約：分類装置は、入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、複数の第１区間音声に基づく複数の群のいずれか又は該複数の群とは異なる新たな群に、複数の第２区間音声を分類する第２分類手段と、を備える。

WO 2024/257308 A1

［続葉有］

IT, LT, LU, LV, MC, ME, 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：

一　国際調査報告（条約第21条(3)）

明　細　書

発明の名称 ： 分類装置、分類方法、記録媒体及び情報表示装置

技術分野

[0001]　この開示は、分類装置、分類方法、記録媒体及び情報表示装置の技術分野に関する。

背景技術

[0002]　この種の装置として、例えば、通話を録音することで生成された音声データ内から、音声認識処理を要することなく、所望のキーワードに対応する音声データの箇所を特定する装置が提案されている（特許文献１参照）。その他、この開示に関連する技術として、特許文献２乃至４が挙げられる。

先行技術文献

特許文献

[0003] 特許文献1：特開２０１１－０５３５６３号公報

特許文献2：特開２０２２－１４４９２７号公報

特許文献3：特開２０１５－０４０９３１号公報

特許文献4：特開２０１３－２２８４７２号公報

発明の概要

発明が解決しようとする課題

[0004]　この開示は、先行技術文献に記載された技術の改良を目的とする音声処理装置、音声処理方法、記録媒体及び情報表示装置を提供することを課題とする。

課題を解決するための手段

[0005]　分類装置の一態様は、入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する第２分

類手段と、を備える。

[0006] 分類方法の一態様は、コンピュータが、入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する。

[0007] 記録媒体の一態様は、コンピュータに、入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する分類方法を実行させるコンピュータプログラムが記録されている。

[0008] 情報表示装置の一態様は、第１入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する第２分類手段と、前記複数の群及び前記新たな群に基づいて、前記複数の区間音声各々の話者を示す第１話者情報を生成する第１生成手段と、前記複数の区間音声の少なくとも一つの区間音声に基づいて、前記少なくとも一つの区間音声に相当する文字列を示す文字情報を生成する第２生成手段と、前記生成された第１話者情報により示される話者と、前記生成された文字情報により示される文字列とを表示する表示手段と、を備える。

図面の簡単な説明

[0009] [図1]分類装置の構成の一例を示すブロック図である。

[図2]分類装置の構成の他の例を示すブロック図である。

[図3]本開示にかかる分類装置の動作を示すフローチャートである。

[図4]音声検出処理の概念を示す図である。

[図5]類似度の一例を示す図である。

[図6]表示画面の一例を示す図である。

[図7]話者入力画面の一例を示す図である。

[図8]表示画面の他の例を示す図である。

発明を実施するための形態

[0010] ＜第１実施形態＞

分類装置、分類方法及び記録媒体に係る実施形態について図１を参照して説明する。以下では、分類装置１０を用いて、分類装置、分類方法及び記録媒体に係る実施形態を説明する。

[0011] 図１において、分類装置１０は、分類部１１及びクラスタリング部１２を備える。分類部１１は、複数の区間音声を、第１音質を有する複数の第１区間音声と、第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する。尚、第１音質は、第２音質より高音質であってよい。

[0012] 複数の区間音声は、入力音声が複数の区間に分割されることによって生成されてよい。入力音声を複数の区間音声に分割する方法には既存の技術を適用可能である。一例としては、入力音声を、所定の時間単位（例えば、２秒単位）で、複数の区間音声に分割する方法が挙げられる。他の例としては、入力音声を、オーバーラップを持たせながら所定の時間単位で、複数の区間音声に分割する方法が挙げられる。尚、入力音声は、例えばマイクロフォン等の集音装置により取得されてよい。

[0013] クラスタリング部１２は、複数の第１区間音声に基づく複数の群のいずれか又は該複数の群とは異なる新たな群に、複数の第２区間音声各々を分類する。ここで、「分類する」とは、所定の基準に従って（例えば、事前に決定されている正解に従って）分けることに限らず、データ間（例えば、一の区間音声と他の区間音声との間）の類似度に基づいて分けることも含む概念である。つまり、本実施形態に係る「分類」には、クラスタリングが含まれていてよい。

[0014] 複数の第１区間音声に基づく複数の群は、例えば以下の方法により生成（

又は、設定）されてよい。分類部１１は、複数の第１区間音声各々の特徴を抽出してよい。区間音声の特徴（即ち、音声の特徴）の一例として、音響特徴であるメル周波数ケプストラム係数（ＭＦＣＣ：Ｍｅｌ－Ｆｒｅｑｕｅｎｃｙ　Ｃｅｐｓｔｒｕｍ　Ｃｏｅｆｆｉｃｉｅｎｔｓ）を用いて算出したｉ－ｖｅｃｔｏｒが挙げられる。区間音声の特徴の他の例として、ＤＮＮ（Ｄｅｅｐ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ）を用いて算出したｘ－ｖｅｃｔｏｒが挙げられる。

[0015] 分類部１１は、複数の第１区間音声の特徴に基づいて、一の第１区間音声と他の第１区間音声との間の類似度を算出してよい。２つの音声間の類似度の一例として、ＰＬＤＡ（Ｐｒｏｂａｂｉｌｉｓｔｉｃ　ｌｉｎｅａｒ　ｄｉｓｃｒｉｍｉｎａｎｔ　ａｎａｌｙｓｉｓ）により計算されたｉ－ｖｅｃｔｏｒ又はｘ－ｖｅｃｔｏｒの類似度が挙げられる。尚、類似度は、類似の程度を示す指標値と称されてもよい。尚、区間音声の特徴の抽出及び類似度の計算は、分類部１１に代えてクラスタリング部１２により行われてよい。

[0016] クラスタリング部１２は、複数の第１区間音声に係る類似度に基づいて、複数の第１区間音声に対してクラスタリング処理を行ってよい。クラスタリング処理の一例として、階層型凝集的クラスタリング（Ａｇｇｌｏｍｅｒａｔｉｖｅ　Ｈｉｅｒａｒｃｈｉｃａｌ　Ｃｌｕｓｔｅｒｉｎｇ：ＡＨＣ）を用いた処理が挙げられる。クラスタリング処理により複数の第１区間音声がグループ分けされることで、上記複数の群（即ち、複数の第１区間音声に基づく複数の群）が生成（又は、設定）されてよい。

[0017] クラスタリング部１２は、例えば以下の方法により、複数の第１区間音声に基づく複数の群のいずれか又は該複数の群とは異なる新たな群に、複数の第２区間音声各々を分類してよい。クラスタリング部１２は、複数の第２区間音声各々の特徴を抽出してよい。複数の第２区間音声各々の特徴は、第１区間音声の特徴と同様の方法により抽出されてよい。

[0018] クラスタリング部１２は、一の第２区間音声の特徴と、複数の第１区間音声に基づく複数の群のうち一の群に含まれる一又は複数の第１区間音声の特

徴とに基づいて、一の第２区間音声と、一又は複数の第１区間音声との間の類似度を計算してよい。クラスタリング部１２は更に、一の第２区間音声の特徴と、複数の第１区間音声に基づく複数の群のうち他の群に含まれる一又は複数の第１区間音声の特徴とに基づいて、一の第２区間音声と、一又は複数の第１区間音声との間の類似度を計算してよい。

[0019] クラスタリング部１２は、上記のように計算された類似度のうち、最大の類似度を特定してよい。クラスタリング部１２は、該特定された最大の類似度となる、一の第２区間音声と一の第１区間音声との組み合わせを特定することにより、該一の第１区間音声が含まれる群（即ち、複数の第１区間音声に基づく群）を特定してよい。クラスタリング部１２は、上記特定された最大の類似度が、所定の閾値以上である場合、上記特定された群に一の第２区間音声を分類してよい。他方で、クラスタリング部１２は、上記特定された最大の類似度が、所定の閾値未満である場合、複数の第１区間音声に基づく複数の群とは異なる新たな群に、一の第２区間音声を分類してよい。

[0020] このように、分類装置１０では、入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、複数の第１区間音声に基づく複数の群のいずれか又は該複数の群とは異なる新たな群に、複数の第２区間音声各々を分類する分類方法が行われる。

[0021] 分類装置１０は、コンピュータが記録媒体に記録されたコンピュータプログラムを読み込むことによって実現されてよい。この場合、記録媒体には、コンピュータに、入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、複数の第１区間音声に基づく複数の群のいずれか又は該複数の群とは異なる新たな群に、複数の第２区間音声各々を分類する分類方法を実行させるコンピュータプログラムが記録されている、と言える。

