

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7696331号  
(P7696331)

(45)発行日 令和7年6月20日(2025.6.20)

(24)登録日 令和7年6月12日(2025.6.12)

|                         |         |       |         |  |
|-------------------------|---------|-------|---------|--|
| (51)国際特許分類              | F I     |       |         |  |
| G 1 0 L 17/02 (2013.01) | G 1 0 L | 17/02 |         |  |
| G 1 0 L 17/00 (2013.01) | G 1 0 L | 17/00 | 2 0 0 C |  |
| G 1 0 L 15/10 (2006.01) | G 1 0 L | 15/10 | 5 0 0 Z |  |

請求項の数 9 (全64頁)

|                   |                             |          |                         |
|-------------------|-----------------------------|----------|-------------------------|
| (21)出願番号          | 特願2022-509394(P2022-509394) | (73)特許権者 | 514136668               |
| (86)(22)出願日       | 令和3年2月15日(2021.2.15)        |          | パナソニック インテレクチュアル プロ     |
| (86)国際出願番号        | PCT/JP2021/005466           |          | パティ コーポレーション オブ アメリカ    |
| (87)国際公開番号        | WO2021/192719               |          | Panasonic Intellec      |
| (87)国際公開日         | 令和3年9月30日(2021.9.30)        |          | tual Property Corpo     |
| 審査請求日             | 令和5年12月20日(2023.12.20)      |          | ration of America       |
| (31)優先権主張番号       | 特願2020-57437(P2020-57437)   |          | アメリカ合衆国 9 0 5 0 4 カリフォル |
| (32)優先日           | 令和2年3月27日(2020.3.27)        |          | ニア州, トーランス, スイート 4 5 0  |
| (33)優先権主張国・地域又は機関 | 日本国(JP)                     | (74)代理人  | 100115381               |
|                   |                             |          | 弁理士 小谷 昌崇               |
|                   |                             | (74)代理人  | 100118049               |
|                   |                             |          | 弁理士 西谷 浩治               |
|                   |                             | (72)発明者  | 板倉 光佑                   |
|                   |                             |          | 大阪府門真市大字門真1 0 0 6 番地 パ  |
|                   |                             |          | 最終頁に続く                  |

(54)【発明の名称】 話者識別方法、話者識別装置、及び話者識別プログラム

(57)【特許請求の範囲】

【請求項1】

コンピュータが、  
識別対象音声データを取得し、  
予め登録されている登録音声データを取得し、  
前記識別対象音声データの特徴量を抽出し、  
前記登録音声データの特徴量を抽出し、  
前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第1話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第2話者識別モデルを選択し、  
前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別し、  
登録対象音声データを取得し、  
前記登録対象音声データの特徴量を抽出し、  
前記登録対象音声データの前記特徴量を用いて前記登録対象音声データの話者の性別を識別し、  
識別した前記性別に対応付けた前記登録対象音声データを前記登録音声データとして登

録し、

前記性別の識別において、

話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを取得し、

前記登録対象音声データの前記特徴量と、予め記憶されている男性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記男性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、

取得した複数の類似度の平均を平均男性類似度として算出し、

前記登録対象音声データの前記特徴量と、予め記憶されている女性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記女性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、

取得した複数の類似度の平均を平均女性類似度として算出し、

前記平均男性類似度が前記平均女性類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、

前記平均男性類似度が前記平均女性類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別する、

話者識別方法。

【請求項 2】

前記第 1 話者識別モデル又は前記第 2 話者識別モデルの選択において、

前記登録音声データの前記話者の性別が男性である場合、前記第 1 話者識別モデルを選択し、

前記登録音声データの前記話者の性別が女性である場合、前記第 2 話者識別モデルを選択する、

請求項 1 記載の話者識別方法。

【請求項 3】

コンピュータが、

識別対象音声データを取得し、

予め登録されている登録音声データを取得し、

前記識別対象音声データの特徴量を抽出し、

前記登録音声データの特徴量を抽出し、

前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第 1 話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第 2 話者識別モデルを選択し、

前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別し、

登録対象音声データを取得し、

前記登録対象音声データの特徴量を抽出し、

前記登録対象音声データの前記特徴量を用いて前記登録対象音声データの話者の性別を識別し、

識別した前記性別に対応付けた前記登録対象音声データを前記登録音声データとして登録し、

前記性別の識別において、

話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを取得し、

前記登録対象音声データの前記特徴量と、予め記憶されている男性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記男性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、

10

20

30

40

50

取得した複数の類似度のうちの最大値を最大男性類似度として算出し、

前記登録対象音声データの前記特微量と、予め記憶されている女性の複数の音声データの特微量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記女性の複数の音声データそれぞれの類似度を前記性別識別モデルから取得し、

取得した複数の類似度のうちの最大値を最大女性類似度として算出し、

前記最大男性類似度が前記最大女性類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、

前記最大男性類似度が前記最大女性類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別する、

話者識別方法。

【請求項 4】

コンピュータが、

識別対象音声データを取得し、

予め登録されている登録音声データを取得し、

前記識別対象音声データの特微量を抽出し、

前記登録音声データの特微量を抽出し、

前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第 1 話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第 2 話者識別モデルを選択し、

前記識別対象音声データの前記特微量と前記登録音声データの前記特微量とを、選択した前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別し、

前記登録音声データは、複数の登録音声データを含み、

前記話者の識別において、

前記識別対象音声データの前記特微量と前記複数の登録音声データの前記特微量それぞれとを、選択した前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかに入力することにより、前記識別対象音声データと前記複数の登録音声データそれぞれの類似度を前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかから取得し、

取得した類似度が最も高い前記登録音声データの話者を前記識別対象音声データの話者として識別する、

話者識別方法。

【請求項 5】

前記登録音声データは、複数の登録音声データを含み、

前記複数の登録音声データは、前記複数の登録音声データそれぞれの話者を識別するための識別情報と対応付けられており、

さらに、前記識別対象音声データの話者を識別するための識別情報を取得し、

前記登録音声データの取得において、前記複数の登録音声データの中から、取得した前記識別情報と一致する識別情報が対応付けられている登録音声データを取得し、

前記話者の識別において、

前記識別対象音声データの前記特微量と前記登録音声データの前記特微量とを、選択した前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかに入力することにより、前記識別対象音声データと前記登録音声データとの類似度を前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかから取得し、

取得した前記類似度が閾値よりも高い場合、前記登録音声データの話者を前記識別対象音声データの話者として識別する、

請求項 1 ～ 3 のいずれか 1 項に記載の話者識別方法。

【請求項 6】

機械学習時に、男性の複数の音声データのうちの 2 つの音声データの特微量の全ての組

10

20

30

40

50

み合わせが前記第1話者識別モデルに入力されることにより、前記2つの音声データの複数の組み合わせそれぞれの類似度が前記第1話者識別モデルから取得され、同一話者の前記2つの音声データの前記類似度と、互いに異なる話者の前記2つの音声データの前記類似度とを識別可能な第1閾値が算出され、

機械学習時に、女性の複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせが前記第2話者識別モデルに入力されることにより、前記2つの音声データの複数の組み合わせそれぞれの類似度が前記第2話者識別モデルから取得され、同一話者の前記2つの音声データの前記類似度と、互いに異なる話者の前記2つの音声データの前記類似度とを識別可能な第2閾値が算出され、

前記話者の識別において、前記第1話者識別モデルから前記類似度を取得した場合、取得した前記類似度から前記第1閾値を減算し、前記第2話者識別モデルから前記類似度を取得した場合、取得した前記類似度から前記第2閾値を減算する、

請求項4記載の話者識別方法。

【請求項7】

コンピュータが、  
識別対象音声データを取得し、  
予め登録されている登録音声データを取得し、  
前記識別対象音声データの特徴量を抽出し、  
前記登録音声データの特徴量を抽出し、  
前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第1話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第2話者識別モデルを選択し、  
前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別し、  
前記登録音声データは、複数の登録音声データを含み、  
前記複数の登録音声データは、前記複数の登録音声データそれぞれの話者を識別するための識別情報と対応付けられており、  
さらに、前記識別対象音声データの話者を識別するための識別情報を取得し、  
前記登録音声データの取得において、前記複数の登録音声データの中から、取得した前記識別情報と一致する識別情報が対応付けられている登録音声データを取得し、  
前記話者の識別において、  
前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データと前記登録音声データとの類似度を前記第1話者識別モデル及び前記第2話者識別モデルのいずれかから取得し、  
取得した前記類似度が閾値よりも高い場合、前記登録音声データの話者を前記識別対象音声データの話者として識別し、  
機械学習時に、男性の複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせが前記第1話者識別モデルに入力されることにより、前記2つの音声データの複数の組み合わせそれぞれの類似度が前記第1話者識別モデルから取得され、同一話者の前記2つの音声データの前記類似度と、互いに異なる話者の前記2つの音声データの前記類似度とを識別可能な第1閾値が算出され、  
機械学習時に、女性の複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせが前記第2話者識別モデルに入力されることにより、前記2つの音声データの複数の組み合わせそれぞれの類似度が前記第2話者識別モデルから取得され、同一話者の前記2つの音声データの前記類似度と、互いに異なる話者の前記2つの音声データの前記類似度とを識別可能な第2閾値が算出され、

10

20

30

40

50

前記話者の識別において、前記第 1 話者識別モデルから前記類似度を取得した場合、取得した前記類似度から前記第 1 閾値を減算し、前記第 2 話者識別モデルから前記類似度を取得した場合、取得した前記類似度から前記第 2 閾値を減算する、  
話者識別方法。

【請求項 8】

識別対象音声データを取得する識別対象音声データ取得部と、  
予め登録されている登録音声データを取得する登録音声データ取得部と、  
前記識別対象音声データの特徴量を抽出する第 1 抽出部と、  
前記登録音声データの特徴量を抽出する第 2 抽出部と、  
前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第 1 話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第 2 話者識別モデルを選択する話者識別モデル選択部と、  
前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第 1 話者識別モデル及び前記第 2 話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別する話者識別部と、  
登録対象音声データを取得する登録対象音声データ取得部と、  
前記登録対象音声データの特徴量を抽出する第 3 抽出部と、  
前記登録対象音声データの前記特徴量を用いて前記登録対象音声データの話者の性別を識別する性別識別部と、  
識別した前記性別に対応付けた前記登録対象音声データを前記登録音声データとして登録する登録部と、

を備え、

前記性別識別部は、  
話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを取得し、  
前記登録対象音声データの前記特徴量と、予め記憶されている男性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記男性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、  
取得した複数の類似度の平均を平均男性類似度として算出し、  
前記登録対象音声データの前記特徴量と、予め記憶されている女性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記女性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、  
取得した複数の類似度の平均を平均女性類似度として算出し、  
前記平均男性類似度が前記平均女性類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、  
前記平均男性類似度が前記平均女性類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別する、  
話者識別装置。

【請求項 9】

識別対象音声データを取得し、  
予め登録されている登録音声データを取得し、  
前記識別対象音声データの特徴量を抽出し、  
前記登録音声データの特徴量を抽出し、  
前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第 1 話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第 2 話者識別モデルを選択し、

前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別し、  
登録対象音声データを取得し、  
前記登録対象音声データの特徴量を抽出し、  
前記登録対象音声データの前記特徴量を用いて前記登録対象音声データの話者の性別を識別し、  
識別した前記性別に対応付けた前記登録対象音声データを前記登録音声データとして登録するようにコンピュータを機能させ、  
前記性別の識別において、  
話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを取得し、  
前記登録対象音声データの前記特徴量と、予め記憶されている男性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記男性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、  
取得した複数の類似度の平均を平均男性類似度として算出し、  
前記登録対象音声データの前記特徴量と、予め記憶されている女性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記女性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、  
取得した複数の類似度の平均を平均女性類似度として算出し、  
前記平均男性類似度が前記平均女性類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、  
前記平均男性類似度が前記平均女性類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別する、

話者識別プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、話者を識別する技術に関するものである。

【背景技術】

【0002】

従来、識別対象の話者の音声データを取得し、取得した音声データに基づいて識別対象の話者が予め登録されている話者であるか否かを識別する技術が知られている。従来の話者識別では、識別対象話者の音声データの特徴量と、登録話者の音声データの特徴量との類似度が算出され、算出された類似度が閾値以上であれば、識別対象話者と登録話者とが同一であると判定される。

【0003】

例えば、非特許文献1では、話者識別のための高精度な特徴量として、*i - v e c t o r*と呼ばれる話者固有の特徴量について開示されている。

【0004】

また、例えば、非特許文献2では、*i - v e c t o r*に代わる特徴量として *x - v e c t o r* が開示されている。*x - v e c t o r* は、ディープラーニングにより生成されたディープニューラルネットワークに音声データが入力されることによって抽出される特徴量である。

【0005】

しかしながら、上記従来の技術では、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させるために、更なる改善が必要とされていた。

【先行技術文献】

【非特許文献】

【0006】

10

20

30

40

50

【文献】Najim Dehak、Patrick Kenny、Reda Dehak、Pierre Dumouchel、Pierre Ouellet、「Front-End Factor Analysis For Speaker Verification」、IEEE Transactions on Audio, Speech and Language Processing、2011年

【文献】David Snyder、Daniel Garcia-Romero、Gregory Sell、Daniel Povey、Sanjeev Khudanpur、「X-Vectors: Robust DNN Embeddings for Speaker Recognition」、IEEE、ICASSP 2018年

【発明の概要】

【0007】

本開示は、上記の問題を解決するためになされたもので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる技術を提供することを目的とするものである。

【0008】

本開示の一態様に係る話者識別方法は、コンピュータが、識別対象音声データを取得し、予め登録されている登録音声データを取得し、前記識別対象音声データの特徴量を抽出し、前記登録音声データの特徴量を抽出し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第1話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第2話者識別モデルを選択し、前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別する。

【0009】

本開示によれば、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

【図面の簡単な説明】

【0010】

【図1】本開示の実施の形態1における話者識別システムの構成を示す図である。

【図2】本開示の実施の形態1における性別識別モデル生成装置の構成を示す図である。

【図3】本開示の実施の形態1における話者識別モデル生成装置の構成を示す図である。

【図4】本実施の形態1における話者識別装置の登録処理の動作について説明するためのフローチャートである。

【図5】図4のステップS3の性別識別処理の動作について説明するためのフローチャートである。

【図6】本実施の形態1における話者識別装置の話者識別処理の動作について説明するための第1のフローチャートである。

【図7】本実施の形態1における話者識別装置の話者識別処理の動作について説明するための第2のフローチャートである。

【図8】本実施の形態1における性別識別モデル生成装置の性別識別モデル生成処理の動作について説明するためのフローチャートである。

【図9】本実施の形態1における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第1のフローチャートである。

【図10】本実施の形態1における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第2のフローチャートである。

【図11】従来の話者識別装置の話者識別性能評価結果と、実施の形態1の話者識別装置の話者識別性能評価結果とを示す図である。

【図12】本実施の形態1の変形例1における性別識別処理の動作について説明するため

10

20

30

40

50

のフローチャートである。

【図 1 3】本実施の形態 1 の変形例 2 における性別識別処理の動作について説明するためのフローチャートである。

【図 1 4】本開示の実施の形態 2 における話者識別システムの構成を示す図である。

【図 1 5】本実施の形態 2 における話者識別装置の登録処理の動作について説明するためのフローチャートである。

【図 1 6】本実施の形態 2 における話者識別装置の話者識別処理の動作について説明するための第 1 のフローチャートである。

【図 1 7】本実施の形態 2 における話者識別装置の話者識別処理の動作について説明するための第 2 のフローチャートである。

【図 1 8】本開示の実施の形態 3 における話者識別モデル生成装置の構成を示す図である。

【図 1 9】本開示の実施の形態 3 における話者識別システムの構成を示す図である。

【図 2 0】本実施の形態 3 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 1 のフローチャートである。

【図 2 1】本実施の形態 3 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 2 のフローチャートである。

【図 2 2】本実施の形態 3 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 3 のフローチャートである。

【図 2 3】本実施の形態 3 における話者識別装置の話者識別処理の動作について説明するための第 1 のフローチャートである。

【図 2 4】本実施の形態 3 における話者識別装置の話者識別処理の動作について説明するための第 2 のフローチャートである。

【図 2 5】本開示の実施の形態 4 における話者識別システムの構成を示す図である。

【図 2 6】本開示の実施の形態 4 における話者識別モデル生成装置の構成を示す図である。

【図 2 7】本実施の形態 4 における話者識別装置の登録処理の動作について説明するためのフローチャートである。

【図 2 8】本実施の形態 4 における話者識別装置の話者識別処理の動作について説明するための第 1 のフローチャートである。

【図 2 9】本実施の形態 4 における話者識別装置の話者識別処理の動作について説明するための第 2 のフローチャートである。

【図 3 0】本実施の形態 4 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 1 のフローチャートである。

【図 3 1】本実施の形態 4 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 2 のフローチャートである。

【図 3 2】本開示の実施の形態 5 における話者識別システムの構成を示す図である。

【図 3 3】本実施の形態 5 における話者識別装置の登録処理の動作について説明するためのフローチャートである。

【図 3 4】本実施の形態 5 における話者識別装置の話者識別処理の動作について説明するための第 1 のフローチャートである。

【図 3 5】本実施の形態 5 における話者識別装置の話者識別処理の動作について説明するための第 2 のフローチャートである。

【図 3 6】本開示の実施の形態 6 における話者識別モデル生成装置の構成を示す図である。

【図 3 7】本開示の実施の形態 6 における話者識別システムの構成を示す図である。

【図 3 8】本実施の形態 6 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 1 のフローチャートである。

【図 3 9】本実施の形態 6 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 2 のフローチャートである。

【図 4 0】本実施の形態 6 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 3 のフローチャートである。

【図 4 1】本実施の形態 6 における話者識別装置の話者識別処理の動作について説明する

10

20

30

40

50

ための第1のフローチャートであ

【図42】本実施の形態6における話者識別装置の話者識別処理の動作について説明するための第2のフローチャートである。

【発明を実施するための形態】

【0011】

(本開示の基礎となった知見)