[0022] （技術的効果）

本願発明者の研究によれば次の事が判明している。一の話者に係る第１音声の第１特徴と、一の話者に係る第２音声の第２特徴とは、話者が同一であるがゆえに互いに似ている。一の話者に係る第１音声の第１特徴と、他の話者に係る第３音声の第３特徴とは、話者が異なるがゆえに似ていない。このため、第１特徴及び第２特徴に基づく第１音声と第２音声との間の類似度は、第１特徴及び第３特徴に基づく第１音声と第３音声との間の類似度に比べて高くなる。言い換えれば、第１特徴及び第３特徴に基づく第１音声と第３音声との間の類似度は、第１特徴及び第２特徴に基づく第１音声と第２音声との間の類似度に比べて低くなる。音声が高音質である場合、２つの音声の話者が同一である場合の類似度と、２つの音声の話者が互いに異なる場合の類似度とは明確に異なる。

[0023] 例えば、第１音声の音質が高音質であり、第２音声の音質が低音質である場合、第１音声と第２音声との間の類似度は、第１音声及び第２音声の音質が高音質である場合に比べて低くなる。他方で、第１音声の音質が高音質であり、第３音声の音質が低音質である場合、第１音声と第３音声との間の類似度は、第１音声及び第３音声が高音質である場合に比べて高くなる。この結果、第２音声の第２特徴が、第１音声の第１特徴に似ているとも言えるし、似ていないとも言えない状態になり得る。同様に、第３音声の第３特徴が、第１音声の第１特徴に似ているとも言えるし、似ていないとも言える状況になり得る。このため、音質が異なる複数の音声に対してクラスタリング処理を行うと、複数の音声を話者毎に適切に分けられない可能性がある。

[0024] これに対して、分類装置１０では、分類部１１が、複数の区間音声を、第１音質を有する複数の第１区間音声と、第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する。そして、クラスタリング部１２が、複数の第１区間音声に基づく複数の群のいずれか又は該複数の群とは異なる新たな群に、複数の第２区間音声各々を分類する。つまり、分類装置１０では、音質によって複数の区間音声各々が分類された後に、音質毎に複数の区間音声各々がグループ分けされる。このため、分類装置１０によれば、複数の

音声を話者毎に適切に分けることができる。

[0025] ＜第２実施形態＞

分類装置、分類方法、記録媒体及び情報表示装置に係る実施形態について図２乃至図７を参照して説明する。以下では、分類装置２０を用いて、分類装置、分類方法、記録媒体及び情報表示装置に係る実施形態を説明する。

[0026] 図２において、分類装置２０は、演算装置２１、記憶装置２２及び通信装置２３を備える。分類装置２０は更に、入力装置２４及び出力装置２５を備えていてよい。尚、分類装置２０は、入力装置２４及び出力装置２５の少なくとも一方を備えていなくてもよい。分類装置２０において、演算装置２１、記憶装置２２、通信装置２３、入力装置２４及び出力装置２５は、データバス２６を介して接続されていてよい。

[0027] 演算装置２１は、例えば、ＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）、ＧＰＵ（Ｇｒａｐｈｉｃｓ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）、ＦＰＧＡ（Ｆｉｅｌｄ Ｐｒｏｇｒａｍｍａｂｌｅ Ｇａｔｅ Ａｒｒａｙ）、ＴＰＵ（ＴｅｎｓｏｒＰｒｏｃｅｓｓｉｎｇＵｎｉｔ）、及び、量子プロセッサのうち少なくとも一つを含んでよい。

[0028] 記憶装置２２は、例えば、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）、ＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）、ハードディスク装置、光磁気ディスク装置、ＳＳＤ（Ｓｏｌｉｄ Ｓｔａｔｅ Ｄｒｉｖｅ）、及び、光ディスクアレイのうち少なくとも一つを含んでよい。つまり、記憶装置２２は、一時的でない記録媒体を含んでよい。記憶装置２２は、所望のデータを記憶可能である。例えば、記憶装置２２は、演算装置２１が実行するコンピュータプログラムを一時的に記憶していてよい。記憶装置２２は、演算装置２１がコンピュータプログラムを実行している場合に演算装置２１が一時的に使用するデータを一時的に記憶してよい。

[0029] 通信装置２３は、分類装置２０の外部の装置と通信可能であってもよい。尚、通信装置２３は、有線通信を行ってもよいし、無線通信を行ってもよい。

[0030] 入力装置２４は、外部から分類装置２０に対する情報の入力を受け付け可能な装置である。入力装置２４は、分類装置２０のオペレータが操作可能な操作装置（例えば、キーボード、マウス、タッチパネル等）を含んでよい。入力装置２４は、例えばＵＳＢ（Ｕｎｉｖｅｒｓａｌ Ｓｅｒｉａｌ Ｂｕｓ）メモリ等の、分類装置２０に着脱可能な記録媒体に記録されている情報を読み取り可能な記録媒体読取装置を含んでよい。尚、分類装置２０に、通信装置２３を介して情報が入力される場合（言い換えれば、分類装置２０が通信装置２３を介して情報を取得する場合）、通信装置２３は入力装置として機能してよい。

[0031] 出力装置２５は、分類装置２０の外部に対して情報を出力可能な装置である。出力装置２５は、上記情報として、文字や画像等の視覚情報を出力してもよいし、音声等の聴覚情報を出力してもよいし、振動等の触覚情報を出力してもよい。出力装置２５は、例えば、ディスプレイ、スピーカ、プリンタ及び振動モータの少なくとも一つを含んでいてよい。出力装置２５は、例えばＵＳＢメモリ等の、分類装置２０に着脱可能な記録媒体に情報を出力可能であってもよい。尚、分類装置２０が通信装置２３を介して情報を出力する場合、通信装置２３は出力装置として機能してよい。

[0032] 演算装置２１は、論理的に実現される機能ブロックとして、又は、物理的に実現される処理回路として、分類部２１１、クラスタリング部２１２、話者情報生成部２１３、文字情報生成部２１４及び表示制御部２１５を有していてよい。

[0033] 尚、分類部２１１、クラスタリング部２１２、話者情報生成部２１３、文字情報生成部２１４及び表示制御部２１５の少なくとも一つは、論理的な機能ブロックと、物理的な処理回路（即ち、ハードウェア）とが混在する形式で実現されてよい。分類部２１１、クラスタリング部２１２、話者情報生成部２１３、文字情報生成部２１４及び表示制御部２１５の少なくとも一部が機能ブロックである場合、分類部２１１、クラスタリング部２１２、話者情報生成部２１３、文字情報生成部２１４及び表示制御部２１５の少なくとも

一部は、演算装置２１が所定のコンピュータプログラムを実行することにより実現されてよい。

[0034] 演算装置２１は、上記所定のコンピュータプログラムを、記憶装置２２から取得してよい（言い換えれば、読み込んでよい）。演算装置２１は、コンピュータで読み取り可能であって且つ一時的でない記録媒体が記憶している上記所定のコンピュータプログラムを、分類装置２０が備える図示しない記録媒体読み取り装置を用いて読み込んでもよい。演算装置２１は、通信装置２３を介して、分類装置２０の外部の図示しない装置から上記所定のコンピュータプログラムを取得してもよい（言い換えれば、ダウンロードしてもよい又は読み込んでもよい）。尚、演算装置２１が実行する上記所定のコンピュータプログラムを記録する記録媒体としては、光ディスク、磁気媒体、光磁気ディスク、半導体メモリ、及び、その他プログラムを格納可能な任意の媒体の少なくとも一つが用いられてよい。

[0035] 尚、「分類部２１１」及び「クラスタリング２１２」は、夫々、上述した第１実施形態における「分類部１１」及び「クラスタリング部１２」に対応する構成要素である。

[0036] 分類部２１１及びクラスタリング部２１２の動作について図３乃至図５を参照して説明する。尚、分類部２１１及びクラスタリング部２１２について、上述した第１実施形態の説明と重複する説明を適宜省略する。

[0037] 図３において、分類部２１１は、入力音声から音声が存在する区間（言い換えれば、音声に対応する音響信号を含む区間）を検出する（ステップＳ１０１）。音声が存在する区間を、以降、適宜「音声区間」と称する。ステップＳ１０１の処理において、分類部２１１は、入力音声に対して音声検出処理を行うことによって、音声区間を検出してよい。尚、音声検出処理には、既存の各種態様を適用可能であるので、その詳細についての説明は省略する。尚、人の発話に係る音声を検出する場合、音声検出処理は、発話検出処理と称されてもよい。

[0038] 入力音声には、音声と、音声以外の音（例えば、物音、音楽及び動物の鳴

声の少なくとも一つ）とが含まれていてよい。また、入力音声には、無音区間が含まれていてよい。ステップＳ１０１の処理の結果、例えば図４に示すように、入力音声から音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５が検出されてよい。尚、図４における区間ＢＳ１、ＢＳ２、ＢＳ３及びＢＳ４は、音声区間以外の区間の一例である。尚、音声区間以外の区間には、例えば、音声以外の音が存在する区間及び無音区間が含まれていてよい。

[0039] 分類部２１１は、一の入力音声から、検出された音声区間（例えば、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５）を抽出することによって、一の入力音声を複数の区間音声に分割してよい。この場合、一の音声区間が一の区間音声に対応してよい。或いは、分類部２１１は、一の入力音声を分割して複数の区間音声を生成した後に、入力音声としての複数の区間音声各々に対して音声検出処理を行ってもよい。この場合、分類部２１１は、複数の区間音声各々から音声区間を抽出してもよい。つまり、分類部２１１は、一の区間音声から音声区間を抽出することによって、一の区間音声を更に分割してもよい。

[0040] ステップＳ１０１の処理の後、分類部２１１は、検出された音声区間（例えば、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５）の話者特徴量を抽出する（ステップＳ１０２）。尚、話者特徴量は、ｉ－ｖｅｃｔｏｒ及びｘ－ｖｅｃｔｏｒの少なくとも一方であってよい。尚、分類部２１１は、音声区間を入力すると、入力された音声区間の話者特徴量を出力する学習モデルを用いて、話者特徴量を抽出してもよい。このような学習モデルは、音声区間に相当する音声データを用いた機械学習により構築されてよい。

[0041] 次に、分類部２１１は、ステップＳ１０２の処理において抽出された話者特徴量に基づいて、複数の音声区間の夫々の間の類似度を計算する（ステップＳ１０３）。尚、音声区間は、一の入力音声の一の区間であると言える。このため、音声区間は区間音声と称されてもよい。このため、複数の音声区間の夫々の間の類似度を計算することは、複数の区間音声の夫々の間の類似度を計算することであると言える。尚、類似度は、ＰＬＤＡにより計算され

たｉ－ｖｅｃｔｏｒ又はｘ－ｖｅｃｔｏｒの類似度であってよい。尚、分類部２１１は、２つの話者特徴量を入力すると、入力された２つの話者特徴量に夫々対応する２つの音声区間の類似度を出力する学習モデルを用いて、類似度を計算してもよい。このような学習モデルは、話者特徴量に相当する学習データを用いた機械学習により構築されてよい。

[0042] 例えば、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５の話者特徴量を、夫々、話者特徴量ＳＦ１、ＳＦ２、ＳＦ３、ＳＦ４及びＳＦ５とする。音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５の夫々の間の類似度の計算結果は、図５に示す表として表すことができる。図５において、例えば、縦軸上の話者特徴量ＳＦ１と、横軸上の話者特徴量ＳＦ２とが交差する位置の値（具体的には、“０．１”）は、音声区間ＶＳ１と音声区間ＶＳ２との間の類似度を示している。本実施形態では、一の音声区間の話者特徴量と他の音声区間の話者特徴量とが似ているほど、類似度の値が大きくなる。

[0043] 尚、分類部２１１は、類似度に代えて、距離を計算してもよい。つまり、分類部２１１は、一の音声区間の話者特徴量と他の音声区間の話者特徴量とが似ている度合いを示す指標として、類似度に代えて距離を計算してもよい。この場合、一の音声区間の話者特徴量と他の音声区間の話者特徴量とが似ているほど、距離を示す値が小さくなってよい。

[0044] ステップＳ１０３の処理の後、分類部２１１は、複数の音声区間を、第１音質を有する音声区間と、第１音質とは異なる第２音質を有する音声区間とに分類する（ステップＳ１０４）。ステップＳ１０４の処理において、分類部２１１は、複数の音声区間各々の音質と基準音質とに基づいて、複数の音声区間を、第１音質を有する音声区間と、第２音質を有する音声区間とに分類してよい。例えば、分類部２１１は、音声区間の音質が基準音質より高い場合、該音声区間を第１音質を有する音声区間に分類してよい。他方で、分類部２１１は、音声区間の音質が基準音質より低い場合、該音声区間を第２音質を有する音声区間に分類してよい。尚、音声区間の音質と基準音質とが「等しい」場合は、どちらかの場合に含めて扱えばよい。このように構成す

れば、複数の音声区間を、第１音質を有する音声区間と第２音質を有する音声区間とに容易に分類することができる。尚、基準音質は、例えば話者特徴量に係る学習モデルを構築するために用いられた音声データの音質の平均値であってよい。尚、第１音質は、第２音質より高音質であってよい。

[0045] 例えば、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５のうち、音声区間ＶＳ１、ＶＳ２及びＶＳ３が高音質な音声区間（言い換えれば、第１音質を有する音声区間）であり、音声区間ＶＳ４及びＶＳ５が低音質な音声区間（言い換えれば、第２音質を有する音声区間）であるものとする。また、音声区間ＶＳ１及びＶＳ３に係る話者が第１話者であり、音声区間ＶＳ２、ＶＳ４及びＶＳ５に係る話者が第２話者であるものとする。この場合、ステップＳ１０４の処理において、分類部２１１は、音声区間ＶＳ１、ＶＳ２及びＶＳ３を第１音質を有する音声区間に分類するとともに、音声区間ＶＳ４及びＶＳ５を第２音質を有する音声区間に分類してよい。

[0046] クラスタリング部２１２は、ステップＳ１０４の処理において第１音質を有する音声区間に分類された複数の音声区間を用いて第１のクラスタリング処理を行う（ステップＳ１０５）。第１のクラスタリング処理の一例として、階層型凝集的クラスタリング（ＡＨＣ）を用いた処理が挙げられる。

[0047] 第１のクラスタリング処理について、図５に示す類似度を用いて具体的に説明する。クラスタリング部２１２は、音声区間ＶＳ１、ＶＳ２及びＶＳ３の夫々の間の類似度のうち、類似度が最大となる音声区間の組合せを特定してよい。図５に示す例では、音声区間ＶＳ１の話者特徴量ＳＦ１と音声区間ＶＳ３の話者特徴量ＳＦ３とが交差する位置の値（即ち、“０．９”）が最大である。このため、クラスタリング部２１２は、類似度が最大となる音声区間の組合せとして、音声区間ＶＳ１と音声区間ＶＳ３との組合せを特定してよい。

[0048] 次に、クラスタリング部２１２は、音声区間ＶＳ１と音声区間ＶＳ３との間の類似度が、第１閾値以上であるか否かを判定してよい。ここで、第１閾値は「０．５」であるものとする。音声区間ＶＳ１と音声区間ＶＳ３との間

の類似度は、「０．９」であり、第１閾値より大きい。このため、クラスタリング部２１２は、音声区間ＶＳ１と音声区間ＶＳ３とを同じグループ（“クラスタ”と称されてもよい）に分けてよい。ここで、音声区間ＶＳ１及びＶＳ３を含むグループを、グループＣ１とする。

[0049] 次に、クラスタリング部２１２は、音声区間ＶＳ１及びＶＳ３と、音声区間ＶＳ２との間の類似度を計算してよい。この場合、クラスタリング部２１２は、音声区間ＶＳ１と音声区間ＶＳ２との間の類似度と、音声区間ＶＳ２と音声区間ＶＳ３との間の類似度との合計値を２で割った値を、音声区間ＶＳ１及びＶＳ３と、音声区間ＶＳ２との間の類似度としてよい。

[0050] 図５において、音声区間ＶＳ１の話者特徴量ＳＦ１と音声区間ＶＳ２の話者特徴量ＳＦ２とが交差する位置の値は「０．１」である。音声区間ＶＳ２の話者特徴量ＳＦ２と音声区間ＶＳ３の話者特徴量ＳＦ３とが交差する位置の値は「０．１」である。このため、音声区間ＶＳ１及びＶＳ３と、音声区間ＶＳ２との間の類似度は、“（０．１＋０．１）／２＝０．１”となる。

[0051] 音声区間ＶＳ１及びＶＳ３と、音声区間ＶＳ２との間の類似度（即ち、“０．１”）は、第１閾値（即ち、“０．５”）より小さい。このため、クラスタリング部２１２は、音声区間ＶＳ２を、音声区間ＶＳ１及びＶＳ３とは異なるグループに分けてよい。ここで、音声区間ＶＳ２を含むグループを、グループＣ２とする。

[0052] 第１音質を有する音声区間としての音声区間ＶＳ１、ＶＳ２及びＶＳ３を用いて第１のクラスタリング処理が行われることによって、グループＣ１及びＣ２が生成（又は、設定）されてよい。このため、グループＣ１及びＣ２は、第１音質を有する音声区間に基づく複数の群である、と言える。