男性と女性とでは、発話した音声データの特徴量の分布が異なる。しかしながら、従来の話者識別装置は、話者の性別に関係なく収集した音声データから生成された話者識別モデルを用いて、識別対象の音声データの話者を識別している。このように、従来技術では、男性の話者及び女性の話者のそれぞれに特化した話者識別については検討されていない。そのため、性別を考慮した話者識別が行われることにより、話者識別の精度が向上する可能性がある。

【0012】

以上の課題を解決するために、本開示の一態様に係る話者識別方法は、コンピュータが、識別対象音声データを取得し、予め登録されている登録音声データを取得し、前記識別対象音声データの特徴量を抽出し、前記登録音声データの特徴量を抽出し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第1話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第2話者識別モデルを選択し、前記識別対象音声データの登録特徴量と前記登録音声データの登録特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの登録話者を識別する。

【0013】

この構成によれば、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、男性用に生成された第1話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、女性用に生成された第2話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

【0014】

したがって、音声データの特徴量の分布が性別によって異なる場合であっても、それぞれの性別に特化した第1話者識別モデル及び第2話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

【0015】

また、上記の話者識別方法において、前記第1話者識別モデル又は前記第2話者識別モデルの選択において、前記登録音声データの登録話者の性別が男性である場合、前記第1話者識別モデルを選択し、前記登録音声データの登録話者の性別が女性である場合、前記第2話者識別モデルを選択してもよい。

【0016】

この構成によれば、登録音声データの話者の性別に応じて、第1話者識別モデル及び第2話者識別モデルのいずれかが選択される。したがって、登録時において予め登録音声データの話者の性別が1回だけ識別されることにより、話者識別毎に識別対象音声データの性別を識別する必要がなくなり、話者識別の処理負荷を軽減することができる。

【0017】

また、上記の話者識別方法において、さらに、登録対象音声データを取得し、さらに、前記登録対象音声データの特徴量を抽出し、さらに、前記登録対象音声データの登録特徴量を用いて前記登録対象音声データの話者の性別を識別し、さらに、識別した前記性別に

10

20

30

40

50

対応付けた前記登録対象音声データを前記登録音声データとして登録してもよい。

【0018】

この構成によれば、登録時において、性別に対応付けた登録音声データを予め登録することができ、話者識別時において、予め登録されている登録音声データに対応付けられている性別を用いて第1話者識別モデル及び第2話者識別モデルのいずれかを容易に選択することができる。

【0019】

また、上記の話者識別方法において、前記性別の識別において、話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを取得し、前記登録対象音声データの前記特徴量を前記性別識別モデルに入力することにより、前記登録対象音声データの前記話者の性別を識別してもよい。

10

【0020】

この構成によれば、話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルに登録対象音声データが入力されるだけで、登録対象音声データの話者の性別を容易に識別することができる。

【0021】

また、上記の話者識別方法において、前記性別の識別において、前記登録対象音声データの前記特徴量と、予め記憶されている男性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記男性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、取得した複数の類似度の平均を平均男性類似度として算出し、前記登録対象音声データの前記特徴量と、予め記憶されている女性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記女性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、取得した複数の類似度の平均を平均女性類似度として算出し、前記平均男性類似度が前記平均女性類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、前記平均男性類似度が前記平均女性類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別してもよい。

20

【0022】

この構成によれば、登録対象音声データの話者が男性である場合、登録対象音声データの特徴量と複数の男性の音声データの特徴量との平均類似度は、登録対象音声データの特徴量と複数の女性の音声データの特徴量との平均類似度よりも高くなる。また、登録対象音声データの話者が女性である場合、登録対象音声データの特徴量と複数の女性の音声データの特徴量との平均類似度は、登録対象音声データの特徴量と複数の男性の音声データの特徴量との平均類似度よりも高くなる。そのため、登録対象音声データの特徴量と複数の男性の音声データの特徴量との平均類似度と、登録対象音声データの特徴量と複数の女性の音声データの特徴量との平均類似度とが比較されることにより、登録対象音声データの話者の性別を容易に識別することができる。

30

【0023】

また、上記の話者識別方法において、前記性別の識別において、前記登録対象音声データの前記特徴量と、予め記憶されている男性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記男性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、取得した複数の類似度のうちの最大値を最大男性類似度として算出し、前記登録対象音声データの前記特徴量と、予め記憶されている女性の複数の音声データの特徴量それぞれとを前記性別識別モデルに入力することにより、前記登録対象音声データと前記女性の複数の音声データそれぞれとの類似度を前記性別識別モデルから取得し、取得した複数の類似度のうちの最大値を最大女性類似度として算出し、前記最大男性類似度が前記最大女性類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、前記最大男性類似度が前記最大女性類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別してもよい。

40

50

## 【0024】

この構成によれば、登録対象音声データの話者が男性である場合、登録対象音声データの特徴量と複数の男性の音声データの特徴量との複数の類似度のうちの最大類似度は、登録対象音声データの特徴量と複数の女性の音声データの特徴量との複数の類似度のうちの最大類似度よりも高くなる。また、登録対象音声データの話者が女性である場合、登録対象音声データの特徴量と複数の女性の音声データの特徴量との複数の類似度のうちの最大類似度は、登録対象音声データの特徴量と複数の男性の音声データの特徴量との複数の類似度のうちの最大類似度よりも高くなる。そのため、登録対象音声データの特徴量と複数の男性の音声データの特徴量との複数の類似度のうちの最大類似度と、登録対象音声データの特徴量と複数の女性の音声データの特徴量との複数の類似度のうちの最大類似度とが比較されることにより、登録対象音声データの話者の性別を容易に識別することができる。

10

## 【0025】

また、上記の話者識別方法において、前記性別の識別において、予め記憶されている男性の複数の音声データの平均特徴量を算出し、前記登録対象音声データの前記特徴量と男性の複数の音声データの前記平均特徴量とを前記性別識別モデルに入力することにより、前記登録対象音声データと男性の音声データ群との第1類似度を前記性別識別モデルから取得し、予め記憶されている女性の複数の音声データの平均特徴量を算出し、前記登録対象音声データの前記特徴量と女性の複数の音声データの前記平均特徴量とを前記性別識別モデルに入力することにより、前記登録対象音声データと女性の音声データ群との第2類似度を前記性別識別モデルから取得し、前記第1類似度が前記第2類似度より高い場合、前記登録対象音声データの前記話者の性別を男性と識別し、前記第1類似度が前記第2類似度より低い場合、前記登録対象音声データの前記話者の性別を女性と識別してもよい。

20

## 【0026】

この構成によれば、登録対象音声データの話者が男性である場合、登録対象音声データの特徴量と複数の男性の音声データの平均特徴量との第1類似度は、登録対象音声データの特徴量と複数の女性の音声データの平均特徴量との第2類似度よりも高くなる。また、登録対象音声データの話者が女性である場合、登録対象音声データの特徴量と複数の女性の音声データの平均特徴量との第2類似度は、登録対象音声データの特徴量と複数の男性の音声データの平均特徴量との第1類似度よりも高くなる。そのため、登録対象音声データの特徴量と複数の男性の音声データの平均特徴量との第1類似度と、登録対象音声データの特徴量と複数の女性の音声データの平均特徴量との第2類似度とが比較されることにより、登録対象音声データの話者の性別を容易に識別することができる。

30

## 【0027】

また、上記の話者識別方法において、前記登録音声データは、複数の登録音声データを含み、前記話者の識別において、前記識別対象音声データの前記特徴量と前記複数の登録音声データの前記特徴量それぞれとを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データと前記複数の登録音声データそれぞれとの類似度を前記第1話者識別モデル及び前記第2話者識別モデルのいずれかから取得し、取得した類似度が最も高い前記登録音声データの話者を前記識別対象音声データの話者として識別してもよい。

40

## 【0028】

この構成によれば、第1話者識別モデル及び第2話者識別モデルのいずれかから、識別対象音声データと複数の登録音声データそれぞれとの類似度が取得され、類似度が最も高い登録音声データの話者が識別対象音声データの話者として識別される。そのため、複数の登録音声データの中から、最も類似する登録音声データの話者を識別対象音声データの話者として識別することができる。

## 【0029】

また、上記の話者識別方法において、前記登録音声データは、複数の登録音声データを含み、前記複数の登録音声データは、前記複数の登録音声データそれぞれの話者を識別するための識別情報と対応付けられており、さらに、前記識別対象音声データの話者を識別

50

するための識別情報を取得し、前記登録音声データの取得において、前記複数の登録音声データの中から、取得した前記識別情報と一致する識別情報が対応付けられている登録音声データを取得し、前記話者の識別において、前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データと前記登録音声データとの類似度を前記第1話者識別モデル及び前記第2話者識別モデルのいずれかから取得し、取得した前記類似度が閾値よりも高い場合、前記登録音声データの話者を前記識別対象音声データの話者として識別してもよい。

#### 【0030】

この構成によれば、複数の登録音声データの中から、識別対象音声データの話者を識別するための識別情報と一致する識別情報が対応付けられている1の登録音声データが取得される。そのため、複数の登録音声データのうちの全ての登録音声データの特徴量と識別対象音声データの特徴量との類似度を算出する必要がなく、複数の登録音声データのうちの1の登録音声データの特徴量と識別対象音声データの特徴量との類似度を算出すればよいので、話者識別の処理負荷を低減することができる。

#### 【0031】

また、上記の話者識別方法において、機械学習時に、男性の複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせが前記第1話者識別モデルに入力されることにより、前記2つの音声データの複数の組み合わせそれぞれの類似度が前記第1話者識別モデルから取得され、同一話者の前記2つの音声データの前記類似度と、互いに異なる話者の前記2つの音声データの前記類似度とを識別可能な第1閾値が算出され、機械学習時に、女性の複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせが前記第2話者識別モデルに入力されることにより、前記2つの音声データの複数の組み合わせそれぞれの類似度が前記第2話者識別モデルから取得され、同一話者の前記2つの音声データの前記類似度と、互いに異なる話者の前記2つの音声データの前記類似度とを識別可能な第2閾値が算出され、前記話者の識別において、前記第1話者識別モデルから前記類似度を取得した場合、取得した前記類似度から前記第1閾値を減算し、前記第2話者識別モデルから前記類似度を取得した場合、取得した前記類似度から前記第2閾値を減算してもよい。

#### 【0032】

2つの異なる第1話者識別モデルと第2話者識別モデルとが用いられる場合、第1話者識別モデルと第2話者識別モデルとの出力値の範囲が異なるおそれがある。そこで、登録時において、第1話者識別モデル及び第2話者識別モデルのそれぞれに対し、同一の話者を識別することが可能な第1閾値及び第2閾値が算出される。また、話者識別時において、算出された識別対象音声データと登録音声データとの類似度から第1閾値又は第2閾値が減算されることにより類似度が補正される。そして、補正された類似度が第1話者識別モデル及び第2話者識別モデルに共通の閾値と比較されることにより、識別対象音声データの話者をより高い精度で識別することができる。

#### 【0033】

本開示の他の態様に係る話者識別装置は、識別対象音声データを取得する識別対象音声データ取得部と、予め登録されている登録音声データを取得する登録音声データ取得部と、前記識別対象音声データの特徴量を抽出する第1抽出部と、前記登録音声データの特徴量を抽出する第2抽出部と、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第1話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第2話者識別モデルを選択する話者識別モデル選択部と、前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別する話者識別部と、を備える。

## 【0034】

この構成によれば、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、男性用に生成された第1話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、女性用に生成された第2話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

## 【0035】

したがって、音声データの特徴量の分布が性別によって異なる場合であっても、それぞれの性別に特化した第1話者識別モデル及び第2話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

10

## 【0036】

本開示の他の態様に係る話者識別プログラムは、識別対象音声データを取得し、予め登録されている登録音声データを取得し、前記識別対象音声データの特徴量を抽出し、前記登録音声データの特徴量を抽出し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第1話者識別モデルを選択し、前記識別対象音声データの話者及び前記登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第2話者識別モデルを選択し、前記識別対象音声データの前記特徴量と前記登録音声データの前記特徴量とを、選択した前記第1話者識別モデル及び前記第2話者識別モデルのいずれかに入力することにより、前記識別対象音声データの前記話者を識別するようにコンピュータを機能させる。

20

## 【0037】

この構成によれば、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、男性用に生成された第1話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、女性用に生成された第2話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

30

## 【0038】

したがって、音声データの特徴量の分布が性別によって異なる場合であっても、それぞれの性別に特化した第1話者識別モデル及び第2話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

## 【0039】

本開示の他の態様に係る性別識別モデル生成方法は、コンピュータが、男性及び女性のいずれかを示す性別ラベルが付与された複数の音声データを取得し、前記複数の音声データのうちの第1音声データ及び第2音声データの各特徴量と、前記第1音声データ及び前記第2音声データの各性別ラベルの類似度とを教師データとして用い、入力を2つの音声データの各特徴量とし、出力を前記2つの音声データの類似度とする性別識別モデルを機械学習により生成する。

40

## 【0040】

この構成によれば、機械学習により生成された性別識別モデルに、登録音声データ又は識別対象音声データの特徴量と、男性の音声データの特徴量とが入力されることにより、2つの音声データの第1類似度が出力される。また、性別識別モデルに、登録音声データ又は識別対象音声データの特徴量と、女性の音声データの特徴量とが入力されることにより、2つの音声データの第2類似度が出力される。そして、第1類似度と第2類似度とが

50

比較されることにより、登録音声データ又は識別対象音声データの話者の性別を容易に推定することができる。

【0041】

本開示の他の態様に係る話者識別モデル生成方法は、コンピュータが、男性である話者を識別するための話者識別ラベルが付与された複数の男性音声データを取得し、前記複数の男性音声データのうちの第1男性音声データ及び第2男性音声データの各特徴量と、前記第1男性音声データ及び前記第2男性音声データの各話者識別ラベルの類似度とを教師データとして用い、入力を2つの音声データの各特徴量とし、出力を前記2つの音声データの類似度とする第1話者識別モデルを機械学習により生成し、女性である話者を識別するための話者識別ラベルが付与された複数の女性音声データを取得し、前記複数の女性音声データのうちの第1女性音声データ及び第2女性音声データの各特徴量と、前記第1女性音声データ及び前記第2女性音声データの各話者識別ラベルの類似度とを教師データとして用い、入力を2つの音声データの各特徴量とし、出力を前記2つの音声データの類似度とする第2話者識別モデルを機械学習により生成する。

10

【0042】

この構成によれば、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、男性用に生成された第1話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、女性用に生成された第2話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

20

【0043】

したがって、音声データの特徴量の分布が性別によって異なる場合であっても、男性に特化した第1話者識別モデル及び女性に特化した第2話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

【0044】

以下添付図面を参照しながら、本開示の実施の形態について説明する。なお、以下の実施の形態は、本開示を具体化した一例であって、本開示の技術的範囲を限定するものではない。

30

【0045】

(実施の形態1)

図1は、本開示の実施の形態1における話者識別システムの構成を示す図である。

【0046】

図1に示す話者識別システムは、マイクロホン1及び話者識別装置2を備える。なお、話者識別装置2は、マイクロホン1を備えていなくてもよいし、備えてもよい。

【0047】

マイクロホン1は、話者が発話した音声を收音し、音声データに変換し、話者識別装置2へ出力する。マイクロホン1は、音声データを予め登録する際に、話者が発話した登録対象音声データを話者識別装置2へ出力する。また、マイクロホン1は、話者を識別する際に、話者が発話した識別対象音声データを話者識別装置2へ出力する。

40

【0048】

話者識別装置2は、登録対象音声データ取得部201、特徴量抽出部202、性別識別モデル記憶部203、性別識別用音声データ記憶部204、性別識別部205、登録部206、識別対象音声データ取得部211、登録音声データ記憶部212、登録音声データ取得部213、特徴量抽出部214、特徴量抽出部215、話者識別モデル記憶部216、モデル選択部217、話者識別部218及び識別結果出力部219を備える。

【0049】

登録対象音声データ取得部201、特徴量抽出部202、性別識別部205、登録部2

50

06、識別対象音声データ取得部211、登録音声データ取得部213、特徴量抽出部214、特徴量抽出部215、モデル選択部217、話者識別部218及び識別結果出力部219は、プロセッサにより実現される。プロセッサは、例えば、CPU（中央演算処理装置）などから構成される。

【0050】

性別識別モデル記憶部203、性別識別用音声データ記憶部204、登録音声データ記憶部212及び話者識別モデル記憶部216は、メモリにより実現される。メモリは、例えば、ROM（Read Only Memory）又はEEPROM（Electrically Erasable Programmable Read Only Memory）などから構成される。

10

【0051】

なお、話者識別装置2は、例えば、コンピュータ、スマートフォン、タブレット型コンピュータ又はサーバであってもよい。

【0052】

登録対象音声データ取得部201は、マイクロホン1から出力された登録対象音声データを取得する。

【0053】

特徴量抽出部202は、登録対象音声データ取得部201によって取得された登録対象音声データの特徴量を抽出する。特徴量は、例えばi-vectorである。i-vectorは、GMM（Gaussian Mixture Model）スーパーベクトルに因子分析を用いることで音声データから抽出される低次元ベクトルの特徴量である。なお、i-vectorの抽出方法については従来技術であるため、詳細な説明は省略する。また、特徴量は、i-vectorに限定されず、例えばx-vectorなどの他の特徴量であってもよい。

20

【0054】

性別識別モデル記憶部203は、話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを予め記憶する。なお、性別識別モデルの生成方法については後述する。

【0055】

性別識別用音声データ記憶部204は、登録対象音声データの話者の性別を識別するために用いる性別識別用音声データの特徴量を予め記憶する。性別識別用音声データは、男性の複数の音声データと、女性の複数の音声データとを含む。なお、性別識別用音声データ記憶部204は、性別識別用音声データの特徴量を予め記憶しているが、本開示は特にこれに限定されず、性別識別用音声データを予め記憶してもよい。この場合、話者識別装置2は、性別識別用音声データの特徴量を抽出する特徴量抽出部を備える。

30

【0056】

性別識別部205は、特徴量抽出部202によって抽出された登録対象音声データの特徴量を用いて登録対象音声データの話者の性別を識別する。性別識別部205は、話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを性別識別モデル記憶部203から取得する。性別識別部205は、登録対象音声データの特徴量を性別識別モデルに入力することにより、登録対象音声データの話者の性別を識別する。

40

【0057】

性別識別部205は、登録対象音声データの特徴量と、性別識別用音声データ記憶部204に予め記憶されている男性の複数の音声データの特徴量それぞれとを性別識別モデルに入力することにより、登録対象音声データと男性の複数の音声データそれぞれの類似度を性別識別モデルから取得する。そして、性別識別部205は、取得した複数の類似度の平均を平均男性類似度として算出する。

【0058】

また、性別識別部205は、登録対象音声データの特徴量と、性別識別用音声データ記

50

憶部 204 に予め記憶されている女性の複数の音声データの特徴量それぞれとを性別識別モデルに入力することにより、登録対象音声データと女性の複数の音声データそれぞれとの類似度を性別識別モデルから取得する。そして、性別識別部 205 は、取得した複数の類似度の平均を平均女性類似度として算出する。

【0059】

性別識別部 205 は、平均男性類似度が平均女性類似度より高い場合、登録対象音声データの話者の性別を男性と識別する。一方、性別識別部 205 は、平均男性類似度が平均女性類似度より低い場合、登録対象音声データの話者の性別を女性と識別する。なお、性別識別部 205 は、平均男性類似度が平均女性類似度と同じである場合、登録対象音声データの話者の性別を男性と識別してもよいし、登録対象音声データの話者の性別を女性と識別してもよい。

【0060】

登録部 206 は、性別識別部 205 によって識別された性別情報を対応付けた登録対象音声データを登録音声データとして登録する。登録部 206 は、登録音声データを登録音声データ記憶部 212 に登録する。

【0061】

なお、話者識別装置 2 は、登録対象音声データの話者に関する情報の入力を受け付ける入力受付部をさらに備えてもよい。そして、登録部 206 は、話者に関する情報に対応付けて登録音声データを登録音声データ記憶部 212 に登録してもよい。話者に関する情報は、例えば、話者の名前などである。

【0062】

識別対象音声データ取得部 211 は、マイクロホン 1 から出力された識別対象音声データを取得する。

【0063】

登録音声データ記憶部 212 は、性別情報が対応付けられた登録音声データを記憶する。登録音声データ記憶部 212 は、複数の登録音声データを記憶する。

【0064】

登録音声データ取得部 213 は、登録音声データ記憶部 212 に登録されている登録音声データを取得する。

【0065】

特徴量抽出部 214 は、識別対象音声データ取得部 211 によって取得された識別対象音声データの特徴量を抽出する。特徴量は、例えば *i - v e c t o r* である。

【0066】

特徴量抽出部 215 は、登録音声データ取得部 213 によって取得された登録音声データの特徴量を抽出する。特徴量は、例えば *i - v e c t o r* である。

【0067】

話者識別モデル記憶部 216 は、男性の話者を識別するために男性の音声データを用いて機械学習された第 1 話者識別モデルと、女性の話者を識別するために女性の音声データを用いて機械学習された第 2 話者識別モデルとを予め記憶する。話者識別モデル記憶部 216 は、後述する話者識別モデル生成装置 4 によって生成された第 1 話者識別モデル及び第 2 話者識別モデルを予め記憶する。なお、第 1 話者識別モデル及び第 2 話者識別モデルの生成方法については後述する。

【0068】

モデル選択部 217 は、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、男性の話者を識別するために男性の音声データを用いて機械学習された第 1 話者識別モデルを選択する。また、モデル選択部 217 は、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、女性の話者を識別するために女性の音声データを用いて機械学習された第 2 話者識別モデルを選択する。

【0069】

本実施の形態 1 において、モデル選択部 2 1 7 は、登録音声データの話者の性別が男性である場合、第 1 話者識別モデルを選択し、登録音声データの話者の性別が女性である場合、第 2 話者識別モデルを選択する。登録音声データには、予め性別が対応付けられている。そのため、モデル選択部 2 1 7 は、登録音声データに対応付けられている性別が男性である場合、第 1 話者識別モデルを選択し、登録音声データに対応付けられている性別が女性である場合、第 2 話者識別モデルを選択する。

【0070】

話者識別部 2 1 8 は、識別対象音声データの特徴量と登録音声データの特徴量とを、モデル選択部 2 1 7 によって選択された第 1 話者識別モデル及び第 2 話者識別モデルのいずれかに入力することにより、識別対象音声データの話者を識別する。

【0071】

話者識別部 2 1 8 は、類似度算出部 2 3 1 及び類似度判定部 2 3 2 を備える。

【0072】

類似度算出部 2 3 1 は、識別対象音声データの特徴量と複数の登録音声データの特徴量それぞれとを、選択された第 1 話者識別モデル及び第 2 話者識別モデルのいずれかに入力することにより、識別対象音声データと複数の登録音声データそれぞれの類似度を第 1 話者識別モデル及び第 2 話者識別モデルのいずれかから取得する。

【0073】

類似度判定部 2 3 2 は、取得した類似度が最も高い登録音声データの話者を識別対象音声データの話者として識別する。

【0074】

なお、類似度判定部 2 3 2 は、最も高い類似度が閾値より大きいかな否かを判定してもよい。識別対象音声データの話者と同じ話者の登録音声データが登録音声データ記憶部 2 1 2 に存在しない場合であっても、識別対象音声データと各登録音声データとの類似度は算出される。そのため、類似度が最も高い登録音声データであっても、当該登録音声データの話者が、識別対象音声データの話者と同じであるとは限らない。そこで、最も高い類似度が閾値より大きいかな否かが判定されることにより、確実に話者を識別することができる。

【0075】

識別結果出力部 2 1 9 は、話者識別部 2 1 8 による識別結果を出力する。識別結果出力部 2 1 9 は、例えばディスプレイ又はスピーカであり、識別対象音声データの話者が識別された場合、識別対象音声データの話者が予め登録されている話者であることを示すメッセージをディスプレイ又はスピーカに出力する。一方、識別結果出力部 2 1 9 は、識別対象音声データの話者が識別されなかった場合、識別対象音声データの話者が予め登録されている話者ではないことを示すメッセージをディスプレイ又はスピーカに出力する。識別結果出力部 2 1 9 は、話者識別部 2 1 8 による識別結果を、話者識別装置 2 以外の他の装置へ出力してもよい。

【0076】

続いて、本開示の実施の形態 1 における性別識別モデル生成装置について説明する。

【0077】

図 2 は、本開示の実施の形態 1 における性別識別モデル生成装置の構成を示す図である。

【0078】

図 2 に示す性別識別モデル生成装置 3 は、性別識別用音声データ記憶部 3 0 1、性別識別用音声データ取得部 3 0 2、特徴量抽出部 3 0 3、性別識別モデル生成部 3 0 4 及び性別識別モデル記憶部 3 0 5 を備える。

【0079】

性別識別用音声データ取得部 3 0 2、特徴量抽出部 3 0 3 及び性別識別モデル生成部 3 0 4 は、プロセッサにより実現される。性別識別用音声データ記憶部 3 0 1 及び性別識別モデル記憶部 3 0 5 は、メモリにより実現される。

【0080】

性別識別用音声データ記憶部 3 0 1 は、男性及び女性のいずれかを示す性別ラベルが付

10

20

30

40

50

与された複数の音声データを予め記憶する。性別識別用音声データ記憶部 301 は、複数の話者毎に互いに異なる複数の音声データを記憶する。

【0081】

性別識別用音声データ取得部 302 は、男性及び女性のいずれかを示す性別ラベルが付与された複数の音声データを性別識別用音声データ記憶部 301 から取得する。なお、本実施の形態 1 では、性別識別用音声データ取得部 302 は、複数の音声データを性別識別用音声データ記憶部 301 から取得しているが、本開示は特にこれに限定されず、ネットワークを介して外部機器から複数の音声データを取得（受信）してもよい。

【0082】

特徴量抽出部 303 は、性別識別用音声データ取得部 302 によって取得された複数の音声データの特徴量を抽出する。特徴量は、例えば *i - v e c t o r* である。

【0083】

性別識別モデル生成部 304 は、複数の音声データのうちの第 1 音声データ及び第 2 音声データの各特徴量と、第 1 音声データ及び第 2 音声データの各性別ラベルの類似度とを教師データとして用い、入力を 2 つの音声データの各特徴量とし、出力を 2 つの音声データの類似度とする性別識別モデルを機械学習により生成する。例えば、性別識別モデルは、第 1 音声データの性別ラベルと第 2 音声データの性別ラベルとが同じであれば、最も高い類似度が出力され、第 1 音声データの性別ラベルと第 2 音声データの性別ラベルとが異なっていれば、最も低い類似度が出力されるように機械学習される。

【0084】

性別識別モデルとしては、確率的線形判別分析（*P r o b a b i l i s t i c L i n e a r D i s c r i m i n a n t A n a l y s i s : P L D A*）によるモデルが用いられる。*P L D A* モデルは、400 次元の *i - v e c t o r* 特徴量から話者の識別に有効な特徴量を自動的に選択し、対数尤度比を類似度として算出する。

【0085】

なお、機械学習としては、例えば、入力情報に対してラベル（出力情報）が付与された教師データを用いて入力と出力との関係を学習する教師あり学習、ラベルのない入力のみからデータの構造を構築する教師なし学習、ラベルありとラベルなしとのどちらも扱う半教師あり学習、報酬を最大化する行動を試行錯誤により学習する強化学習なども挙げられる。また、機械学習の具体的な手法としては、ニューラルネットワーク（多層のニューラルネットワークを用いた深層学習を含む）、遺伝的プログラミング、決定木、ベイジアン・ネットワーク、又はサポート・ベクター・マシン（*S V M*）などが存在する。性別識別モデルの機械学習においては、以上で挙げた具体例のいずれかをを用いればよい。

【0086】

性別識別モデル記憶部 305 は、性別識別モデル生成部 304 によって生成された性別識別モデルを記憶する。

【0087】

なお、性別識別モデル生成装置 3 は、性別識別モデル記憶部 305 に記憶された性別識別モデルを話者識別装置 2 へ送信してもよい。話者識別装置 2 は、受信した性別識別モデルを性別識別モデル記憶部 203 に記憶してもよい。また、話者識別装置 2 の製造時において、性別識別モデル生成装置 3 によって生成された性別識別モデルが話者識別装置 2 に記憶されてもよい。

【0088】

また、本実施の形態 1 の性別識別モデル生成装置 3 においては、複数の音声データには、男性及び女性のいずれかを示す性別ラベルが付与されているが、本開示は特にこれに限定されず、話者を識別するための識別情報がラベルとして付与されてもよい。この場合、性別識別モデル生成部 304 は、複数の音声データのうちの第 1 音声データ及び第 2 音声データの各特徴量と、第 1 音声データ及び第 2 音声データの各識別情報の類似度とを教師データとして用い、入力を 2 つの音声データの各特徴量とし、出力を 2 つの音声データの類似度とする性別識別モデルを機械学習により生成する。例えば、性別識別モデルは、第

10

20

30

40

50

1 音声データの識別情報と第 2 音声データの識別情報とが同じであれば、最も高い類似度が出力され、第 1 音声データの識別情報と第 2 音声データの識別情報とが異なっていれば、最も低い類似度が出力されるように機械学習される。

【0089】

続いて、本開示の実施の形態 1 における話者識別モデル生成装置について説明する。

【0090】

図 3 は、本開示の実施の形態 1 における話者識別モデル生成装置の構成を示す図である。

【0091】

図 3 に示す話者識別モデル生成装置 4 は、男性音声データ記憶部 401、男性音声データ取得部 402、特徴量抽出部 403、第 1 話者識別モデル生成部 404、第 1 話者識別モデル記憶部 405、女性音声データ記憶部 411、女性音声データ取得部 412、特徴量抽出部 413、第 2 話者識別モデル生成部 414 及び第 2 話者識別モデル記憶部 415 を備える。

【0092】

男性音声データ取得部 402、特徴量抽出部 403、第 1 話者識別モデル生成部 404、女性音声データ取得部 412、特徴量抽出部 413 及び第 2 話者識別モデル生成部 414 は、プロセッサにより実現される。男性音声データ記憶部 401、第 1 話者識別モデル記憶部 405、女性音声データ記憶部 411 及び第 2 話者識別モデル記憶部 415 は、メモリにより実現される。

【0093】

男性音声データ記憶部 401 は、男性である話者を識別するための話者識別ラベルが付与された複数の男性音声データを記憶する。男性音声データ記憶部 401 は、複数の話者毎に互いに異なる複数の男性音声データを記憶する。

【0094】

男性音声データ取得部 402 は、男性である話者を識別するための話者識別ラベルが付与された複数の男性音声データを男性音声データ記憶部 401 から取得する。なお、本実施の形態 1 では、男性音声データ取得部 402 は、複数の男性音声データを男性音声データ記憶部 401 から取得しているが、本開示は特にこれに限定されず、ネットワークを介して外部機器から複数の男性音声データを取得（受信）してもよい。

【0095】

特徴量抽出部 403 は、男性音声データ取得部 402 によって取得された複数の男性音声データの特徴量を抽出する。特徴量は、例えば *i - v e c t o r* である。

【0096】

第 1 話者識別モデル生成部 404 は、複数の男性音声データのうちの第 1 男性音声データ及び第 2 男性音声データの各特徴量と、第 1 男性音声データ及び第 2 男性音声データの各話者識別ラベルの類似度とを教師データとして用い、入力を 2 つの音声データの各特徴量とし、出力を 2 つの音声データの類似度とする第 1 話者識別モデルを機械学習により生成する。例えば、第 1 話者識別モデルは、第 1 男性音声データの話者識別ラベルと第 2 男性音声データの話者識別ラベルとが同じであれば、最も高い類似度が出力され、第 1 男性音声データの話者識別ラベルと第 2 男性音声データの話者識別ラベルとが異なっていれば、最も低い類似度が出力されるように機械学習される。

【0097】

第 1 話者識別モデルとしては、PLDA によるモデルが用いられる。PLDA モデルは、400 次元の *i - v e c t o r*（特徴量）から話者の識別に有効な特徴量を自動的に選択し、対数尤度比を類似度として算出する。

【0098】

なお、機械学習としては、例えば、入力情報に対してラベル（出力情報）が付与された教師データを用いて入力と出力との関係を学習する教師あり学習、ラベルのない入力のみからデータの構造を構築する教師なし学習、ラベルありとラベルなしとのどちらも扱う半教師あり学習、報酬を最大化する行動を試行錯誤により学習する強化学習なども挙げられ

10

20

30

40

50

る。また、機械学習の具体的な手法としては、ニューラルネットワーク（多層のニューラルネットワークを用いた深層学習を含む）、遺伝的プログラミング、決定木、ベイジアン・ネットワーク、又はサポート・ベクター・マシン（SVM）などが存在する。第1話者識別モデルの機械学習においては、以上で挙げた具体例のいずれかを用いればよい。

【0099】

第1話者識別モデル記憶部405は、第1話者識別モデル生成部404によって生成された第1話者識別モデルを記憶する。

【0100】

女性音声データ記憶部411は、女性である話者を識別するための話者識別ラベルが付与された複数の女性音声データを記憶する。女性音声データ記憶部411は、複数の話者毎に互いに異なる複数の女性音声データを記憶する。

【0101】

女性音声データ取得部412は、女性である話者を識別するための話者識別ラベルが付与された複数の女性音声データを女性音声データ記憶部411から取得する。なお、本実施の形態1では、女性音声データ取得部412は、複数の女性音声データを女性音声データ記憶部411から取得しているが、本開示は特にこれに限定されず、ネットワークを介して外部機器から複数の女性音声データを取得（受信）してもよい。

【0102】

特徴量抽出部413は、女性音声データ取得部412によって取得された複数の女性音声データの特徴量を抽出する。特徴量は、例えば *i - v e c t o r* である。

【0103】

第2話者識別モデル生成部414は、複数の女性音声データのうちの第1女性音声データ及び第2女性音声データの各特徴量と、第1女性音声データ及び第2女性音声データの各話者識別ラベルの類似度とを教師データとして用い、入力を2つの音声データの各特徴量とし、出力を2つの音声データの類似度とする第2話者識別モデルを機械学習により生成する。例えば、第2話者識別モデルは、第1女性音声データの話者識別ラベルと第2女性音声データの話者識別ラベルとが同じであれば、最も高い類似度が出力され、第1女性音声データの話者識別ラベルと第2女性音声データの話者識別ラベルとが異なっていれば、最も低い類似度が出力されるように機械学習される。

【0104】

第2話者識別モデルとしては、PLDAによるモデルが用いられる。PLDAモデルは、400次元の *i - v e c t o r*（特徴量）から話者の識別に有効な特徴量を自動的に選択し、対数尤度比を類似度として算出する。

【0105】

なお、機械学習としては、例えば、入力情報に対してラベル（出力情報）が付与された教師データを用いて入力と出力との関係を学習する教師あり学習、ラベルのない入力のみからデータの構造を構築する教師なし学習、ラベルありとラベルなしとのどちらも扱う半教師あり学習、報酬を最大化する行動を試行錯誤により学習する強化学習なども挙げられる。また、機械学習の具体的な手法としては、ニューラルネットワーク（多層のニューラルネットワークを用いた深層学習を含む）、遺伝的プログラミング、決定木、ベイジアン・ネットワーク、又はサポート・ベクター・マシン（SVM）などが存在する。第2話者識別モデルの機械学習においては、以上で挙げた具体例のいずれかを用いればよい。