[0053] ステップＳ１０５の処理の後、クラスタリング部２１２は、第２のクラスタリング処理を行う（ステップＳ１０６）。ステップＳ１０６の処理において、クラスタリング部２１２は、第２のクラスタリング処理を行うことによって、第２音質を有する音声区間を、第１音質を有する音声区間に基づく複数の群（例えば、グループＣ１及びＣ２）のいずれか、又は、該複数の群と

は異なる新たな群に分類する。

[0054] 第２のクラスタリング処理について、図５に示す類似度を用いて具体的に説明する。上述したように、音声区間ＶＳ４及びＶＳ５が第２音質を有する音声区間の一例に相当する。従って、クラスタリング部２１２は、第２のクラスタリング処理を行うことによって、音声区間ＶＳ４及びＶＳ５各々を、上述したグループＣ１及びＣ２のいずれか、又は、グループＣ１及びＣ２とは異なる新たなグループに分類する。

[0055] 第２のクラスタリング処理において、クラスタリング部２１２は、音声区間ＶＳ４とグループＣ１及びＣ２各々との間の類似度、並びに、音声区間ＶＳ５とグループＣ１及びＣ２各々との間の類似度を計算してよい。ここで、音声区間とグループとの間の類似度は、音声区間と、グループに含まれる一又は複数の音声区間との間の類似度に基づいて計算されてよい。

[0056] 例えば、音声区間ＶＳ４とグループＣ１との間の類似度は、音声区間ＶＳ４と、グループＣ１に含まれる音声区間ＶＳ１及びＶＳ３各々との間の類似度に基づいて計算されてよい。この場合、クラスタリング部２１２は、音声区間ＶＳ１と音声区間ＶＳ４との間の類似度と、音声区間ＶＳ３と音声区間ＶＳ４との間の類似度との合計値を２で割った値を、音声区間ＶＳ４とグループＣ１との間の類似度としてよい。

[0057] 図５において、音声区間ＶＳ１の話者特徴量ＳＦ１と音声区間ＶＳ４の話者特徴量ＳＦ４とが交差する位置の値は「０．５」である。音声区間ＶＳ３の話者特徴量ＳＦ３と音声区間ＶＳ４の話者特徴量ＳＦ４とが交差する位置の値は「０．４」である。このため、音声区間ＶＳ４とグループＣ１との間の類似度は、“（０．５＋０．４）／２＝０．４５”となる。同様の計算により、音声区間ＶＳ４とグループＣ２との間の類似度は「０．４」となる。音声区間ＶＳ５とグループＣ１との間の類似度は「０．４」となる。音声区間ＶＳ５とグループＣ２との間の類似度は「０．５」となる。

[0058] クラスタリング部２１２は、計算された類似度のうち類似度が最大となる組合せを特定してよい。ここでは、音声区間ＶＳ５とグループＣ２との組合

せが、類似度が最大となる組合せとして特定されてよい。次に、クラスタリング部２１２は、音声区間ＶＳ５とグループＣ２との間の類似度が第２閾値以上であるか否かを判定してよい。第２閾値は、第１閾値（即ち、第１のクラスタリング処理に係る閾値）と同じであってもよいし、異なっていてもよい。ここでは、第２閾値は「０．５」であるものとする。

[0059] 音声区間ＶＳ５とグループＣ２との間の類似度は「０．５」であり、第２閾値と等しい。このため、クラスタリング部２１２は、音声区間ＶＳ５をグループＣ２に分類してよい。仮に、音声区間ＶＳ５とグループＣ２との間の類似度が第２閾値より小さい場合、クラスタリング部２１２は、音声区間ＶＳ５を、グループＣ１及びＣ２とは異なる新たなグループＣ３に分類してよい。

[0060] 次に、クラスタリング部２１２は、音声区間ＶＳ４とグループＣ１及びＣ２各々との間の類似度を再度計算してよい。グループＣ２に音声区間ＶＳ２及びＶＳ５が含まれるので、音声区間ＶＳ４とグループＣ２との間の類似度は「０．５５」となる。クラスタリング部２１２は、音声区間ＶＳ４とグループＣ２との間の類似度が第２閾値以上であるか否かを判定してよい。音声区間ＶＳ４とグループＣ２との間の類似度は「０．５５」であり、第２閾値より大きい。このため、クラスタリング部２１２は、音声区間ＶＳ４をグループＣ２に分類してよい。

[0061] 第１のクラスタリング処理及び第２のクラスタリング処理の結果、音声区間ＶＳ１及びＶＳ３がグループＣ１に分類され、音声区間ＶＳ２、ＶＳ４及びＶＳ５がグループＣ２に分類される。上述したように、音声区間ＶＳ１及びＶＳ３に係る話者が第１話者であり、音声区間ＶＳ２、ＶＳ４及びＶＳ５に係る話者が第２話者である。このため、第１のクラスタリング処理及び第２のクラスタリング処理により、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５が話者毎に適切に分けられたことがわかる。

[0062] 話者情報生成部２１３は、クラスタリング部２１２によるクラスタリング処理（具体的には、第１のクラスタリング処理及び第２のクラスタリング処

理）の結果に基づいて、音声区間（例えば、音声区間ＶＳ１、ＶＳ２、ＶＳ３及びＶＳ４）の話者を示す話者情報を生成する。話者情報生成部２１３は、一の話者と他の話者とを区別（又は、識別）可能な態様で話者を示す話者情報を生成してよい。話者情報生成部２１３は、話者を、例えば「話者１」及び「話者２」等と表してよい。

[0063] 話者情報生成部２１３は、第１のクラスタリング処理によって一の群（例えば、グループＣ１及びＣ２の一方）に分類された一の音声区間（即ち、第１音質を有する音声区間）の話者と、第２のクラスタリング処理によって上記一の群に分類された他の音声区間（即ち、第２音質を有する音声区間）の話者との表示態様が互いに異なる話者情報を生成してもよい。この場合、話者情報は、一の音声区間の話者である一の話者を示す文字と、他の音声区間の話者として、上記一の話者を示唆する文字とを示してよい。「一の話者を示唆する文字」は、例えば、一の話者を示す文字が「話者１」である場合、「話者１かも」及び「話者１（推測）」の一方であってよい。このように構成すれば、クラスタリングの確からしさを、分類装置２０のユーザに提示することができる。尚、第２のクラスタリング処理によって、新たなグループ（例えば、グループＣ１及びＣ２とは異なるグループ）に分類された音声区間が存在する場合、話者情報生成部２１３は、新たなグループに分類された音声区間の話者を示す話者情報を生成しなくてもよい。或いは、第２のクラスタリング処理によって、新たなグループ（例えば、グループＣ１及びＣ２とは異なるグループ）に分類された音声区間が存在する場合、話者情報生成部２１３は、新たなグループに分類された音声区間の話者を、例えば「Ｕｎｋｏｗｎ」等と表してよい。尚、「Ｕｎｋｏｗｎ」は一例であり、これに限定されるものではない。

[0064] 尚、話者情報生成部２１３は、クラスタリング部２１２によるクラスタリング処理の結果に基づいて、一の話者に係る音声として分けられた一又は複数の音声区間に、話者認識処理及び声認証処理の少なくとも一方を施すことにより、一の話者に対応する具体的な人物を特定してもよい。この場合、話

者情報生成部２１３は、話者を具体的な人名で表してよい。

[0065] 図３のステップＳ１０２からＳ１０６の処理と並行して、文字情報生成部２１４は、ステップＳ１０１の処理において検出された音声区間に係る音声に基づいて、音声に相当する文字列を示す文字情報を生成する。尚、音声から文字列を示す文字情報を生成する方法（いわゆる文字起こし）には、既存の各種態様を適用可能であるので、その詳細についての説明は省略する。

[0066] 表示制御部２１５は、話者情報生成部２１３により生成された話者情報により示される話者と、文字情報生成部２１４により生成された文字情報により示される文字列とを互いに対応づけて表示するように、表示装置として機能可能な出力装置２５を制御する。その結果、出力装置２５は、例えば図６に示す画像を表示してよい。尚、表示制御部２１５は、話者を示す文字に代えて又は加えて、画像（例えば、アイコン）を表示するように、出力装置２５を制御してもよい。尚、分類装置２０は、話者情報により示される話者と文字情報により示される文字列とを表示するので、情報表示装置と称されてもよい。

[0067] 尚、話者情報生成部２１３が、話者認識処理及び声認証処理を行わない場合、表示制御部２１５は、例えば図７に示す話者入力画面を表示するように出力装置２５を制御してよい。分類装置２０のユーザは、入力装置２４を介して、話者入力画面に表示される話者のうち少なくとも一人の話者の人名を入力してよい。話者情報生成部２１３は、入力された人名に基づいて、話者情報により示される話者を変更してよい。表示制御部２１５は、話者情報生成部２１３により変更された話者情報により示される話者と、文字情報により示される文字列とを互いに対応づけて表示するように、出力装置２５を制御してよい。