【0106】

第2話者識別モデル記憶部415は、第2話者識別モデル生成部414によって生成された第2話者識別モデルを記憶する。

【0107】

なお、話者識別モデル生成装置4は、第1話者識別モデル記憶部405に記憶された第1話者識別モデル及び第2話者識別モデル記憶部415に記憶された第2話者識別モデルを話者識別装置2へ送信してもよい。話者識別装置2は、受信した第1話者識別モデル及び第2話者識別モデルを話者識別モデル記憶部216に記憶してもよい。また、話者識別

10

20

30

40

50

装置 2 の製造時において、話者識別モデル生成装置 4 によって生成された第 1 話者識別モデル及び第 2 話者識別モデルが話者識別装置 2 に記憶されてもよい。

【0108】

続いて、本実施の形態 1 における話者識別装置 2 の登録処理及び話者識別処理のそれぞれの動作について説明する。

【0109】

図 4 は、本実施の形態 1 における話者識別装置の登録処理の動作について説明するためのフローチャートである。

【0110】

まず、ステップ S 1 において、登録対象音声データ取得部 201 は、マイクロホン 1 から出力された登録対象音声データを取得する。自身が発話した音声データの登録を希望する話者は、マイクロホン 1 に向かって所定の文章を発話する。このとき、登録対象音声データの文章は、識別対象音声データの文章よりも長いことが好ましい。文字数が比較的多い登録対象音声データが取得されることにより、話者識別の精度を向上させることができる。また、話者識別装置 2 は、予め決められている複数の文章を登録対象話者に提示してもよい。この場合、登録対象話者は、提示された複数の文章を発話する。

【0111】

次に、ステップ S 2 において、特徴量抽出部 202 は、登録対象音声データ取得部 201 によって取得された登録対象音声データの特徴量を抽出する。

【0112】

次に、ステップ S 3 において、性別識別部 205 は、特徴量抽出部 202 によって抽出された登録対象音声データの特徴量を用いて登録対象音声データの話者の性別を識別する性別識別処理を行う。なお、性別識別処理については後述する。

【0113】

次に、ステップ S 4 において、登録部 206 は、性別識別部 205 によって識別された性別情報を対応付けた登録対象音声データを登録音声データとして登録音声データ記憶部 212 に記憶する。

【0114】

図 5 は、図 4 のステップ S 3 の性別識別処理の動作について説明するためのフローチャートである。

【0115】

まず、ステップ S 11 において、性別識別部 205 は、性別識別モデルを性別識別モデル記憶部 203 から取得する。性別識別モデルは、複数の音声データのうちの第 1 音声データ及び第 2 音声データの各特徴量と、第 1 音声データ及び第 2 音声データの各性別ラベルの類似度とを教師データとして用いて機械学習された性別識別モデルである。なお、性別識別モデルは、複数の音声データのうちの第 1 音声データ及び第 2 音声データの各特徴量と、第 1 音声データ及び第 2 音声データの各識別情報の類似度とを教師データとして用いて機械学習された性別識別モデルであってもよい。

【0116】

次に、ステップ S 12 において、性別識別部 205 は、男性音声データの特徴量を性別識別用音声データ記憶部 204 から取得する。

【0117】

次に、ステップ S 13 において、性別識別部 205 は、登録対象音声データの特徴量と、男性音声データの特徴量とを性別識別モデルに入力することにより、登録対象音声データと男性音声データとの類似度を算出する。

【0118】

次に、ステップ S 14 において、性別識別部 205 は、登録対象音声データと全ての男性音声データとの類似度が算出されたか否かを判定する。ここで、登録対象音声データと全ての男性音声データとの類似度が算出されていないと判定された場合（ステップ S 14 で NO）、ステップ S 12 に処理が戻る。そして、性別識別部 205 は、複数の男性音声

10

20

30

40

50

データの特徴量の中から、類似度を算出していない男性音声データの特徴量を性別識別用音声データ記憶部 204 から取得する。

【0119】

一方、登録対象音声データと全ての男性音声データとの類似度が算出されたと判定された場合（ステップ S14 で YES）、ステップ S15 において、性別識別部 205 は、算出した複数の類似度の平均を平均男性類似度として算出する。

【0120】

次に、ステップ S16 において、性別識別部 205 は、女性音声データの特徴量を性別識別用音声データ記憶部 204 から取得する。

【0121】

次に、ステップ S17 において、性別識別部 205 は、登録対象音声データの特徴量と、女性音声データの特徴量とを性別識別モデルに入力することにより、登録対象音声データと女性音声データとの類似度を算出する。

【0122】

次に、ステップ S18 において、性別識別部 205 は、登録対象音声データと全ての女性音声データとの類似度が算出されたか否かを判定する。ここで、登録対象音声データと全ての女性音声データとの類似度が算出されていないと判定された場合（ステップ S18 で NO）、ステップ S16 に処理に戻る。そして、性別識別部 205 は、複数の女性音声データの特徴量の中から、類似度を算出していない女性音声データの特徴量を性別識別用音声データ記憶部 204 から取得する。

【0123】

一方、登録対象音声データと全ての女性音声データとの類似度が算出されたと判定された場合（ステップ S18 で YES）、ステップ S19 において、性別識別部 205 は、算出した複数の類似度の平均を平均女性類似度として算出する。

【0124】

次に、ステップ S20 において、性別識別部 205 は、平均男性類似度と平均女性類似度とのうちの高い方の性別を識別結果として登録部 206 へ出力する。

【0125】

図 6 は、本実施の形態 1 における話者識別装置の話者識別処理の動作について説明するための第 1 のフローチャートであり、図 7 は、本実施の形態 1 における話者識別装置の話者識別処理の動作について説明するための第 2 のフローチャートである。

【0126】

まず、ステップ S31 において、識別対象音声データ取得部 211 は、マイクロホン 1 から出力された識別対象音声データを取得する。識別対象話者は、マイクロホン 1 に向かって発話する。マイクロホン 1 は、識別対象話者によって発話された音声を集集し、識別対象音声データを出力する。

【0127】

次に、ステップ S32 において、特徴量抽出部 214 は、識別対象音声データ取得部 211 によって取得された識別対象音声データの特徴量を抽出する。

【0128】

次に、ステップ S33 において、登録音声データ取得部 213 は、登録音声データを登録音声データ記憶部 212 から取得する。このとき、登録音声データ取得部 213 は、登録音声データ記憶部 212 に登録されている複数の登録音声データの中から、1 の登録音声データを取得する。

【0129】

次に、ステップ S34 において、特徴量抽出部 215 は、登録音声データ取得部 213 によって取得された登録音声データの特徴量を抽出する。

【0130】

次に、ステップ S35 において、モデル選択部 217 は、登録音声データ取得部 213 によって取得された登録音声データに対応付けられている性別を取得する。

10

20

30

40

50

## 【0131】

次に、ステップS36において、モデル選択部217は、取得した性別が男性であるか否かを判定する。ここで、取得した性別が男性であると判定された場合（ステップS36でYES）、ステップS37において、モデル選択部217は、第1話者識別モデルを選択する。モデル選択部217は、選択した第1話者識別モデルを話者識別モデル記憶部216から取得し、取得した第1話者識別モデルを類似度算出部231へ出力する。

## 【0132】

一方、取得した性別が男性ではないと判定された場合、すなわち取得した性別が女性であると判定された場合（ステップS36でNO）、ステップS38において、モデル選択部217は、第2話者識別モデルを選択する。モデル選択部217は、選択した第2話者識別モデルを話者識別モデル記憶部216から取得し、取得した第2話者識別モデルを類似度算出部231へ出力する。

10

## 【0133】

次に、ステップS39において、類似度算出部231は、識別対象音声データの特徴量と登録音声データの特徴量とを、選択された第1話者識別モデル及び第2話者識別モデルのいずれかに入力することにより、識別対象音声データと登録音声データとの類似度を算出する。

## 【0134】

次に、ステップS40において、類似度算出部231は、識別対象音声データと、登録音声データ記憶部212に記憶されている全ての登録音声データとの類似度が算出されたか否かを判定する。ここで、識別対象音声データと全ての登録音声データとの類似度が算出されていないと判定された場合（ステップS40でNO）、ステップS33に処理に戻る。そして、登録音声データ取得部213は、登録音声データ記憶部212に記憶されている複数の登録音声データの中から、類似度が算出されていない登録音声データを取得する。

20

## 【0135】

一方、識別対象音声データと全ての登録音声データとの類似度が算出されたと判定された場合（ステップS40でYES）、ステップS41において、類似度判定部232は、最も高い類似度が閾値より大きいと判定する。

## 【0136】

ここで、最も高い類似度が閾値より大きいと判定された場合（ステップS41でYES）、ステップS42において、類似度判定部232は、類似度が最も高い登録音声データの話者を識別対象音声データの話者として識別する。

30

## 【0137】

一方、最も高い類似度が閾値以下であると判定された場合（ステップS41でNO）、ステップS43において、類似度判定部232は、複数の登録音声データの中に、識別対象音声データと話者が同一である登録音声データが存在しないと判定する。

## 【0138】

次に、ステップS44において、識別結果出力部219は、話者識別部218による識別結果を出力する。識別結果出力部219は、識別対象音声データの話者が識別された場合、識別対象音声データの話者が予め登録されている話者であることを示すメッセージを出力する。一方、識別結果出力部219は、識別対象音声データの話者が識別されなかった場合、識別対象音声データの話者が予め登録されている話者ではないことを示すメッセージを出力する。

40

## 【0139】

このように、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、男性用に生成された第1話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、女性用

50

に生成された第2話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

【0140】

したがって、音声データの特徴量の分布が性別によって異なる場合であっても、それぞれの性別に特化した第1話者識別モデル及び第2話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

【0141】

なお、本実施の形態1において、モデル選択部217は、登録音声データの話者の性別に基づいて、第1話者識別モデル及び第2話者識別モデルのいずれかを選択しているが、本開示は特にこれに限定されない。モデル選択部217は、識別対象音声データの話者の性別に基づいて、第1話者識別モデル及び第2話者識別モデルのいずれかを選択してもよい。この場合、話者識別装置2は、識別対象音声データの話者の性別を識別する性別識別部と、話者の性別を識別するために男性及び女性の音声データを用いて機械学習された性別識別モデルを予め記憶する性別識別モデル記憶部と、識別対象音声データの話者の性別を識別するために用いる性別識別用音声データの特徴量を予め記憶する性別識別用音声データ記憶部とを備える。性別識別部、性別識別モデル記憶部及び性別識別用音声データ記憶部は、上記の性別識別部205、性別識別モデル記憶部203及び性別識別用音声データ記憶部204と同じ構成である。また、識別対象音声データの話者の性別が識別される場合、性別識別部205、性別識別モデル記憶部203及び性別識別用音声データ記憶部204は不要となる。

【0142】

続いて、本実施の形態1における性別識別モデル生成装置3の性別識別モデル生成処理の動作について説明する。

【0143】

図8は、本実施の形態1における性別識別モデル生成装置の性別識別モデル生成処理の動作について説明するためのフローチャートである。

【0144】

まず、ステップS51において、性別識別用音声データ取得部302は、男性及び女性のいずれかを示す性別ラベルが付与された複数の音声データを性別識別用音声データ記憶部301から取得する。

【0145】

次に、ステップS52において、特徴量抽出部303は、性別識別用音声データ取得部302によって取得された複数の音声データの特徴量を抽出する。

【0146】

次に、ステップS53において、性別識別モデル生成部304は、複数の音声データのうちの第1音声データ及び第2音声データの各特徴量と、第1音声データ及び第2音声データの各性別ラベルの類似度とを教師データとして取得する。

【0147】

次に、ステップS54において、性別識別モデル生成部304は、取得した教師データを用い、入力を2つの音声データの特徴量とし、出力を2つの音声データの類似度とする性別識別モデルを機械学習する。

【0148】

次に、ステップS55において、性別識別モデル生成部304は、複数の音声データのうちの全ての音声データの組み合わせを用いて性別識別モデルを機械学習させたか否かを判定する。ここで、全ての音声データの組み合わせを用いて性別識別モデルを機械学習させていないと判定された場合（ステップS55でNO）、ステップS53に処理に戻る。そして、性別識別モデル生成部304は、複数の音声データのうちの機械学習に用いていない組み合わせの第1音声データ及び第2音声データの各特徴量と、第1音声データ及び第2音声データの各性別ラベルの類似度とを教師データとして取得する。

## 【0149】

一方、全ての音声データの組み合わせを用いて性別識別モデルを機械学習させたと判定された場合（ステップS55でYES）、ステップS56において、性別識別モデル生成部304は、機械学習により生成した性別識別モデルを性別識別モデル記憶部305に記憶する。

## 【0150】

このように、機械学習により生成された性別識別モデルに、登録音声データ又は識別対象音声データの特徴量と、男性の音声データの特徴量とが入力されることにより、2つの音声データの第1類似度が出力される。また、性別識別モデルに、登録音声データ又は識別対象音声データの特徴量と、女性の音声データの特徴量とが入力されることにより、2つの音声データの第2類似度が出力される。そして、第1類似度と第2類似度とが比較されることにより、登録音声データ又は識別対象音声データの話者の性別を容易に推定することができる。

10

## 【0151】

続いて、本実施の形態1における話者識別モデル生成装置4の話者識別モデル生成処理の動作について説明する。

## 【0152】

図9は、本実施の形態1における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第1のフローチャートであり、図10は、本実施の形態1における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第2のフローチャートである。

20

## 【0153】

まず、ステップS61において、男性音声データ取得部402は、男性である話者を識別するための話者識別ラベルが付与された複数の男性音声データを男性音声データ記憶部401から取得する。

## 【0154】

次に、ステップS62において、特徴量抽出部403は、男性音声データ取得部402によって取得された複数の男性音声データの特徴量を抽出する。

## 【0155】

次に、ステップS63において、第1話者識別モデル生成部404は、複数の男性音声データのうちの第1男性音声データ及び第2男性音声データの各特徴量と、第1男性音声データ及び第2男性音声データの各話者識別ラベルの類似度とを教師データとして取得する。

30

## 【0156】

次に、ステップS64において、第1話者識別モデル生成部404は、取得した教師データを用い、入力を2つの音声データの各特徴量とし、出力を2つの音声データの類似度とする第1話者識別モデルを機械学習する。

## 【0157】

次に、ステップS65において、第1話者識別モデル生成部404は、複数の男性音声データのうちの全ての男性音声データの組み合わせを用いて第1話者識別モデルを機械学習させたか否かを判定する。ここで、全ての男性音声データの組み合わせを用いて第1話者識別モデルを機械学習させていないと判定された場合（ステップS65でNO）、ステップS63に処理が戻る。そして、第1話者識別モデル生成部404は、複数の男性音声データのうちの機械学習に用いていない組み合わせの第1男性音声データ及び第2男性音声データの各特徴量と、第1男性音声データ及び第2男性音声データの各話者識別ラベルの類似度とを教師データとして取得する。

40

## 【0158】

一方、全ての男性音声データの組み合わせを用いて第1話者識別モデルを機械学習させたと判定された場合（ステップS65でYES）、ステップS66において、第1話者識別モデル生成部404は、機械学習により生成された第1話者識別モデルを第1話者識別

50

モデル記憶部 405 に記憶する。

【0159】

次に、ステップ S67 において、女性音声データ取得部 412 は、女性である話者を識別するための話者識別ラベルが付与された複数の女性音声データを女性音声データ記憶部 411 から取得する。

【0160】

次に、ステップ S68 において、特徴量抽出部 413 は、女性音声データ取得部 412 によって取得された複数の女性音声データの特徴量を抽出する。

【0161】

次に、ステップ S69 において、第 2 話者識別モデル生成部 414 は、複数の女性音声データのうちの第 1 女性音声データ及び第 2 女性音声データの各特徴量と、第 1 女性音声データ及び第 2 女性音声データの各話者識別ラベルの類似度とを教師データとして取得する。

10

【0162】

次に、ステップ S70 において、第 2 話者識別モデル生成部 414 は、取得した教師データを用い、入力を 2 つの音声データの各特徴量とし、出力を 2 つの音声データの類似度とする第 2 話者識別モデルを機械学習する。

【0163】

次に、ステップ S71 において、第 2 話者識別モデル生成部 414 は、複数の女性音声データのうちの全ての女性音声データの組み合わせを用いて第 2 話者識別モデルを機械学習させたか否かを判定する。ここで、全ての女性音声データの組み合わせを用いて第 2 話者識別モデルを機械学習させていないと判定された場合（ステップ S71 で NO）、ステップ S69 に処理が戻る。そして、第 2 話者識別モデル生成部 414 は、複数の女性音声データのうちの機械学習に用いていない組み合わせの第 1 女性音声データ及び第 2 女性音声データの各特徴量と、第 1 女性音声データ及び第 2 女性音声データの各話者識別ラベルの類似度とを教師データとして取得する。

20

【0164】

一方、全ての女性音声データの組み合わせを用いて第 2 話者識別モデルを機械学習させたと判定された場合（ステップ S71 で YES）、ステップ S72 において、第 2 話者識別モデル生成部 414 は、機械学習により生成された第 2 話者識別モデルを第 2 話者識別モデル記憶部 415 に記憶する。

30

【0165】

このように、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が男性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、男性用に生成された第 1 話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者のいずれかの性別が女性である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、女性用に生成された第 2 話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

【0166】

40

したがって、音声データの特徴量の分布が性別によって異なる場合であっても、男性に特化した第 1 話者識別モデル及び女性に特化した第 2 話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができる。

【0167】

続いて、本実施の形態 1 における話者識別装置 2 の話者識別性能の評価について説明する。

【0168】

図 11 は、従来の話者識別装置の話者識別性能評価結果と、実施の形態 1 の話者識別装置の話者識別性能評価結果とを示す図である。

50

【0169】

図11に示す性能評価結果は、SRE19 progress dataset及びSRE19 Evaluation datasetにより提供される音声データの話者を従来の話者識別装置及び本実施の形態1の話者識別装置2により識別した結果を表している。

【0170】

SRE19は、アメリカ国立標準技術研究所（NIST）が主催する話者識別のコンペティションである。SRE19 progress dataset及びSRE19 Evaluation datasetは、いずれもSRE19によって提供されるデータセットである。

【0171】

従来の話者識別装置は、男性と女性とを区別することなく生成された話者識別モデルを用いて、音声データの話者を識別した。

【0172】

また、本実施の形態1の話者識別装置2は、男性用の第1話者識別モデルと女性用の第2話者識別モデルとを用いて、音声データの話者を識別した。

【0173】

評価結果は、話者識別の評価に一般的に用いられるEER（Equal Error Rate）（％）と、NISTによって定義されたコストであるminC（minimum Cost）と、NISTによって定義されたコストであるactC（actual Cost）とで表した。なお、EER、minC及びactCは、いずれも値が小さいほど性能が高いことを表している。

【0174】

図11に示すように、本実施の形態1の話者識別装置2のEER、minC及びactCは、いずれも、従来の話者識別装置のEER、minC及びactCよりも小さくなっている。これにより、本実施の形態1の話者識別装置2の話者識別性能は、従来の話者識別装置の話者識別性能よりも高いことが分かる。

【0175】

続いて、本実施の形態1の変形例1における話者識別装置2について説明する。

【0176】

本実施の形態1の変形例1における話者識別装置2は、性別識別部205の性別識別処理が上記と異なる。

【0177】

本実施の形態1の変形例1における性別識別部205は、登録対象音声データの特徴量と、性別識別用音声データ記憶部204に予め記憶されている男性の複数の音声データの特徴量それぞれとを性別識別モデルに入力することにより、登録対象音声データと男性の複数の音声データそれぞれの類似度を性別識別モデルから取得する。そして、性別識別部205は、取得した複数の類似度のうちの最大値を最大男性類似度として算出する。

【0178】

また、性別識別部205は、登録対象音声データの特徴量と、性別識別用音声データ記憶部204に予め記憶されている女性の複数の音声データの特徴量それぞれとを性別識別モデルに入力することにより、登録対象音声データと女性の複数の音声データそれぞれの類似度を性別識別モデルから取得する。そして、性別識別部205は、取得した複数の類似度のうちの最大値を最大女性類似度として算出する。

【0179】

性別識別部205は、最大男性類似度が最大女性類似度より高い場合、登録対象音声データの話者の性別を男性と識別する。一方、性別識別部205は、最大男性類似度が最大女性類似度より低い場合、登録対象音声データの話者の性別を女性と識別する。なお、性別識別部205は、最大男性類似度が最大女性類似度と同じである場合、登録対象音声データの話者の性別を男性と識別してもよいし、登録対象音声データの話者の性別を女性と識別してもよい。

10

20

30

40

50

## 【0180】

続いて、本実施の形態1の変形例1における性別識別処理の動作について説明する。

## 【0181】

図12は、本実施の形態1の変形例1における性別識別処理の動作について説明するためのフローチャートである。本実施の形態1の変形例1における性別識別処理は、図4のステップS3の性別識別処理の他の動作例である。

## 【0182】

ステップS81～ステップS84の処理は、図5に示すステップS11～ステップS14の処理と同じであるので、説明を省略する。

## 【0183】

登録対象音声データと全ての男性音声データとの類似度が算出されたと判定された場合（ステップS84でYES）、ステップS85において、性別識別部205は、算出した複数の類似度のうちの最大値を最大男性類似度として算出する。

## 【0184】

ステップS86～ステップS88の処理は、図5に示すステップS16～ステップS18の処理と同じであるので、説明を省略する。

## 【0185】

登録対象音声データと全ての女性音声データとの類似度が算出されたと判定された場合（ステップS88でYES）、ステップS89において、性別識別部205は、算出した複数の類似度のうちの最大値を最大女性類似度として算出する。

## 【0186】

次に、ステップS90において、性別識別部205は、最大男性類似度と最大女性類似度とのうちの高い方の性別を識別結果として登録部206へ出力する。

## 【0187】

続いて、本実施の形態1の変形例2における話者識別装置2について説明する。

## 【0188】

本実施の形態1の変形例2における話者識別装置2は、性別識別部205の性別識別処理が上記と異なる。

## 【0189】

本実施の形態1の変形例2における性別識別部205は、予め記憶されている男性の複数の音声データの平均特徴量を算出する。そして、性別識別部205は、登録対象音声データの特徴量と男性の複数の音声データの平均特徴量とを性別識別モデルに入力することにより、登録対象音声データと男性の音声データ群との第1類似度を性別識別モデルから取得する。

## 【0190】

また、性別識別部205は、予め記憶されている女性の複数の音声データの平均特徴量を算出する。そして、性別識別部205は、登録対象音声データの特徴量と女性の複数の音声データの平均特徴量とを性別識別モデルに入力することにより、登録対象音声データと女性の音声データ群との第2類似度を性別識別モデルから取得する。

## 【0191】

性別識別部205は、第1類似度が第2類似度より高い場合、登録対象音声データの話者の性別を男性と識別する。一方、性別識別部205は、第1類似度が第2類似度より低い場合、登録対象音声データの話者の性別を女性と識別する。なお、性別識別部205は、第1類似度が第2類似度と同じである場合、登録対象音声データの話者の性別を男性と識別してもよいし、登録対象音声データの話者の性別を女性と識別してもよい。

## 【0192】

続いて、本実施の形態1の変形例2における性別識別処理の動作について説明する。

## 【0193】

図13は、本実施の形態1の変形例2における性別識別処理の動作について説明するためのフローチャートである。本実施の形態1の変形例2における性別識別処理は、図4の

10

20

30

40

50

ステップ S 3 の性別識別処理の他の動作例である。

【0194】

まず、ステップ S 101 において、性別識別部 205 は、性別識別モデルを性別識別モデル記憶部 203 から取得する。性別識別モデルは、複数の音声データのうちの第 1 音声データ及び第 2 音声データの各特微量と、第 1 音声データ及び第 2 音声データの各性別ラベルの類似度とを教師データとして用いて機械学習された性別識別モデルである。なお、性別識別モデルは、複数の音声データのうちの第 1 音声データ及び第 2 音声データの各特微量と、第 1 音声データ及び第 2 音声データの各識別情報の類似度とを教師データとして用いて機械学習された性別識別モデルであってもよい。

【0195】

次に、ステップ S 102 において、性別識別部 205 は、複数の男性音声データの特微量を性別識別用音声データ記憶部 204 から取得する。

【0196】

次に、ステップ S 103 において、性別識別部 205 は、取得した複数の男性音声データの平均特微量を算出する。

【0197】

次に、ステップ S 104 において、性別識別部 205 は、登録対象音声データの特微量と、複数の男性音声データの平均特微量とを性別識別モデルに入力することにより、登録対象音声データと男性音声データ群との第 1 類似度を算出する。

【0198】

次に、ステップ S 105 において、性別識別部 205 は、複数の女性音声データの特微量を性別識別用音声データ記憶部 204 から取得する。

【0199】

次に、ステップ S 106 において、性別識別部 205 は、取得した複数の女性音声データの平均特微量を算出する。

【0200】

次に、ステップ S 107 において、性別識別部 205 は、登録対象音声データの特微量と、複数の女性音声データの平均特微量とを性別識別モデルに入力することにより、登録対象音声データと女性音声データ群との第 2 類似度を算出する。

【0201】

次に、ステップ S 108 において、性別識別部 205 は、第 1 類似度と第 2 類似度とのうちの高い方の性別を識別結果として登録部 206 へ出力する。

【0202】

なお、本実施の形態 1 における性別識別処理では、登録音声データと男性音声データとの類似度と、登録音声データと女性音声データとの類似度が算出され、算出された 2 つの類似度が比較されることにより、登録音声データの性別が判定されるが、本開示は特にこれに限定されない。例えば、性別識別モデル生成部 304 は、複数の音声データのうちの 1 の音声データの特微量と、1 の音声データの性別ラベルとを教師データとして用い、入力を音声データの特微量とし、出力を男性及び女性のいずれかとする性別識別モデルを機械学習により生成してもよい。この場合、性別識別モデルは、例えば、ディープニューラルネットワークモデルであり、機械学習は、例えば、ディープラーニングである。

【0203】

また、性別識別モデルは、入力された音声データに対して、男性である確率及び女性である確率を算出してもよい。この場合、性別識別部 205 は、男性である確率及び女性である確率のうちの高い方の確率の性別を識別結果として出力してもよい。

【0204】

また、話者識別装置 2 は、登録対象音声データが取得される際に、登録対象音声データの話者の性別の入力を受け付ける入力受付部を備えてもよい。この場合、登録部 206 は、登録対象音声データ取得部 201 によって取得された登録対象音声データを、入力受付部によって受け付けられた性別と対応付けて登録音声データ記憶部 212 に記憶してもよ

10

20

30

40

50

い。これにより、特徴量抽出部 202、性別識別モデル記憶部 203、性別識別用音声データ記憶部 204 及び性別識別部 205 が不要となり、話者識別装置 2 の構成を簡略化することができるとともに、話者識別装置 2 の登録処理にかかる負荷を低減することができる。

#### 【0205】

##### （実施の形態 2）

上記の実施の形態 1 では、登録音声データ記憶部 212 に記憶されている全ての登録音声データそれぞれと、識別対象音声データとの類似度が算出され、類似度が最も高い登録音声データの話者が識別対象音声データの話者として識別される。一方、実施の形態 2 では、識別対象音声データの話者の識別情報が入力され、登録音声データ記憶部 212 に記憶されている複数の登録音声データのうち、当該識別情報に予め対応付けられている 1 の登録音声データが取得される。そして、1 の登録音声データと識別対象音声データとの類似度が算出され、類似度が閾値より高い場合、登録音声データの話者が識別対象音声データの話者として識別される。

#### 【0206】

図 14 は、本開示の実施の形態 2 における話者識別システムの構成を示す図である。

#### 【0207】

図 14 に示す話者識別システムは、マイクロホン 1 及び話者識別装置 21 を備える。なお、話者識別装置 21 は、マイクロホン 1 を備えていなくてもよいし、備えてもよい。

#### 【0208】

なお、本実施の形態 2 において、実施の形態 1 と同じ構成については同じ符号を付し、説明を省略する。

#### 【0209】

話者識別装置 21 は、登録対象音声データ取得部 201、特徴量抽出部 202、性別識別モデル記憶部 203、性別識別用音声データ記憶部 204、性別識別部 205、登録部 2061、識別対象音声データ取得部 211、登録音声データ記憶部 2121、登録音声データ取得部 2131、特徴量抽出部 214、特徴量抽出部 215、話者識別モデル記憶部 216、モデル選択部 217、話者識別部 2181、識別結果出力部 219、入力受付部 221 及び識別情報取得部 222 を備える。

#### 【0210】

入力受付部 221 は、例えばキーボード、マウス及びタッチパネルなどの入力装置である。音声データの登録時において、入力受付部 221 は、音声データを登録する話者を識別するための識別情報の話者による入力を受け付ける。また、音声データの識別時において、入力受付部 221 は、識別する音声データの話者を識別するための識別情報の話者による入力を受け付ける。なお、入力受付部 221 は、カードリーダー又は R F I D（Radio Frequency Identification）リーダーなどであってもよい。この場合、話者は、識別情報が記録されたカードをカードリーダーに読み取らせる、又は識別情報が記録された R F I D タグを R F I D リーダーに読み取らせる。

#### 【0211】

識別情報取得部 222 は、入力受付部 221 によって受け付けられた識別情報を取得する。音声データの登録時において、識別情報取得部 222 は、登録対象音声データの話者を識別するための識別情報を取得し、取得した識別情報を登録部 2061 へ出力する。また、音声データの識別時において、識別情報取得部 222 は、識別対象音声データの話者を識別するための識別情報を取得し、取得した識別情報を登録音声データ取得部 2131 へ出力する。

#### 【0212】

登録部 2061 は、性別識別部 205 によって識別された性別情報及び識別情報取得部 222 によって取得された識別情報を対応付けた登録対象音声データを登録音声データとして登録する。登録部 2061 は、登録音声データを登録音声データ記憶部 2121 に登録する。

## 【0213】

登録音声データ記憶部2121は、性別情報及び識別情報が対応付けられた登録音声データを記憶する。登録音声データ記憶部2121は、複数の登録音声データを記憶する。複数の登録音声データは、複数の登録音声データそれぞれの話者を識別するための識別情報と対応付けられている。

## 【0214】

登録音声データ取得部2131は、登録音声データ記憶部2121に登録されている複数の登録音声データの中から、識別情報取得部222によって取得された識別情報と一致する識別情報が対応付けられている登録音声データを取得する。

## 【0215】

話者識別部2181は、類似度算出部2311及び類似度判定部2321を備える。

## 【0216】

類似度算出部2311は、識別対象音声データの特徴量と登録音声データの特徴量とを、選択した第1話者識別モデル及び第2話者識別モデルのいずれかに入力することにより、識別対象音声データと登録音声データとの類似度を第1話者識別モデル及び第2話者識別モデルのいずれかから取得する。

## 【0217】

類似度判定部2321は、取得した類似度が閾値よりも高い場合、登録音声データの話者を識別対象音声データの話者として識別する。

## 【0218】

続いて、本実施の形態2における話者識別装置21の登録処理及び話者識別処理のそれぞれの動作について説明する。

## 【0219】

図15は、本実施の形態2における話者識別装置の登録処理の動作について説明するためのフローチャートである。

## 【0220】

まず、ステップS121において、識別情報取得部222は、入力受付部221によって受け付けられた話者の識別情報を取得する。入力受付部221は、音声データを登録する話者を識別するための識別情報の話者による入力を受け付け、受け付けた識別情報を識別情報取得部222へ出力する。識別情報取得部222は、登録対象音声データの話者を識別するための識別情報を登録部2061へ出力する。

## 【0221】

次に、ステップS122において、登録対象音声データ取得部201は、マイクロホン1から出力された登録対象音声データを取得する。識別情報を入力した、自身が発話した音声データの登録を希望する話者は、マイクロホン1に向かって所定の文章を発話する。

## 【0222】

ステップS123及びステップS124の処理は、図4に示すステップS2及びステップS3の処理と同じであるので、説明を省略する。

## 【0223】

次に、ステップS125において、登録部2061は、性別識別部205によって識別された性別情報及び識別情報取得部222によって取得された識別情報を対応付けた登録対象音声データを登録音声データとして登録音声データ記憶部2121に記憶する。これにより、登録音声データ記憶部2121は、性別情報及び識別情報を対応付けた登録音声データを記憶する。

## 【0224】

図16は、本実施の形態2における話者識別装置の話者識別処理の動作について説明するための第1のフローチャートであり、図17は、本実施の形態2における話者識別装置の話者識別処理の動作について説明するための第2のフローチャートである。

## 【0225】

まず、ステップS131において、識別情報取得部222は、入力受付部221によ

10

20

30

40

50

て受け付けられた話者の識別情報を取得する。入力受付部 221 は、識別する音声データの話者を識別するための識別情報の話者による入力を受け付け、受け付けた識別情報を識別情報取得部 222 へ出力する。識別情報取得部 222 は、識別対象音声データの話者を識別するための識別情報を登録音声データ取得部 2131 へ出力する。

【0226】

次に、ステップ S132 において、登録音声データ取得部 2131 は、識別情報取得部 222 によって取得された識別情報と一致する識別情報が登録音声データ記憶部 2121 にあるか否かを判定する。ここで、取得された識別情報と一致する識別情報が登録音声データ記憶部 2121 にないと判定された場合（ステップ S132 で NO）、話者識別処理が終了する。なお、取得された識別情報と一致する識別情報が登録音声データ記憶部 2121 にないと判定された場合、識別結果出力部 219 は、入力された識別情報が登録されていないことを話者に通知するための通知情報を出力してもよい。また、取得された識別情報と一致する識別情報が登録音声データ記憶部 2121 にないと判定された場合、識別結果出力部 219 は、音声データの登録を話者に促すための通知情報を出力してもよい。

【0227】

一方、取得された識別情報と一致する識別情報が登録音声データ記憶部 2121 にあると判定された場合（ステップ S132 で YES）、ステップ S133 において、識別対象音声データ取得部 211 は、マイクロホン 1 から出力された識別対象音声データを取得する。

【0228】

次に、ステップ S134 において、特徴量抽出部 214 は、識別対象音声データ取得部 211 によって取得された識別対象音声データの特徴量を抽出する。

【0229】

次に、ステップ S135 において、登録音声データ取得部 2131 は、登録音声データ記憶部 2121 に登録されている複数の登録音声データの中から、識別情報取得部 222 によって取得された識別情報と一致する識別情報に対応付けられている登録音声データを取得する。

【0230】

ステップ S136 ～ステップ S141 の処理は、図 6 に示すステップ S34 ～ステップ S39 の処理と同じであるので、説明を省略する。

【0231】

次に、ステップ S142 において、類似度判定部 2321 は、類似度算出部 2311 によって算出された類似度が閾値より大きいと判定する。

【0232】

ここで、類似度算出部 2311 によって算出された類似度が閾値より大きいと判定された場合（ステップ S142 で YES）、ステップ S143 において、類似度判定部 2321 は、登録音声データの話者を識別対象音声データの話者として識別する。

【0233】

一方、類似度算出部 2311 によって算出された類似度が閾値以下であると判定された場合（ステップ S142 で NO）、ステップ S144 において、類似度判定部 2321 は、識別対象音声データの話者が登録音声データの話者ではないと判定する。

【0234】

次に、ステップ S145 において、識別結果出力部 219 は、話者識別部 2181 による識別結果を出力する。識別結果出力部 219 は、識別対象音声データの話者が識別された場合、識別対象音声データの話者が予め登録されている話者であることを示すメッセージを出力する。一方、識別結果出力部 219 は、識別対象音声データの話者が識別されなかった場合、識別対象音声データの話者が予め登録されている話者ではないことを示すメッセージを出力する。

【0235】

このように、本実施の形態 2 では、識別情報に対応付けられている登録音声データと識

10

20

30

40

50

別対象音声データとの類似度のみが算出される。したがって、複数の登録音声データそれぞれと識別対象音声データとの複数の類似度が算出される実施の形態 1 と比較して、実施の形態 2 では、類似度算出の処理負荷を低減することができる。