[0068] （技術的効果）

音質が異なる複数の音声に対してクラスタリング処理（上述した第１のクラスタリング処理に相当）を行うと、複数の音声を話者毎に適切に分けられない可能性がある。例えば、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及び

ＶＳ５に対して、上述した第１のクラスタリング処理を行うと、音声区間ＶＳ１及びＶＳ２がグループＣ１に分類され、音声区間ＶＳ２がグループＣ２に分類され、音声区間ＶＳ４及びＶＳ５が、グループＣ１及びＣ２とは異なるグループＣ３に分類される。尚、計算過程の説明は省略する。

[0069] 上述したように、音声区間ＶＳ１及びＶＳ３に係る話者が第１話者であり、音声区間ＶＳ２、ＶＳ４及びＶＳ５に係る話者が第２話者である。従って、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５に対して、上述した第１のクラスタリング処理を行うと、音声区間ＶＳ１、ＶＳ２、ＶＳ３、ＶＳ４及びＶＳ５が話者毎に分けられていないことがわかる。

[0070] これに対して、分類装置２０では、分類部２１１により複数の音声区間が、第１音質を有する音声区間と、第２音質を有する音声区間とに分類される。クラスタリング部２１２により、第１音質を有する音声区間に分類された複数の音声区間に対して第１のクラスタリング処理が行われる。その後、クラスタリング部２１２により、第２音質を有する音声区間に分類された音声区間が、第１音質を有する音声区間に基づく複数のグループのいずれか、又は、該複数のグループとは異なる新たなグループに分類される。このように、複数の音声区間を音質で分類した上で、音質毎にクラスタリング処理を行うことによって、音質が異なる複数の音声区間を話者毎に適切に分けることができる。

[0071] 尚、第２音質より高音質である第１音質を有する複数の音声区間に対して第１のクラスタリング処理が行われることにより、第２音質を有する複数の音声区間に対して第１のクラスタリング処理が行われる場合に比べて、適切にグループ分けをすることができる。音声に相当する文字列が話者に対応づけられて表示されるので、実用上有用である。

[0072] ＜第３実施形態＞

分類装置、分類方法、記録媒体及び情報表示装置に係る実施形態について図８を参照して説明する。以下では、分類装置２０を用いて、分類装置、分類方法、記録媒体及び情報表示装置に係る実施形態を説明する。第３実施形

態では、表示制御部２１５の動作について主に説明する。その他の構成については、上述した第２実施形態と同様であってよい。第３実施形態について、第２実施形態の説明と重複する説明を適宜省略する。

[0073] 入力装置２４は、マイクロフォンを有していてよい。分類部２１１は、入力装置２４としてのマイクロフォンにより逐次取得される音に対して音声検出処理を行うことによって、音声区間を検出してよい。文字情報生成部２１４は、検出された音声区間に係る音声に基づいて、音声に相当する文字列を示す第１文字情報を生成してよい。表示制御部２１５は、文字情報生成部２１４により生成された第１文字情報により示される文字列を表示するように出力装置２５を制御してよい。その結果、出力装置２５は、例えば図８に示す画像を表示してよい。この場合、話者情報生成部２１３により話者情報が生成されていないので、話者を示す情報（例えば、文字又は画像）は表示されない。

[0074] 上記処理と並行して、分類部２１１は、複数の音声区間各々の話者特徴量を抽出してよい。分類部２１１は、抽出された話者特徴量に基づいて、複数の音声区間の夫々の間の類似度を計算してよい。その後、分類部２１１は、複数の音声区間を、第１音質を有する音声区間と、第２音質を有する音声区間とに分類してよい。クラスタリング部２１２は、第１音質を有する音声区間に分類された複数の音声区間に対して第１のクラスタリング処理を行ってよい。クラスタリング部２１２は、第２のクラスタリング処理を行うことによって、第２音質を有する音声区間に分類された音声区間を、第１音質を有する音声区間に基づく複数のグループのいずれか、又は、該複数のグループとは異なる新たなグループに分類してよい。

[0075] 話者情報生成部２１３は、クラスタリング処理（具体的には、第１のクラスタリング処理及び第２のクラスタリング処理）の結果に基づいて、複数の音声区間各々の話者を示す第１話者情報を生成してよい。表示制御部２１５は、話者情報生成部２１３により生成された第１話者情報により示される話者と、文字情報生成部２１４により生成された第１文字情報により示される

文字列を表示するように出力装置２５を制御してよい。その結果、出力装置２５は、例えば図６に示す画像を表示してよい。

[0076] 尚、分類装置２０は、第１の期間にマイクロフォンにより取得された音を一の入力音声とし、第１の期間に続く第２の期間にマイクロフォンにより取得された音を他の入力音声として扱ってよい。或いは、分類装置２０は、第１の期間にマイクロフォンにより取得された音と、第１の期間に続く第２の期間にマイクロフォンにより取得された音とを合わせて一の入力音声として扱ってよい。この場合、マイクロフォンにより新たな音が取得されることによって、一の入力音声が更新されることとなる。

[0077] 分類装置２０が、第１の期間にマイクロフォンにより取得された音と、第１の期間に続く第２の期間にマイクロフォンにより取得された音とを合わせて一の入力音声として扱う場合の動作について説明する。

[0078] 分類部２１１は、第１の期間にマイクロフォンにより取得された音である一の入力音声に対して音声検出処理を行うことによって、音声区間を検出してよい。分類部２１１は、複数の音声区間各々の話者特徴量を抽出してよい。分類部２１１は、抽出された話者特徴量に基づいて、複数の音声区間の夫々の間の類似度を計算してよい。その後、分類部２１１は、複数の音声区間を、第１音質を有する音声区間と、第２音質を有する音声区間とに分類してよい。クラスタリング部２１２は、第１音質を有する音声区間に分類された複数の音声区間に対して第１のクラスタリング処理を行ってよい。クラスタリング部２１２は、第２のクラスタリング処理を行うことによって、第２音質を有する音声区間に分類された音声区間を、第１音質を有する音声区間に基づく複数のグループのいずれか、又は、該複数のグループとは異なる新たなグループに分類してよい。話者情報生成部２１３は、クラスタリング処理の結果に基づいて、複数の音声区間各々の話者を示す第１話者情報を生成してよい。文字情報生成部２１４は、検出された音声区間に係る音声に基づいて、音声に相当する文字列を示す第１文字情報を生成してよい。表示制御部２１５は、話者情報生成部２１３により生成された第１話者情報により示さ

れる話者と、文字情報生成部２１４により生成された第１文字情報により示される文字列を表示するように出力装置２５を制御してよい。

[0079] 第１の期間に続く第２の期間にマイクロフォンにより取得された音が入力された場合、分類装置２０は、第１の期間にマイクロフォンにより取得された音と、第２の期間にマイクロフォンにより取得された音とを合わせることによって、一の入力音声を更新してよい。

[0080] 分類部２１１は、更新された一の入力音声に対して音声検出処理を行うことによって、音声区間を検出してよい。この場合、分類部２１１は、更新された一の入力音声に含まれる第１の期間にマイクロフォンにより取得された音に対して再度音声検出処理を行ってよい。尚、分類部２１１は、更新された一の入力音声に含まれる第１の期間にマイクロフォンにより取得された音に対して音声検出処理を行わなくてもよい。分類部２１１は、複数の音声区間各々の話者特徴量を抽出してよい。分類部２１１は、抽出された話者特徴量に基づいて、複数の音声区間の夫々の間の類似度を計算してよい。その後、分類部２１１は、複数の音声区間を、第１音質を有する音声区間と、第２音質を有する音声区間とに分類してよい。クラスタリング部２１２は、第１音質を有する音声区間に分類された複数の音声区間に対して第１のクラスタリング処理を行ってよい。クラスタリング部２１２は、第２のクラスタリング処理を行うことによって、第２音質を有する音声区間に分類された音声区間を、第１音質を有する音声区間に基づく複数のグループのいずれか、又は、該複数のグループとは異なる新たなグループに分類してよい。話者情報生成部２１３は、クラスタリング処理の結果に基づいて、複数の音声区間各々の話者を示す第２話者情報を生成してよい。文字情報生成部２１４は、検出された音声区間に係る音声に基づいて、音声に相当する文字列を示す第２文字情報を生成してよい。表示制御部２１５は、話者情報生成部２１３により生成された第２話者情報により示される話者と、文字情報生成部２１４により生成された第２文字情報により示される文字列を表示するように出力装置２５を制御してよい。