【0236】

(実施の形態 3)

上記の実施の形態 1 及び実施の形態 2 では、登録音声データの話者の性別に応じて第 1 話者識別モデルと第 2 話者識別モデルとのいずれかが選択されるが、2 つの異なる第 1 話者識別モデルと第 2 話者識別モデルとが用いられる場合、第 1 話者識別モデルと第 2 話者識別モデルとの出力値の範囲が異なるおそれがある。そこで、実施の形態 3 では、音声データの登録時において、第 1 話者識別モデル及び第 2 話者識別モデルのそれぞれに対し、同一の話者を識別することが可能な第 1 閾値及び第 2 閾値が算出され、算出された第 1 閾値及び第 2 閾値が記憶される。また、音声データの識別時において、算出された識別対象音声データと登録音声データとの類似度から第 1 閾値又は第 2 閾値が減算されることにより類似度が補正される。そして、補正された類似度が第 1 話者識別モデル及び第 2 話者識別モデルに共通の第 3 閾値と比較されることにより、識別対象音声データの話者が識別される。

【0237】

まず、本開示の実施の形態 3 における話者識別モデル生成装置について説明する。

【0238】

図 18 は、本開示の実施の形態 3 における話者識別モデル生成装置の構成を示す図である。

【0239】

図 18 に示す話者識別モデル生成装置 41 は、男性音声データ記憶部 401、男性音声データ取得部 402、特徴量抽出部 403、第 1 話者識別モデル生成部 404、第 1 話者識別モデル記憶部 405、第 1 話者識別部 406、第 1 閾値算出部 407、閾値記憶部 408、女性音声データ記憶部 411、女性音声データ取得部 412、特徴量抽出部 413、第 2 話者識別モデル生成部 414、第 2 話者識別モデル記憶部 415、第 2 話者識別部 416 及び第 2 閾値算出部 417 を備える。

【0240】

なお、本実施の形態 3 において、実施の形態 1 及び実施の形態 2 と同じ構成については同じ符号を付し、説明を省略する。

【0241】

第 1 話者識別部 406 は、男性の複数の音声データのうちの 2 つの音声データの特徴量の全ての組み合わせを第 1 話者識別モデルに入力することにより、2 つの音声データの複数の組み合わせそれぞれの類似度を第 1 話者識別モデルから取得する。

【0242】

第 1 閾値算出部 407 は、同一話者の 2 つの音声データの類似度と、互いに異なる話者の 2 つの音声データの類似度とを識別可能な第 1 閾値を算出する。第 1 閾値算出部 407 は、第 1 話者識別部 406 によって算出された複数の類似度を回帰分析することにより、第 1 閾値を算出する。

【0243】

第 2 話者識別部 416 は、女性の複数の音声データのうちの 2 つの音声データの特徴量の全ての組み合わせを第 2 話者識別モデルに入力することにより、2 つの音声データの複数の組み合わせそれぞれの類似度を第 2 話者識別モデルから取得する。

【0244】

第 2 閾値算出部 417 は、同一話者の 2 つの音声データの類似度と、互いに異なる話者の 2 つの音声データの類似度とを識別可能な第 2 閾値を算出する。第 2 閾値算出部 417 は、第 2 話者識別部 416 によって算出された複数の類似度を回帰分析することにより、第 2 閾値を算出する。

【0245】

閾値記憶部 408 は、第 1 閾値算出部 407 によって算出された第 1 閾値と、第 2 閾値算出部 417 によって算出された第 2 閾値とを記憶する。

【0246】

続いて、本開示の実施の形態 3 における話者識別システムについて説明する。

【0247】

図 19 は、本開示の実施の形態 3 における話者識別システムの構成を示す図である。

【0248】

図 19 に示す話者識別システムは、マイクロホン 1 及び話者識別装置 22 を備える。なお、話者識別装置 22 は、マイクロホン 1 を備えていなくてもよいし、備えてもよい。

【0249】

なお、本実施の形態 3 において、実施の形態 1 及び実施の形態 2 と同じ構成については同じ符号を付し、説明を省略する。

【0250】

話者識別装置 22 は、登録対象音声データ取得部 201、特徴量抽出部 202、性別識別モデル記憶部 203、性別識別用音声データ記憶部 204、性別識別部 205、登録部 2061、識別対象音声データ取得部 211、登録音声データ記憶部 2121、登録音声データ取得部 2131、特徴量抽出部 214、特徴量抽出部 215、話者識別モデル記憶部 216、モデル選択部 217、話者識別部 2182、識別結果出力部 219、入力受付部 221、識別情報取得部 222 及び閾値記憶部 223 を備える。

【0251】

話者識別部 2182 は、類似度算出部 2311、類似度補正部 233 及び類似度判定部 2322 を備える。

【0252】

類似度補正部 233 は、第 1 話者識別モデルから類似度を取得した場合、取得した類似度から第 1 閾値を減算する。また、類似度補正部 233 は、第 2 話者識別モデルから類似度を取得した場合、取得した類似度から第 2 閾値を減算する。類似度補正部 233 は、類似度算出部 2311 によって第 1 話者識別モデルを用いて類似度が算出された場合、閾値記憶部 223 から第 1 閾値を読み出し、算出された類似度から第 1 閾値を減算する。また、類似度補正部 233 は、類似度算出部 2311 によって第 2 話者識別モデルを用いて類似度が算出された場合、閾値記憶部 223 から第 2 閾値を読み出し、算出された類似度から第 2 閾値を減算する。

【0253】

閾値記憶部 223 は、第 1 話者識別モデルを用いて算出された類似度を補正するための第 1 閾値と、第 2 話者識別モデルを用いて算出された類似度を補正するための第 2 閾値とを予め記憶する。閾値記憶部 223 は、話者識別モデル生成装置 41 によって生成された第 1 閾値及び第 2 閾値を予め記憶する。

【0254】

なお、話者識別モデル生成装置 41 は、閾値記憶部 408 に記憶された第 1 閾値及び第 2 閾値を話者識別装置 22 へ送信してもよい。話者識別装置 22 は、受信した第 1 閾値及び第 2 閾値を閾値記憶部 223 に記憶してもよい。また、話者識別装置 22 の製造時において、話者識別モデル生成装置 41 によって生成された第 1 閾値及び第 2 閾値が閾値記憶部 223 に記憶されてもよい。

【0255】

類似度判定部 2322 は、類似度補正部 233 によって第 1 閾値又は第 2 閾値が減算された類似度が第 3 閾値よりも高い場合、登録音声データの話者を識別対象音声データの話者として識別する。

【0256】

続いて、本実施の形態 3 における話者識別モデル生成装置 41 の話者識別モデル生成処理の動作について説明する。

【0257】

図20は、本実施の形態3における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第1のフローチャートであり、図21は、本実施の形態3における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第2のフローチャートであり、図22は、本実施の形態3における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第3のフローチャートである。

【0258】

ステップS151～ステップS156の処理は、図9に示すステップS61～ステップS66の処理と同じであるので、説明を省略する。

【0259】

次に、ステップS157において、第1話者識別部406は、第1話者識別モデル記憶部405から第1話者識別モデルを取得する。

【0260】

次に、ステップS158において、第1話者識別部406は、特徴量抽出部403によって抽出された複数の男性音声データの特徴量の中から、2つの男性音声データの特徴量を取得する。

【0261】

次に、ステップS159において、第1話者識別部406は、取得した2つの男性音声データの特徴量を第1話者識別モデルに入力することにより、2つの男性音声データの類似度を算出する。なお、2つの男性音声データは、1人の話者によって発話された2つの男性音声データと、2人の話者によって発話された2つの男性音声データとのいずれかである。このとき、2つの男性音声データが1人の話者によって発話された男性音声データである場合の類似度は、2つの男性音声データが2人の話者によって発話された男性音声データである場合の類似度よりも高くなる。

【0262】

次に、ステップS160において、第1話者識別部406は、全ての男性音声データの組み合わせの類似度が算出されたか否かを判定する。ここで、全ての男性音声データの組み合わせの類似度が算出されていないと判定された場合（ステップS160でNO）、ステップS158に処理が戻る。そして、第1話者識別部406は、複数の男性音声データの特徴量の中から、類似度を算出していない2つの男性音声データの特徴量を特徴量抽出部403から取得する。

【0263】

一方、全ての男性音声データの組み合わせの類似度が算出されたと判定された場合（ステップS160でYES）、ステップS161において、第1閾値算出部407は、第1話者識別部406によって算出された複数の類似度を回帰分析することにより、同一話者の2つの男性音声データの類似度と、互いに異なる話者の2つの男性音声データの類似度とを識別可能な第1閾値を算出する。

【0264】

次に、ステップS162において、第1閾値算出部407は、算出した第1閾値を閾値記憶部408に記憶する。

【0265】

ステップS163～ステップS168の処理は、図9及び図10に示すステップS67～ステップS72の処理と同じであるので、説明を省略する。

【0266】

次に、ステップS169において、第2話者識別部416は、第2話者識別モデル記憶部415から第2話者識別モデルを取得する。

【0267】

次に、ステップS170において、第2話者識別部416は、特徴量抽出部413によって抽出された複数の女性音声データの特徴量の中から、2つの女性音声データの特徴量を取得する。

10

20

30

40

50

## 【0268】

次に、ステップS171において、第2話者識別部416は、取得した2つの女性音声データの特徴量を第2話者識別モデルに入力することにより、2つの女性音声データの類似度を算出する。なお、2つの女性音声データは、1人の話者によって発話された2つの女性音声データと、2人の話者によって発話された2つの女性音声データとのいずれかである。このとき、2つの女性音声データが1人の話者によって発話された女性音声データである場合の類似度は、2つの女性音声データが2人の話者によって発話された女性音声データである場合の類似度よりも高くなる。

## 【0269】

次に、ステップS172において、第2話者識別部416は、全ての女性音声データの組み合わせの類似度が算出されたか否かを判定する。ここで、全ての女性音声データの組み合わせの類似度が算出されていないと判定された場合（ステップS172でNO）、ステップS170に処理が戻る。そして、第2話者識別部416は、複数の女性音声データの特徴量の中から、類似度を算出していない2つの女性音声データの特徴量の特徴量抽出部413から取得する。

10

## 【0270】

一方、全ての女性音声データの組み合わせの類似度が算出されたと判定された場合（ステップS172でYES）、ステップS173において、第2閾値算出部417は、第2話者識別部416によって算出された複数の類似度を回帰分析することにより、同一話者の2つの女性音声データの類似度と、互いに異なる話者の2つの女性音声データの類似度とを識別可能な第2閾値を算出する。

20

## 【0271】

次に、ステップS174において、第2閾値算出部417は、算出した第2閾値を閾値記憶部408に記憶する。

## 【0272】

続いて、本実施の形態3における話者識別装置22の話者識別処理の動作について説明する。

## 【0273】

図23は、本実施の形態3における話者識別装置の話者識別処理の動作について説明するための第1のフローチャートであり、図24は、本実施の形態3における話者識別装置の話者識別処理の動作について説明するための第2のフローチャートである。

30

## 【0274】

ステップS181～ステップS191の処理は、図16及び図17に示すステップS131～ステップS141の処理と同じであるので、説明を省略する。

## 【0275】

次に、ステップS192において、類似度補正部233は、閾値記憶部223から第1閾値又は第2閾値を取得する。このとき、モデル選択部217によって第1話者識別モデルが選択された場合、類似度補正部233は、閾値記憶部223から第1閾値を取得する。また、モデル選択部217によって第2話者識別モデルが選択された場合、類似度補正部233は、閾値記憶部223から第2閾値を取得する。

40

## 【0276】

次に、ステップS193において、類似度補正部233は、取得した第1閾値又は第2閾値を用いて、類似度算出部2311によって算出された類似度を補正する。このとき、類似度補正部233は、類似度算出部2311によって算出された類似度から、第1閾値又は第2閾値を減算する。

## 【0277】

次に、ステップS194において、類似度判定部2322は、類似度補正部233によって補正された類似度が第3閾値より大きいかな否かを判定する。なお、第3閾値は、例えば0である。類似度判定部2322は、補正された類似度が0より大きい場合、識別対象音声データが予め登録されている登録音声データと一致すると判定し、補正された類似度

50

が0以下である場合、識別対象音声データが予め登録されている登録音声データと一致しないと判定する。

【0278】

ここで、類似度補正部233によって補正された類似度が第3閾値より大きいと判定された場合（ステップS194でYES）、ステップS195において、類似度判定部232は、登録音声データの話者を識別対象音声データの話者として識別する。

【0279】

一方、類似度補正部233によって補正された類似度が第3閾値以下であると判定された場合（ステップS194でNO）、ステップS196において、類似度判定部232は、識別対象音声データの話者が登録音声データの話者ではないと判定する。

10

【0280】

ステップS197の処理は、図17に示すステップS145の処理と同じであるので、説明を省略する。

【0281】

2つの異なる第1話者識別モデルと第2話者識別モデルとが用いられる場合、第1話者識別モデルと第2話者識別モデルとの出力値の範囲が異なるおそれがある。そこで、本実施の形態3では、登録時において、第1話者識別モデル及び第2話者識別モデルのそれぞれに対し、同一の話者を識別することが可能な第1閾値及び第2閾値が算出される。また、話者識別時において、算出された識別対象音声データと登録音声データとの類似度から第1閾値又は第2閾値が減算されることにより類似度が補正される。そして、補正された類似度が第1話者識別モデル及び第2話者識別モデルに共通の第3閾値と比較されることにより、識別対象音声データの話者をより高い精度で識別することができる。

20

【0282】

（実施の形態4）

発話時間が長いほど発話内の情報量が多いため、話者の識別が容易であり、本人同士の登録音声データと識別対象音声データとの類似度が高くなりやすい。一方、発話時間が短いほど発話内の情報量が少なくなるため、話者の識別が困難であり、本人同士の登録音声データと識別対象音声データとであっても類似度が低くなるおそれがある。そのため、発話時間が長い音声データを用いて機械学習した話者識別モデルにより、発話時間が短い識別対象音声データの話者を識別した場合、話者識別の精度が低下するおそれがある。

30

【0283】

そこで、実施の形態4に係る話者識別方法は、コンピュータが、識別対象音声データを取得し、予め登録されている登録音声データを取得し、識別対象音声データの特徴量を抽出し、登録音声データの特徴量を抽出し、識別対象音声データの話者及び登録音声データの話者の少なくとも一方の発話時間が所定時間以上である場合、発話時間が所定時間以上の話者を識別するために発話時間が所定時間以上である音声データを用いて機械学習された第3話者識別モデルを選択し、識別対象音声データの話者及び登録音声データの話者の少なくとも一方の発話時間が所定時間より短い場合、発話時間が所定時間より短い話者を識別するために発話時間が所定時間より短い音声データを用いて機械学習された第4話者識別モデルを選択し、識別対象音声データの特徴量と登録音声データの特徴量とを、選択した第3話者識別モデル及び第4話者識別モデルのいずれかに入力することにより、識別対象音声データの話者を識別する。

40

【0284】

図25は、本開示の実施の形態4における話者識別システムの構成を示す図である。

【0285】

図25に示す話者識別システムは、マイクロホン1及び話者識別装置24を備える。なお、話者識別装置24は、マイクロホン1を備えていなくてもよいし、備えてもよい。

【0286】

なお、本実施の形態4において、実施の形態1と同じ構成については同じ符号を付し、説明を省略する。

50

## 【0287】

話者識別装置24は、登録対象音声データ取得部201、発話時間計測部207、登録部2064、識別対象音声データ取得部211、登録音声データ記憶部2124、登録音声データ取得部213、特徴量抽出部214、特徴量抽出部215、話者識別モデル記憶部2164、モデル選択部2174、話者識別部2184及び識別結果出力部219を備える。

## 【0288】

発話時間計測部207は、登録対象音声データ取得部201によって取得された登録対象音声データの発話時間を計測する。なお、発話時間は、登録対象音声データ取得部201によって登録対象音声データの取得が開始された時刻から登録対象音声データの取得が終了された時刻までの時間である。

10

## 【0289】

登録部2064は、発話時間計測部207によって計測された発話時間を示す発話時間情報を対応付けた登録対象音声データを登録音声データとして登録する。登録部2064は、登録音声データを登録音声データ記憶部2124に登録する。

## 【0290】

なお、話者識別装置24は、登録対象音声データの話者に関する情報の入力を受け付ける入力受付部をさらに備えてもよい。そして、登録部2064は、話者に関する情報に対応付けて登録音声データを登録音声データ記憶部2124に登録してもよい。話者に関する情報は、例えば、話者の名前などである。

20

## 【0291】

登録音声データ記憶部2124は、発話時間情報が対応付けられた登録音声データを記憶する。登録音声データ記憶部2124は、複数の登録音声データを記憶する。

## 【0292】

話者識別モデル記憶部2164は、発話時間が所定時間以上の話者を識別するために発話時間が所定時間以上である音声データを用いて機械学習された第3話者識別モデルと、発話時間が所定時間より短い話者を識別するために発話時間が所定時間より短い音声データを用いて機械学習された第4話者識別モデルとを予め記憶する。話者識別モデル記憶部2164は、後述する話者識別モデル生成装置44によって生成された第3話者識別モデル及び第4話者識別モデルを予め記憶する。なお、第3話者識別モデル及び第4話者識別モデルの生成方法については後述する。

30

## 【0293】

モデル選択部2174は、識別対象音声データの話者及び登録音声データの話者の少なくとも一方の発話時間が所定時間以上である場合、発話時間が所定時間以上の話者を識別するために発話時間が所定時間以上である音声データを用いて機械学習された第3話者識別モデルを選択する。また、モデル選択部2174は、識別対象音声データの話者及び登録音声データの話者の少なくとも一方の発話時間が所定時間より短い場合、発話時間が所定時間より短い話者を識別するために発話時間が所定時間より短い音声データを用いて機械学習された第4話者識別モデルを選択する。

## 【0294】

本実施の形態4において、モデル選択部2174は、登録音声データの話者の発話時間が所定時間以上である場合、第3話者識別モデルを選択し、登録音声データの話者の発話時間が所定時間より短い場合、第4話者識別モデルを選択する。登録音声データには、予め発話時間が対応付けられている。そのため、モデル選択部2174は、登録音声データに対応付けられている発話時間が所定時間以上である場合、第3話者識別モデルを選択し、登録音声データに対応付けられている発話時間が所定時間より短い場合、第4話者識別モデルを選択する。なお、所定時間は、例えば60秒である。

40

## 【0295】

話者識別部2184は、識別対象音声データの特徴量と登録音声データの特徴量とを、モデル選択部2174によって選択された第3話者識別モデル及び第4話者識別モデルの

50

いずれかに入力することにより、識別対象音声データの話者を識別する。

【0296】

話者識別部2184は、類似度算出部2314及び類似度判定部232を備える。

【0297】

類似度算出部2314は、識別対象音声データの特徴量と複数の登録音声データの特徴量それぞれとを、選択された第3話者識別モデル及び第4話者識別モデルのいずれかに入力することにより、識別対象音声データと複数の登録音声データそれぞれの類似度を第3話者識別モデル及び第4話者識別モデルのいずれかから取得する。

【0298】

続いて、本開示の実施の形態4における話者識別モデル生成装置について説明する。

10

【0299】

図26は、本開示の実施の形態4における話者識別モデル生成装置の構成を示す図である。

【0300】

図26に示す話者識別モデル生成装置44は、長時間音声データ記憶部421、長時間音声データ取得部422、特徴量抽出部423、第3話者識別モデル生成部424、第3話者識別モデル記憶部425、短時間音声データ記憶部431、短時間音声データ取得部432、特徴量抽出部433、第4話者識別モデル生成部434及び第4話者識別モデル記憶部435を備える。

【0301】

20

長時間音声データ取得部422、特徴量抽出部423、第3話者識別モデル生成部424、短時間音声データ取得部432、特徴量抽出部433及び第4話者識別モデル生成部434は、プロセッサにより実現される。長時間音声データ記憶部421、第3話者識別モデル記憶部425、短時間音声データ記憶部431及び第4話者識別モデル記憶部435は、メモリにより実現される。

【0302】

長時間音声データ記憶部421は、話者を識別するための話者識別ラベルが付与された、発話時間が所定時間以上である複数の長時間音声データを記憶する。長時間音声データは、発話時間が所定時間以上である音声データである。長時間音声データ記憶部421は、複数の話者毎に互いに異なる複数の長時間音声データを記憶する。

30

【0303】

長時間音声データ取得部422は、話者を識別するための話者識別ラベルが付与された複数の長時間音声データを長時間音声データ記憶部421から取得する。なお、本実施の形態4では、長時間音声データ取得部422は、複数の長時間音声データを長時間音声データ記憶部421から取得しているが、本開示は特にこれに限定されず、ネットワークを介して外部機器から複数の長時間音声データを取得（受信）してもよい。

【0304】

特徴量抽出部423は、長時間音声データ取得部422によって取得された複数の長時間音声データの特徴量を抽出する。特徴量は、例えば *vector* である。

【0305】

40

第3話者識別モデル生成部424は、複数の長時間音声データのうちの第1長時間音声データ及び第2長時間音声データの各特徴量と、第1長時間音声データ及び第2長時間音声データの各話者識別ラベルの類似度とを教師データとして用い、入力を2つの音声データの各特徴量とし、出力を2つの音声データの類似度とする第3話者識別モデルを機械学習により生成する。例えば、第3話者識別モデルは、第1長時間音声データの話者識別ラベルと第2長時間音声データの話者識別ラベルとが同じであれば、最も高い類似度が出力され、第1長時間音声データの話者識別ラベルと第2長時間音声データの話者識別ラベルとが異なっていれば、最も低い類似度が出力されるように機械学習される。

【0306】

第3話者識別モデルとしては、PLDAによるモデルが用いられる。PLDAモデルは

50

、400次元の *i - v e c t o r*（特徴量）から話者の識別に有効な特徴量を自動的に選択し、対数尤度比を類似度として算出する。

【0307】

なお、機械学習としては、例えば、入力情報に対してラベル（出力情報）が付与された教師データを用いて入力と出力との関係を学習する教師あり学習、ラベルのない入力のみからデータの構造を構築する教師なし学習、ラベルありとラベルなしとのどちらも扱う半教師あり学習、報酬を最大化する行動を試行錯誤により学習する強化学習なども挙げられる。また、機械学習の具体的な手法としては、ニューラルネットワーク（多層のニューラルネットワークを用いた深層学習を含む）、遺伝的プログラミング、決定木、ベイジアン・ネットワーク、又はサポート・ベクター・マシン（SVM）などが存在する。第3話者識別モデルの機械学習においては、以上で挙げた具体例のいずれかを用いればよい。

10

【0308】

第3話者識別モデル記憶部425は、第3話者識別モデル生成部424によって生成された第3話者識別モデルを記憶する。

【0309】

短時間音声データ記憶部431は、話者を識別するための話者識別ラベルが付与された、発話時間が所定時間より短い複数の短時間音声データを記憶する。短時間音声データは、発話時間が所定時間より短い音声データである。短時間音声データ記憶部431は、複数の話者毎に互いに異なる複数の短時間音声データを記憶する。

【0310】

短時間音声データ取得部432は、話者を識別するための話者識別ラベルが付与された複数の短時間音声データを短時間音声データ記憶部431から取得する。なお、本実施の形態4では、短時間音声データ取得部432は、複数の短時間音声データを短時間音声データ記憶部431から取得しているが、本開示は特にこれに限定されず、ネットワークを介して外部機器から複数の短時間音声データを取得（受信）してもよい。

20

【0311】

特徴量抽出部433は、短時間音声データ取得部432によって取得された複数の短時間音声データの特徴量を抽出する。特徴量は、例えば *i - v e c t o r* である。

【0312】

第4話者識別モデル生成部434は、複数の短時間音声データのうちの第1短時間音声データ及び第2短時間音声データの各特徴量と、第1短時間音声データ及び第2短時間音声データの各話者識別ラベルの類似度とを教師データとして用い、入力を2つの音声データの各特徴量とし、出力を2つの音声データの類似度とする第4話者識別モデルを機械学習により生成する。例えば、第4話者識別モデルは、第1短時間音声データの話者識別ラベルと第2短時間音声データの話者識別ラベルとが同じであれば、最も高い類似度が出力され、第1短時間音声データの話者識別ラベルと第2短時間音声データの話者識別ラベルとが異なっていれば、最も低い類似度が出力されるように機械学習される。

30

【0313】

第4話者識別モデルとしては、PLDAによるモデルが用いられる。PLDAモデルは、400次元の *i - v e c t o r*（特徴量）から話者の識別に有効な特徴量を自動的に選択し、対数尤度比を類似度として算出する。

40

【0314】

なお、機械学習としては、例えば、入力情報に対してラベル（出力情報）が付与された教師データを用いて入力と出力との関係を学習する教師あり学習、ラベルのない入力のみからデータの構造を構築する教師なし学習、ラベルありとラベルなしとのどちらも扱う半教師あり学習、報酬を最大化する行動を試行錯誤により学習する強化学習なども挙げられる。また、機械学習の具体的な手法としては、ニューラルネットワーク（多層のニューラルネットワークを用いた深層学習を含む）、遺伝的プログラミング、決定木、ベイジアン・ネットワーク、又はサポート・ベクター・マシン（SVM）などが存在する。第4話者識別モデルの機械学習においては、以上で挙げた具体例のいずれかを用いればよい。

50

## 【0315】

第4話者識別モデル記憶部435は、第4話者識別モデル生成部434によって生成された第4話者識別モデルを記憶する。

## 【0316】

なお、話者識別モデル生成装置44は、第3話者識別モデル記憶部425に記憶された第3話者識別モデル及び第4話者識別モデル記憶部435に記憶された第4話者識別モデルを話者識別装置24へ送信してもよい。話者識別装置24は、受信した第3話者識別モデル及び第4話者識別モデルを話者識別モデル記憶部2164に記憶してもよい。また、話者識別装置24の製造時において、話者識別モデル生成装置44によって生成された第3話者識別モデル及び第4話者識別モデルが話者識別装置24に記憶されてもよい。

10

## 【0317】

続いて、本実施の形態4における話者識別装置24の登録処理及び話者識別処理のそれぞれの動作について説明する。

## 【0318】

図27は、本実施の形態4における話者識別装置の登録処理の動作について説明するためのフローチャートである。

## 【0319】

まず、ステップS201において、登録対象音声データ取得部201は、マイクロホン1から出力された登録対象音声データを取得する。自身が発話した音声データの登録を希望する話者は、マイクロホン1に向かって所定の文章を発話する。このとき、所定の文章は、発話時間が所定時間以上になる文章、及び発話時間が所定時間より短くなる文章のいずれかである。話者識別装置24は、予め決められている複数の文章を登録対象話者に提示してもよい。この場合、登録対象話者は、提示された複数の文章を発話する。

20

## 【0320】

次に、ステップS202において、発話時間計測部207は、登録対象音声データ取得部201によって取得された登録対象音声データの発話時間を計測する。

## 【0321】

次に、ステップS203において、登録部2064は、発話時間計測部207によって計測された発話時間を示す発話時間情報に対応付けた登録対象音声データを登録音声データとして登録音声データ記憶部2124に記憶する。

30

## 【0322】

図28は、本実施の形態4における話者識別装置の話者識別処理の動作について説明するための第1のフローチャートであり、図29は、本実施の形態4における話者識別装置の話者識別処理の動作について説明するための第2のフローチャートである。

## 【0323】

ステップS211～ステップS214の処理は、図6に示すステップS31～ステップS34の処理と同じであるので、説明を省略する。

## 【0324】

次に、ステップS215において、モデル選択部2174は、登録音声データ取得部213によって取得された登録音声データに対応付けられている発話時間を取得する。

40

## 【0325】

次に、ステップS216において、モデル選択部2174は、取得した発話時間が所定時間以上であるか否かを判定する。ここで、取得した発話時間が所定時間以上であると判定された場合（ステップS216でYES）、ステップS217において、モデル選択部2174は、第3話者識別モデルを選択する。モデル選択部2174は、選択した第3話者識別モデルを話者識別モデル記憶部2164から取得し、取得した第3話者識別モデルを類似度算出部2314へ出力する。

## 【0326】

一方、取得した発話時間が所定時間以上ではないと判定された場合、すなわち取得した発話時間が所定時間より短いと判定された場合（ステップS216でNO）、ステップS

50

218において、モデル選択部2174は、第4話者識別モデルを選択する。モデル選択部2174は、選択した第4話者識別モデルを話者識別モデル記憶部2164から取得し、取得した第4話者識別モデルを類似度算出部2314へ出力する。

【0327】

次に、ステップS219において、類似度算出部2314は、識別対象音声データの特徴量と登録音声データの特徴量とを、選択された第3話者識別モデル及び第4話者識別モデルのいずれかに入力することにより、識別対象音声データと登録音声データとの類似度を算出する。

【0328】

次に、ステップS220において、類似度算出部2314は、識別対象音声データと、登録音声データ記憶部2124に記憶されている全ての登録音声データとの類似度が算出されたか否かを判定する。ここで、識別対象音声データと全ての登録音声データとの類似度が算出されていないと判定された場合（ステップS220でNO）、ステップS213に処理が戻る。そして、登録音声データ取得部213は、登録音声データ記憶部2124に記憶されている複数の登録音声データの中から、類似度が算出されていない登録音声データを取得する。

10

【0329】

一方、識別対象音声データと全ての登録音声データとの類似度が算出されたと判定された場合（ステップS220でYES）、ステップS221において、類似度判定部232は、最も高い類似度が閾値より大きいかな否かを判定する。

20

【0330】

なお、ステップS221～ステップS224の処理は、図7に示すステップS41～ステップS44の処理と同じであるので、説明を省略する。

【0331】

このように、識別対象音声データの話者及び登録音声データの話者の少なくとも一方の発話時間が所定時間以上である場合、識別対象音声データの特徴量と登録音声データの特徴量とが、発話時間が所定時間以上である音声データを用いて機械学習された第3話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。また、識別対象音声データの話者及び登録音声データの話者の少なくとも一方の発話時間が所定時間より短い場合、識別対象音声データの特徴量と登録音声データの特徴量とが、発話時間が所定時間より短い音声データを用いて機械学習された第4話者識別モデルに入力されることにより、識別対象音声データの話者が識別される。

30

【0332】

したがって、識別対象音声データ及び登録音声データの少なくとも一方の発話時間の長さに応じた第3話者識別モデル及び第4話者識別モデルにより、識別対象音声データの話者が識別されるので、識別対象の話者が予め登録されている話者であるかな否かを識別する精度を向上させることができる。

【0333】

なお、本実施の形態4において、モデル選択部2174は、登録音声データに対応付けられている発話時間に基づいて、第3話者識別モデル及び第4話者識別モデルのいずれかを選択しているが、本開示は特にこれに限定されない。話者識別装置24は、登録音声データ取得部213によって取得された登録音声データの話者の発話時間を計測する発話時間計測部を備えてもよい。発話時間計測部は、計測した発話時間をモデル選択部2174へ出力してもよい。なお、登録音声データ取得部213によって取得された登録音声データの話者の発話時間が計測される場合、発話時間計測部207は不要となり、登録部2064は、登録対象音声データ取得部201によって取得された登録対象音声データのみを登録音声データとして登録音声データ記憶部2124に記憶してもよい。

40

【0334】

また、本実施の形態4において、モデル選択部2174は、登録音声データの話者の発話時間に基づいて、第3話者識別モデル及び第4話者識別モデルのいずれかを選択してい

50

るが、本開示は特にこれに限定されない。モデル選択部 2174 は、識別対象音声データの話者の発話時間に基づいて、第3話者識別モデル及び第4話者識別モデルのいずれかを選択してもよい。この場合、話者識別装置 24 は、識別対象音声データの話者の発話時間を計測する発話時間計測部を備えてもよい。発話時間計測部は、識別対象音声データ取得部 211 によって取得された識別対象音声データの発話時間を計測し、計測した発話時間をモデル選択部 2174 へ出力してもよい。モデル選択部 2174 は、識別対象音声データの話者の発話時間が所定時間以上である場合、第3話者識別モデルを選択し、識別対象音声データの話者の発話時間が所定時間より短い場合、第4話者識別モデルを選択してもよい。なお、所定時間は、例えば30秒である。また、識別対象音声データの話者の発話時間が計測される場合、発話時間計測部 207 は不要となり、登録部 2064 は、登録対象音声データ取得部 201 によって取得された登録対象音声データのみを登録音声データとして登録音声データ記憶部 2124 に記憶してもよい。

10

#### 【0335】

また、本実施の形態4において、モデル選択部 2174 は、登録音声データの話者の発話時間と識別対象音声データの話者の発話時間との両方に基づいて、第3話者識別モデル及び第4話者識別モデルのいずれかを選択してもよい。モデル選択部 2174 は、登録音声データの話者の発話時間と識別対象音声データの話者の発話時間との両方が所定時間以上である場合、第3話者識別モデルを選択してもよい。また、モデル選択部 2174 は、登録音声データの話者の発話時間と識別対象音声データの話者の発話時間との少なくとも一方が所定時間より短い場合、第4話者識別モデルを選択してもよい。なお、所定時間は、例えば20秒である。この場合、話者識別装置 24 は、識別対象音声データの話者の発話時間を計測する発話時間計測部をさらに備えてもよい。また、話者識別装置 24 は、発話時間計測部 207 を備えずに、登録音声データ取得部 213 によって取得された登録音声データの話者の発話時間を計測する発話時間計測部を備えてもよい。

20

#### 【0336】

さらに、本実施の形態4において、モデル選択部 2174 は、登録音声データの話者の発話時間と識別対象音声データの話者の発話時間とのいずれか一方が所定時間以上である場合、第3話者識別モデルを選択してもよい。また、モデル選択部 2174 は、登録音声データの話者の発話時間と識別対象音声データの話者の発話時間との両方が所定時間より短い場合、第4話者識別モデルを選択してもよい。なお、所定時間は、例えば100秒である。この場合、話者識別装置 24 は、識別対象音声データの話者の発話時間を計測する発話時間計測部をさらに備えてもよい。また、話者識別装置 24 は、発話時間計測部 207 を備えずに、登録音声データ取得部 213 によって取得された登録音声データの話者の発話時間を計測する発話時間計測部を備えてもよい。

30

#### 【0337】

続いて、本実施の形態4における話者識別モデル生成装置 44 の話者識別モデル生成処理の動作について説明する。

#### 【0338】

図30は、本実施の形態4における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第1のフローチャートであり、図31は、本実施の形態4における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第2のフローチャートである。

40

#### 【0339】

まず、ステップ S231 において、長時間音声データ取得部 422 は、話者を識別するための話者識別ラベルが付与された、発話時間が所定時間以上である複数の長時間音声データを長時間音声データ記憶部 421 から取得する。

#### 【0340】

次に、ステップ S232 において、特徴量抽出部 423 は、長時間音声データ取得部 422 によって取得された複数の長時間音声データの特徴量を抽出する。

#### 【0341】

50

次に、ステップ S 2 3 3 において、第 3 話者識別モデル生成部 4 2 4 は、複数の長時間音声データのうちの第 1 長時間音声データ及び第 2 長時間音声データの各特徴量と、第 1 長時間音声データ及び第 2 長時間音声データの各話者識別ラベルの類似度とを教師データとして取得する。

【0 3 4 2】

次に、ステップ S 2 3 4 において、第 3 話者識別モデル生成部 4 2 4 は、取得した教師データを用い、入力を 2 つの音声データの各特徴量とし、出力を 2 つの音声データの類似度とする第 3 話者識別モデルを機械学習する。