[0081] 第１文字情報及び第２文字情報の両方に含まれる一の文字列（例えば、第１の期間にマイクロフォンにより取得された音に含まれる一の音声に相当する一の文字列）の話者が、第１話者情報により示される話者と、第２話者情報により示される話者とで異なる場合、表示制御部２１５は、第２話者情報が生成された際に、上記一の文字列に対応づけて表示される話者を、第２話者情報により示される話者から、第２話者情報により示される話者に変更するように、出力装置２５を制御してよい。その結果、出力装置２５は、一の文字列に対応づけて表示される話者を、例えば「話者１」から「話者２」に変更してよい。

[0082] （技術的効果）

本実施形態に係る分類装置２０によれば、発話が行われている状況において、発話に相当する文字列をリアルタイムに表示することができる。このため、例えば会議の議事録をリアルタイムに生成することができる。

[0083] ＜付記＞

以上説明した実施形態に関して、更に以下の付記を開示する。

[0084] （付記１）

入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する第２分類手段と、

を備える分類装置。

[0085] （付記２）

前記第２分類手段は、前記複数の区間音声のうち一の区間音声と、前記複数の区間音声各々との類似の程度を夫々示す複数の第１指標値のうち、前記複数の第２区間音声のうち一の第２区間音声と、前記複数の群のうち一の群に含まれる、前記複数の第１区間音声の一又は複数の第１区間音声との類似の程度を示す一又は複数の第１指標値に基づいて、前記一の第２区間音声と

前記一の群との類似の程度を示す第２指標値を算出し、

前記第２分類手段は、前記複数の群各々について算出された複数の第２指標値のうち、最大の第２指標値であって、所定の閾値以上の第２指標値に対応する群に、前記一の第２区間音声を分類する

付記１に記載の分類装置。

[0086] （付記３）

前記第２分類手段は、前記複数の区間音声のうち一の区間音声と、前記複数の区間音声各々との類似の程度を夫々示す複数の第１指標値のうち、前記複数の第２区間音声のうち一の第２区間音声と、前記複数の群のうち一の群に含まれる、前記複数の第１区間音声の一又は複数の第１区間音声との類似の程度を示す一又は複数の第１指標値に基づいて、前記一の第２区間音声と前記一の群との類似の程度を示す第２指標値を算出し、

前記第２分類手段は、前記複数の群各々について算出された複数の第２指標値のいずれもが所定の閾値より小さい場合、前記新たな群に、前記一の第２区間音声を分類する

付記１又は２に記載の分類装置。

[0087] （付記４）

前記分類手段は、前記複数の区間音声各々の音質と、基準音質とに基づいて、前記複数の区間音声を、前記複数の第１区間音声と前記複数の第２区間音声とに分類する

付記１乃至３のいずれか一項に記載の分類装置。

[0088] （付記５）

前記第２音質は、前記第１音質より低い音質である

付記１乃至４のいずれか一項に記載の分類装置。

[0089] （付記６）

コンピュータが、

入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第

２音質を有する複数の第２区間音声とに分類し、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する

分類方法。

[0090] （付記７）

コンピュータに、

入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する

分類方法を実行させるコンピュータプログラムが記録されている記録媒体。

[0091] （付記８）

第１入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する第２分類手段と、

前記複数の群及び前記新たな群に基づいて、前記複数の区間音声各々の話者を示す第１話者情報を生成する第１生成手段と、

前記複数の区間音声の少なくとも一つの区間音声に基づいて、前記少なくとも一つの区間音声に相当する文字列を示す文字情報を生成する第２生成手段と、

前記生成された第１話者情報により示される話者と、前記生成された文字情報により示される文字列とを表示する表示手段と、

を備える情報表示装置。

[0092] （付記９）

前記表示手段は、前記少なくとも一つの区間音声のうち、前記複数の群及び前記新たな群のうち第１群に分類された第１区間音声に該当する区間音声に相当する文字列を、前記生成された第１話者情報により示される話者と対応付けて表示する

付記８に記載の情報表示装置。

[0093] （付記１０）

前記第１話者情報は、前記複数の群のうち第２群に分類された第１区間音声の話者である第１話者を示すとともに、前記第２群に分類された第２区間音声の話者として、前記第１話者を示唆する文字を示す

請求項８又は９に記載の情報表示装置。

[0094] （付記１１）

前記第１入力音声が複数の区間に分割されることによって生成された複数の区間音声に、第２入力音声が複数の区間に分割されることによって生成された他の複数の区間音声が加えられた場合、

前記第１分類手段は、前記他の複数の区間音声を、前記第１音質を有する複数の第３区間音声と、前記第２音質を有する複数の第４区間音声とに分類し、

前記第２分類手段は、前記複数の第１区間音声及び前記複数の第３区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声及び前記複数の第４区間音声を分類し、

前記第１生成手段は、前記複数の群及び前記新たな群に基づいて、前記複数の区間音声及び前記他の複数の区間音声各々の話者を示す第２話者情報を生成し、

前記生成された文字情報により示される一の文字列に相当する一の区間音声の話者が、前記第１話者情報により示される話者と、前記第２話者情報により示される話者とで異なる場合、前記表示手段は、前記第２話者情報が生成された際に、前記一の文字列に対応づけられる話者を、前記第１話者情報により示される話者から、前記第２話者情報により示される話者に変更する

付記８乃至１０のいずれか一項に記載の情報表示装置。

[0095] この開示は、上述した実施形態に限られるものではなく、請求の範囲及び明細書全体から読み取れる発明の要旨或いは思想に反しない範囲で適宜変更可能であり、そのような変更を伴う分類装置、分類方法、記録媒体及び情報表示装置もまたこの開示の技術的範囲に含まれるものである。

符号の説明

[0096] １０、２０　分類装置

１１、２１１　分類部

１２、２１２　クラスタリング部

２１３　話者情報生成部

２１４　文字情報生成部

２１５　表示制御部

請求の範囲

[請求項1] 入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する第２分類手段と、

を備える分類装置。

[請求項2] 前記第２分類手段は、前記複数の区間音声のうち一の区間音声と、前記複数の区間音声各々との類似の程度を夫々示す複数の第１指標値のうち、前記複数の第２区間音声のうち一の第２区間音声と、前記複数の群のうち一の群に含まれる、前記複数の第１区間音声の一又は複数の第１区間音声との類似の程度を示す一又は複数の第１指標値に基づいて、前記一の第２区間音声と前記一の群との類似の程度を示す第２指標値を算出し、

前記第２分類手段は、前記複数の群各々について算出された複数の第２指標値のうち、最大の第２指標値であって、所定の閾値以上の第２指標値に対応する群に、前記一の第２区間音声を分類する

請求項１に記載の分類装置。

[請求項3] 前記第２分類手段は、前記複数の区間音声のうち一の区間音声と、前記複数の区間音声各々との類似の程度を夫々示す複数の第１指標値のうち、前記複数の第２区間音声のうち一の第２区間音声と、前記複数の群のうち一の群に含まれる、前記複数の第１区間音声の一又は複数の第１区間音声との類似の程度を示す一又は複数の第１指標値に基づいて、前記一の第２区間音声と前記一の群との類似の程度を示す第２指標値を算出し、

前記第２分類手段は、前記複数の群各々について算出された複数の

第２指標値のいずれもが所定の閾値より小さい場合、前記新たな群に、前記一の第２区間音声を分類する

請求項１又は２に記載の分類装置。

[請求項4] 前記分類手段は、前記複数の区間音声各々の音質と、基準音質とに基づいて、前記複数の区間音声を、前記複数の第１区間音声と前記複数の第２区間音声とに分類する

請求項１乃至３のいずれか一項に記載の分類装置。

[請求項5] 前記第２音質は、前記第１音質より低い音質である

請求項１乃至４のいずれか一項に記載の分類装置。

[請求項6] コンピュータが、

入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する

分類方法。

[請求項7] コンピュータに、

入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類し、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声を分類する

分類方法を実行させるコンピュータプログラムが記録されている記録媒体。

[請求項8] 第１入力音声が複数の区間に分割されることによって生成された複数の区間音声を、第１音質を有する複数の第１区間音声と、前記第１音質とは異なる第２音質を有する複数の第２区間音声とに分類する第１分類手段と、

前記複数の第１区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新たな群に、前記複数の第２区間音声各々を分類する第２分類手段と、

前記複数の群及び前記新たな群に基づいて、前記複数の区間音声の話者を示す第１話者情報を生成する第１生成手段と、

前記複数の区間音声の少なくとも一つの区間音声に基づいて、前記少なくとも一つの区間音声に相当する文字列を示す文字情報を生成する第２生成手段と、

前記生成された第１話者情報により示される話者と、前記生成された文字情報により示される文字列とを表示する表示手段と、

を備える情報表示装置。

[請求項9] 前記表示手段は、前記少なくとも一つの区間音声のうち、前記複数の群及び前記新たな群のうち第１群に分類された第１区間音声に該当する区間音声に相当する文字列を、前記生成された第１話者情報により示される話者と対応付けて表示する

請求項８に記載の情報表示装置。

[請求項10] 前記第１話者情報は、前記複数の群のうち第２群に分類された第１区間音声の話者である第１話者を示すとともに、前記第２群に分類された第２区間音声の話者として、前記第１話者を示唆する文字を示す

請求項８又は９に記載の情報表示装置。

[請求項11] 前記第１入力音声が複数の区間に分割されることによって生成された複数の区間音声に、第２入力音声が複数の区間に分割されることによって生成された他の複数の区間音声が加えられた場合、

前記第１分類手段は、前記他の複数の区間音声を、前記第１音質を有する複数の第３区間音声と、前記第２音質を有する複数の第４区間音声とに分類し、

前記第２分類手段は、前記複数の第１区間音声及び前記複数の第３区間音声に基づく複数の群のいずれか又は前記複数の群とは異なる新

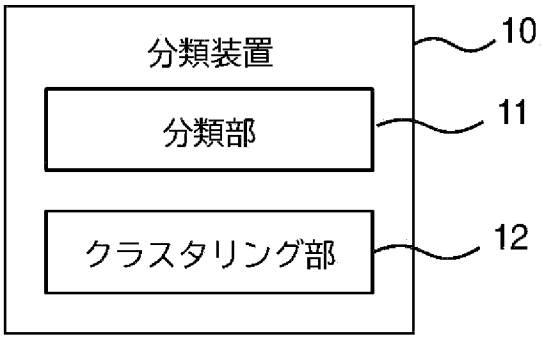
たな群に、前記複数の第２区間音声及び前記複数の第４区間音声を分類し、

前記第１生成手段は、前記複数の群及び前記新たな群に基づいて、前記複数の区間音声及び前記他の複数の区間音声各々の話者を示す第２話者情報を生成し、

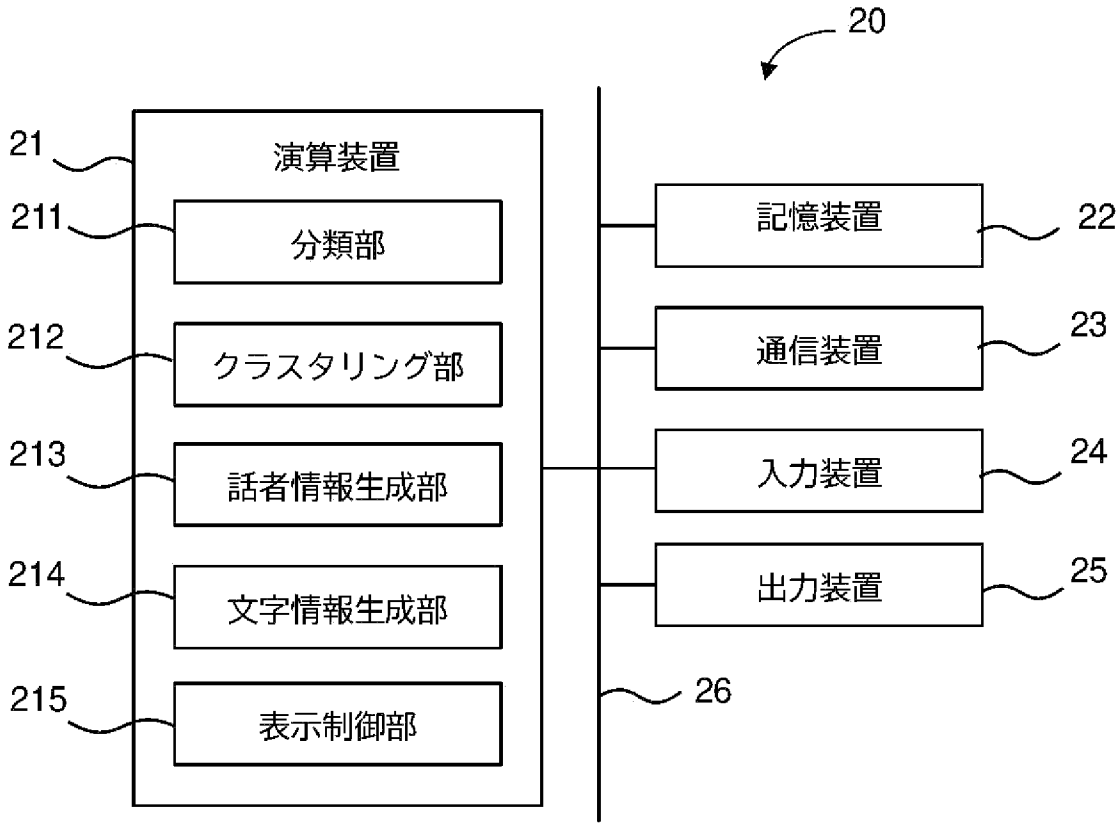
前記生成された文字情報により示される一の文字列に相当する一の区間音声の話者が、前記第１話者情報により示される話者と、前記第２話者情報により示される話者とで異なる場合、前記表示手段は、前記第２話者情報が生成された際に、前記一の文字列に対応づけられる話者を、前記第１話者情報により示される話者から、前記第２話者情報により示される話者に変更する

請求項８乃至１０のいずれか一項に記載の情報表示装置。

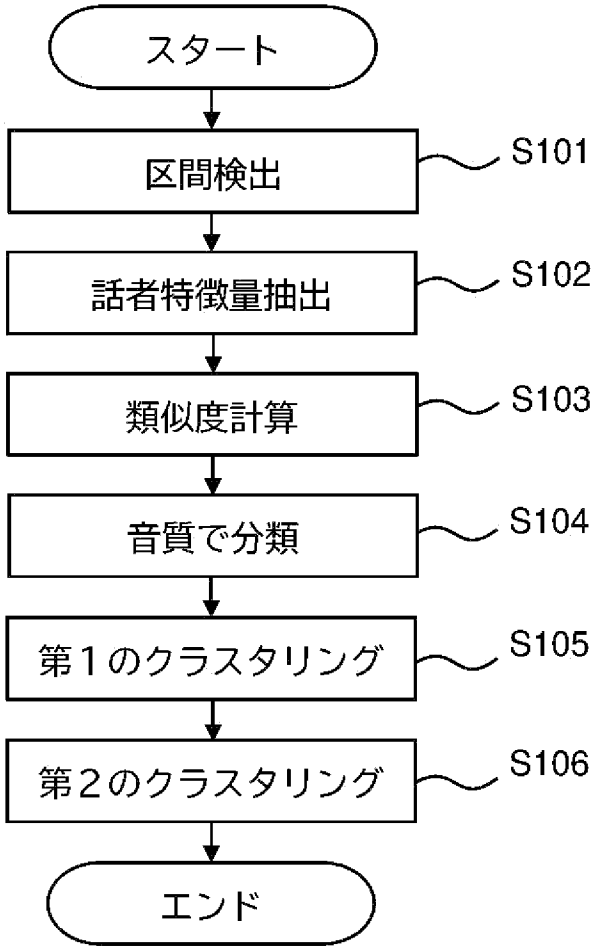
[図1]


分類装置
10
分類部
11
クラスタリング部
12

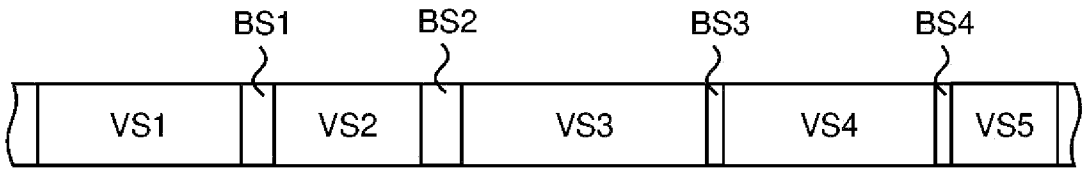
[図2]


20
21
演算装置
211
分類部
212
クラスタリング部
213
話者情報生成部
214
文字情報生成部
215
表示制御部
記憶装置
22
通信装置
23
入力装置
24
出力装置
25
26

[図3]



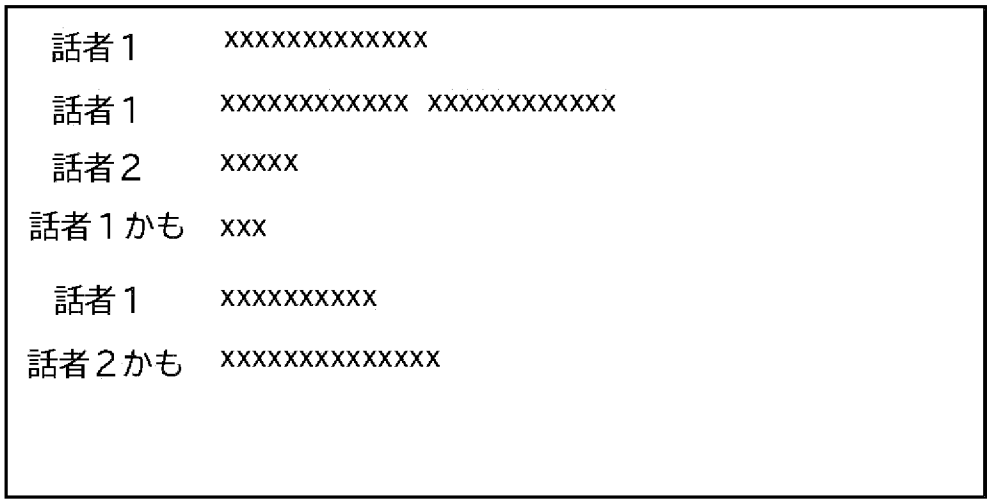
[図4]



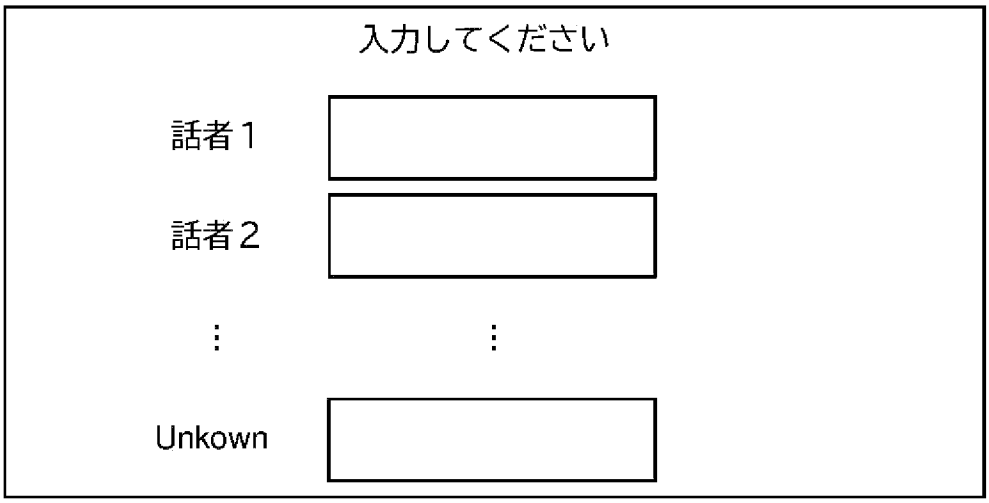
[図5]

	SF1	SF2	SF3	SF4	SF5
SF1	1	0.1	0.9	0.5	0.4
SF2	0.1	1	0.1	0.4	0.5
SF3	0.9	0.1	1	0.4	0.4
SF4	0.5	0.4	0.4	1	0.7
SF5	0.4	0.5	0.4	0.7	1

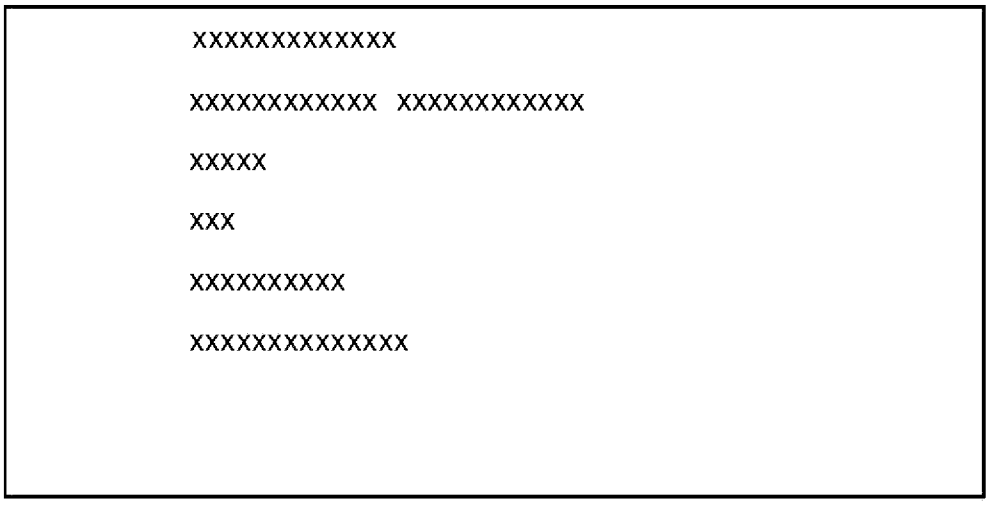
[図6]


話者１ xxxxxxxxxxxxx
話者１ xxxxxxxxxxxx xxxxxxxxxxxx
話者２ xxxxx
話者１かも xxx
話者１ xxxxxxxxxx
話者２かも xxxxxxxxxxxxx

[図7]


入力してください
話者１
話者２
Unkown

[図8]


xxxxxxxxxxxxx
xxxxxxxxxxxx xxxxxxxxxxxx
xxxxx
xxx
xxxxxxxxxx
xxxxxxxxxxxxx

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2023/022279

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/04(2013.01)i; ***G10L 17/02***(2013.01)i
FI: G10L15/04 200; G10L17/02

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)

G10L15/00-17/26

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

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2023
Registered utility model specifications of Japan 1996-2023
Published registered utility model applications of Japan 1994-2023

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.
A	WO 2018/163279 A1 (NEC CORPORATION) 13 September 2018 (2018-09-13) entire text	1-11
A	JP 2008-175955 A (KABUSHIKI KAISHA TOSHIBA) 31 July 2008 (2008-07-31) entire text	1-11
A	JP 2022-109867 A (NAVER CORP.) 28 July 2022 (2022-07-28) entire text	1-11
A	JP 2022-133118 A (PANASONIC INTELLECTUAL PROPERTY MANAGEMENT CO., LTD.) 13 September 2022 (2022-09-13) entire text	1-11
A	JP 2019-8131 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 17 January 2019 (2019-01-17) entire text	1-11
A	WO 2022/113218 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 02 June 2022 (2022-06-02) entire text	1-11

☐ 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	Date of mailing of the international search report
28 August 2023	**05 September 2023**
Name and mailing address of the ISA/JP **Japan Patent Office (ISA/JP)** **3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915** **Japan**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2023/022279

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2018/163279 A1	13 September 2018	US 2021/0134300 A1 entire text	
JP 2008-175955 A	31 July 2008	US 2008/0215324 A1 entire text	
JP 2022-109867 A	28 July 2022	US 2022/0230648 A1 entire text KR 10-2022-0103507 A	
JP 2022-133118 A	13 September 2022	(Family: none)	
JP 2019-8131 A	17 January 2019	(Family: none)	
WO 2022/113218 A1	02 June 2022	(Family: none)	

国際調査報告

国際出願番号
PCT/JP2023/022279

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G10L 15/04(2013.01)i; G10L 17/02(2013.01)i
FI: G10L15/04 200; G10L17/02

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））

G10L15/00-17/26

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	１９７１－２０２３年
日本国実用新案登録公報	１９９６－２０２３年
日本国登録実用新案公報	１９９４－２０２３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
A	WO 2018/163279 A1 (日本電気株式会社) 13.09.2018 (2018-09-13) 全文	1-11
A	JP 2008-175955 A (株式会社東芝) 31.07.2008 (2008-07-31) 全文	1-11
A	JP 2022-109867 A (ネイバー コーポレーション) 28.07.2022 (2022-07-28) 全文	1-11
A	JP 2022-133118 A (パナソニックＩＰマネジメント株式会社) 13.09.2022 (2022-09-13) 全文	1-11
A	JP 2019-8131 A (日本電信電話株式会社) 17.01.2019 (2019-01-17) 全文	1-11
A	WO 2022/113218 A1 (日本電信電話株式会社) 02.06.2022 (2022-06-02) 全文	1-11

☐ 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抵触するものではなく、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	国際調査報告の発送日
28.08.2023	05.09.2023

名称及びあて先
日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）
山下　剛史 5Z 8946

電話番号 03-3581-1101 内線 3591

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2023/022279

引用文献	公表日	パテントファミリー文献	公表日
WO 2018/163279 A1	13.09.2018	US 2021/0134300 A1 全文	
JP 2008-175955 A	31.07.2008	US 2008/0215324 A1 全文	
JP 2022-109867 A	28.07.2022	US 2022/0230648 A1 全文 KR 10-2022-0103507 A	
JP 2022-133118 A	13.09.2022	（ファミリーなし）	
JP 2019-8131 A	17.01.2019	（ファミリーなし）	
WO 2022/113218 A1	02.06.2022	（ファミリーなし）	