【0 3 4 3】

次に、ステップ S 2 3 5 において、第 3 話者識別モデル生成部 4 2 4 は、複数の長時間音声データのうちの全ての長時間音声データの組み合わせを用いて第 3 話者識別モデルを機械学習させたか否かを判定する。ここで、全ての長時間音声データの組み合わせを用いて第 3 話者識別モデルを機械学習させていないと判定された場合（ステップ S 2 3 5 で N O）、ステップ S 2 3 3 に処理が戻る。そして、第 3 話者識別モデル生成部 4 2 4 は、複数の長時間音声データのうちの機械学習に用いていない組み合わせの第 1 長時間音声データ及び第 2 長時間音声データの各特徴量と、第 1 長時間音声データ及び第 2 長時間音声データの各話者識別ラベルの類似度とを教師データとして取得する。

【0 3 4 4】

一方、全ての長時間音声データの組み合わせを用いて第 3 話者識別モデルを機械学習させたと判定された場合（ステップ S 2 3 5 で Y E S）、ステップ S 2 3 6 において、第 3 話者識別モデル生成部 4 2 4 は、機械学習により生成された第 3 話者識別モデルを第 3 話者識別モデル記憶部 4 2 5 に記憶する。

【0 3 4 5】

次に、ステップ S 2 3 7 において、短時間音声データ取得部 4 3 2 は、話者を識別するための話者識別ラベルが付与された、発話時間が所定時間より短い複数の短時間音声データを短時間音声データ記憶部 4 3 1 から取得する。

【0 3 4 6】

次に、ステップ S 2 3 8 において、特徴量抽出部 4 3 3 は、短時間音声データ取得部 4 3 2 によって取得された複数の短時間音声データの特徴量を抽出する。

【0 3 4 7】

次に、ステップ S 2 3 9 において、第 4 話者識別モデル生成部 4 3 4 は、複数の短時間音声データのうちの第 1 短時間音声データ及び第 2 短時間音声データの各特徴量と、第 1 短時間音声データ及び第 2 短時間音声データの各話者識別ラベルの類似度とを教師データとして取得する。

【0 3 4 8】

次に、ステップ S 2 4 0 において、第 4 話者識別モデル生成部 4 3 4 は、取得した教師データを用い、入力を 2 つの音声データの各特徴量とし、出力を 2 つの音声データの類似度とする第 4 話者識別モデルを機械学習する。

【0 3 4 9】

次に、ステップ S 2 4 1 において、第 4 話者識別モデル生成部 4 3 4 は、複数の短時間音声データのうちの全ての短時間音声データの組み合わせを用いて第 4 話者識別モデルを機械学習させたか否かを判定する。ここで、全ての短時間音声データの組み合わせを用いて第 4 話者識別モデルを機械学習させていないと判定された場合（ステップ S 2 4 1 で N O）、ステップ S 2 3 9 に処理が戻る。そして、第 4 話者識別モデル生成部 4 3 4 は、複数の短時間音声データのうちの機械学習に用いていない組み合わせの第 1 短時間音声データ及び第 2 短時間音声データの各特徴量と、第 1 短時間音声データ及び第 2 短時間音声データの各話者識別ラベルの類似度とを教師データとして取得する。

【0 3 5 0】

一方、全ての短時間音声データの組み合わせを用いて第 4 話者識別モデルを機械学習させたと判定された場合（ステップ S 2 4 1 で Y E S）、ステップ S 2 4 2 において、第 4

10

20

30

40

50

話者識別モデル生成部 434 は、機械学習により生成された第 4 話者識別モデルを第 4 話者識別モデル記憶部 435 に記憶する。

【0351】

(実施の形態 5)

上記の実施の形態 4 では、登録音声データ記憶部 2124 に記憶されている全ての登録音声データそれぞれと、識別対象音声データとの類似度が算出され、類似度が最も高い登録音声データの話者が識別対象音声データの話者として識別される。一方、実施の形態 5 では、識別対象音声データの話者の識別情報が入力され、登録音声データ記憶部 212 に記憶されている複数の登録音声データのうち、当該識別情報に予め対応付けられている 1 の登録音声データが取得される。そして、1 の登録音声データと識別対象音声データとの類似度が算出され、類似度が閾値より高い場合、登録音声データの話者が識別対象音声データの話者として識別される。

10

【0352】

図 32 は、本開示の実施の形態 5 における話者識別システムの構成を示す図である。

【0353】

図 32 に示す話者識別システムは、マイクロホン 1 及び話者識別装置 25 を備える。なお、話者識別装置 25 は、マイクロホン 1 を備えていなくてもよいし、備えてもよい。

【0354】

なお、本実施の形態 5 において、実施の形態 1～4 と同じ構成については同じ符号を付し、説明を省略する。

20

【0355】

話者識別装置 25 は、登録対象音声データ取得部 201、発話時間計測部 207、登録部 2065、識別対象音声データ取得部 211、登録音声データ記憶部 2125、登録音声データ取得部 2135、特徴量抽出部 214、特徴量抽出部 215、話者識別モデル記憶部 2164、モデル選択部 2174、話者識別部 2185、識別結果出力部 219、入力受付部 221 及び識別情報取得部 222 を備える。

【0356】

登録部 2065 は、発話時間計測部 207 によって計測された発話時間を示す発話時間情報及び識別情報取得部 222 によって取得された識別情報に対応付けた登録対象音声データを登録音声データとして登録する。登録部 2065 は、登録音声データを登録音声データ記憶部 2125 に登録する。

30

【0357】

登録音声データ記憶部 2125 は、発話時間情報及び識別情報に対応付けられた登録音声データを記憶する。登録音声データ記憶部 2125 は、複数の登録音声データを記憶する。複数の登録音声データは、複数の登録音声データそれぞれの話者を識別するための識別情報と対応付けられている。

【0358】

登録音声データ取得部 2135 は、登録音声データ記憶部 2125 に登録されている複数の登録音声データの中から、識別情報取得部 222 によって取得された識別情報と一致する識別情報に対応付けられている登録音声データを取得する。

40

【0359】

話者識別部 2185 は、類似度算出部 2315 及び類似度判定部 2325 を備える。

【0360】

類似度算出部 2315 は、識別対象音声データの特徴量と登録音声データの特徴量とを、選択した第 3 話者識別モデル及び第 4 話者識別モデルのいずれかに入力することにより、識別対象音声データと登録音声データとの類似度を第 3 話者識別モデル及び第 4 話者識別モデルのいずれかから取得する。

【0361】

類似度判定部 2325 は、取得した類似度が閾値よりも高い場合、登録音声データの話者を識別対象音声データの話者として識別する。

50

## 【0362】

続いて、本実施の形態5における話者識別装置25の登録処理及び話者識別処理のそれぞれの動作について説明する。

## 【0363】

図33は、本実施の形態5における話者識別装置の登録処理の動作について説明するためのフローチャートである。

## 【0364】

ステップS251及びステップS252の処理は、図15に示すステップS121及びステップS122の処理と同じであるので、説明を省略する。また、ステップS253の処理は、図27に示すステップS202の処理と同じであるので、説明を省略する。

10

## 【0365】

次に、ステップS254において、登録部2065は、発話時間計測部207によって計測された発話時間情報及び識別情報取得部222によって取得された識別情報を対応付けた登録対象音声データを登録音声データとして登録音声データ記憶部2125に記憶する。これにより、登録音声データ記憶部2125は、発話時間情報及び識別情報を対応付けた登録音声データを記憶する。

## 【0366】

図34は、本実施の形態5における話者識別装置の話者識別処理の動作について説明するための第1のフローチャートであり、図35は、本実施の形態5における話者識別装置の話者識別処理の動作について説明するための第2のフローチャートである。

20

## 【0367】

ステップS261～ステップS264の処理は、図16に示すステップS131～ステップS134の処理と同じであるので、説明を省略する。

## 【0368】

次に、ステップS265において、登録音声データ取得部2135は、登録音声データ記憶部2125に登録されている複数の登録音声データの中から、識別情報取得部222によって取得された識別情報と一致する識別情報が対応付けられている登録音声データを取得する。

## 【0369】

ステップS266～ステップS270の処理は、図28に示すステップS214～ステップS218の処理と同じであるので、説明を省略する。

30

## 【0370】

次に、ステップS271において、類似度算出部2315は、識別対象音声データの特徴量と登録音声データの特徴量とを、選択した第3話者識別モデル及び第4話者識別モデルのいずれかに入力することにより、識別対象音声データと登録音声データとの類似度を第3話者識別モデル及び第4話者識別モデルのいずれかから取得する。

## 【0371】

次に、ステップS272において、類似度判定部2325は、類似度算出部2315によって算出された類似度が閾値より大きいかな否かを判定する。

## 【0372】

ここで、類似度算出部2315によって算出された類似度が閾値より大きいと判定された場合（ステップS272でYES）、ステップS273において、類似度判定部2325は、登録音声データの話者を識別対象音声データの話者として識別する。

40

## 【0373】

一方、類似度算出部2315によって算出された類似度が閾値以下であると判定された場合（ステップS272でNO）、ステップS274において、類似度判定部2325は、識別対象音声データの話者が登録音声データの話者ではないと判定する。

## 【0374】

ステップS275の処理は、図7に示すステップS44の処理と同じであるので、説明を省略する。

50

## 【0375】

このように、本実施の形態5では、識別情報に対応付けられている登録音声データと識別対象音声データとの類似度のみが算出される。したがって、複数の登録音声データそれぞれと識別対象音声データとの複数の類似度が算出される実施の形態4と比較して、実施の形態5では、類似度算出の処理負荷を低減することができる。

## 【0376】

## (実施の形態6)

上記の実施の形態4及び実施の形態5では、登録音声データの話者の発話時間に応じて第3話者識別モデルと第4話者識別モデルとのいずれかが選択されるが、2つの異なる第3話者識別モデルと第4話者識別モデルとが用いられる場合、第3話者識別モデルと第4話者識別モデルとの出力値の範囲が異なるおそれがある。そこで、実施の形態6では、音声データの登録時において、第3話者識別モデル及び第4話者識別モデルのそれぞれに対し、同一の話者を識別することが可能な第3閾値及び第4閾値が算出され、算出された第3閾値及び第4閾値が記憶される。また、音声データの識別時において、算出された識別対象音声データと登録音声データとの類似度から第3閾値又は第4閾値が減算されることにより類似度が補正される。そして、補正された類似度が第3話者識別モデル及び第4話者識別モデルに共通の第5閾値と比較されることにより、識別対象音声データの話者が識別される。

## 【0377】

まず、本開示の実施の形態6における話者識別モデル生成装置について説明する。

## 【0378】

図36は、本開示の実施の形態6における話者識別モデル生成装置の構成を示す図である。

## 【0379】

図36に示す話者識別モデル生成装置46は、長時間音声データ記憶部421、長時間音声データ取得部422、特徴量抽出部423、第3話者識別モデル生成部424、第3話者識別モデル記憶部425、第3話者識別部426、第3閾値算出部427、閾値記憶部428、短時間音声データ記憶部431、短時間音声データ取得部432、特徴量抽出部433、第4話者識別モデル生成部434、第4話者識別モデル記憶部435、第4話者識別部436及び第4閾値算出部437を備える。

## 【0380】

なお、本実施の形態6において、実施の形態1～5と同じ構成については同じ符号を付し、説明を省略する。

## 【0381】

第3話者識別部426は、発話時間が所定時間以上である複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせを第3話者識別モデルに入力することにより、2つの音声データの複数の組み合わせそれぞれの類似度を第3話者識別モデルから取得する。

## 【0382】

第3閾値算出部427は、同一話者の2つの音声データの類似度と、互いに異なる話者の2つの音声データの類似度とを識別可能な第3閾値を算出する。第3閾値算出部427は、第3話者識別部426によって算出された複数の類似度を回帰分析することにより、第3閾値を算出する。

## 【0383】

第4話者識別部436は、発話時間が所定の時間よりも短い複数の音声データのうちの2つの音声データの特徴量の全ての組み合わせを第4話者識別モデルに入力することにより、2つの音声データの複数の組み合わせそれぞれの類似度を第4話者識別モデルから取得する。

## 【0384】

第4閾値算出部437は、同一話者の2つの音声データの類似度と、互いに異なる話者

10

20

30

40

50

の２つの音声データの類似度とを識別可能な第４閾値を算出する。第４閾値算出部４３７は、第４話者識別部４３６によって算出された複数の類似度を回帰分析することにより、第４閾値を算出する。

【０３８５】

閾値記憶部４２８は、第３閾値算出部４２７によって算出された第３閾値と、第４閾値算出部４３７によって算出された第４閾値とを記憶する。

【０３８６】

続いて、本開示の実施の形態６における話者識別システムについて説明する。

【０３８７】

図３７は、本開示の実施の形態６における話者識別システムの構成を示す図である。

10

【０３８８】

図３７に示す話者識別システムは、マイクロホン１及び話者識別装置２６を備える。なお、話者識別装置２６は、マイクロホン１を備えていなくてもよいし、備えてもよい。

【０３８９】

なお、本実施の形態６において、実施の形態１～５と同じ構成については同じ符号を付し、説明を省略する。

【０３９０】

話者識別装置２６は、登録対象音声データ取得部２０１、発話時間計測部２０７、登録部２０６５、識別対象音声データ取得部２１１、登録音声データ記憶部２１２５、登録音声データ取得部２１３５、特徴量抽出部２１４、特徴量抽出部２１５、話者識別モデル記憶部２１６４、モデル選択部２１７４、話者識別部２１８６、識別結果出力部２１９、入力受付部２２１、識別情報取得部２２２及び閾値記憶部２２３６を備える。

20

【０３９１】

話者識別部２１８６は、類似度算出部２３１５、類似度補正部２３３６及び類似度判定部２３２６を備える。

【０３９２】

類似度補正部２３３６は、第３話者識別モデルから類似度を取得した場合、取得した類似度から第３閾値を減算する。また、類似度補正部２３３６は、第４話者識別モデルから類似度を取得した場合、取得した類似度から第４閾値を減算する。類似度補正部２３３６は、類似度算出部２３１５によって第３話者識別モデルを用いて類似度が算出された場合、閾値記憶部２２３６から第３閾値を読み出し、算出された類似度から第３閾値を減算する。また、類似度補正部２３３６は、類似度算出部２３１５によって第４話者識別モデルを用いて類似度が算出された場合、閾値記憶部２２３６から第４閾値を読み出し、算出された類似度から第４閾値を減算する。

30

【０３９３】

閾値記憶部２２３６は、第３話者識別モデルを用いて算出された類似度を補正するための第３閾値と、第４話者識別モデルを用いて算出された類似度を補正するための第４閾値とを予め記憶する。閾値記憶部２２３６は、話者識別モデル生成装置４６によって生成された第３閾値及び第４閾値を予め記憶する。

【０３９４】

なお、話者識別モデル生成装置４６は、閾値記憶部４２８に記憶された第３閾値及び第４閾値を話者識別装置２６へ送信してもよい。話者識別装置２６は、受信した第３閾値及び第４閾値を閾値記憶部２２３６に記憶してもよい。また、話者識別装置２６の製造時において、話者識別モデル生成装置４６によって生成された第３閾値及び第４閾値が閾値記憶部２２３６に記憶されてもよい。

40

【０３９５】

類似度判定部２３２６は、類似度補正部２３３６によって第３閾値又は第４閾値が減算された類似度が第５閾値よりも高い場合、登録音声データの話者を識別対象音声データの話者として識別する。

【０３９６】

50

続いて、本実施の形態 6 における話者識別モデル生成装置 4 6 の話者識別モデル生成処理の動作について説明する。

【0397】

図 3 8 は、本実施の形態 6 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 1 のフローチャートであり、図 3 9 は、本実施の形態 6 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 2 のフローチャートであり、図 4 0 は、本実施の形態 6 における話者識別モデル生成装置の話者識別モデル生成処理の動作について説明するための第 3 のフローチャートである。

【0398】

ステップ S 2 8 1 ～ステップ S 2 8 6 の処理は、図 3 0 に示すステップ S 2 3 1 ～ステップ S 2 3 6 の処理と同じであるので、説明を省略する。

【0399】

次に、ステップ S 2 8 7 において、第 3 話者識別部 4 2 6 は、第 3 話者識別モデル記憶部 4 2 5 から第 3 話者識別モデルを取得する。

【0400】

次に、ステップ S 2 8 8 において、第 3 話者識別部 4 2 6 は、特徴量抽出部 4 2 3 によって抽出された複数の長時間音声データの特徴量の中から、2 つの長時間音声データの特徴量を取得する。

【0401】

次に、ステップ S 2 8 9 において、第 3 話者識別部 4 2 6 は、取得した 2 つの長時間音声データの特徴量を第 3 話者識別モデルに入力することにより、2 つの長時間音声データの類似度を算出する。なお、2 つの長時間音声データは、1 人の話者によって発話された 2 つの長時間音声データと、2 人の話者によって発話された 2 つの長時間音声データとのいずれかである。このとき、2 つの長時間音声データが 1 人の話者によって発話された長時間音声データである場合の類似度は、2 つの長時間音声データが 2 人の話者によって発話された長時間音声データである場合の類似度よりも高くなる。

【0402】

次に、ステップ S 2 9 0 において、第 3 話者識別部 4 2 6 は、全ての長時間音声データの組み合わせの類似度が算出されたか否かを判定する。ここで、全ての長時間音声データの組み合わせの類似度が算出されていないと判定された場合（ステップ S 2 9 0 で N O）、ステップ S 2 8 8 に処理が戻る。そして、第 3 話者識別部 4 2 6 は、複数の長時間音声データの特徴量の中から、類似度を算出していない 2 つの長時間音声データの特徴量の特徴量抽出部 4 2 3 から取得する。

【0403】

一方、全ての長時間音声データの組み合わせの類似度が算出されたと判定された場合（ステップ S 2 9 0 で Y E S）、ステップ S 2 9 1 において、第 3 閾値算出部 4 2 7 は、第 3 話者識別部 4 2 6 によって算出された複数の類似度を回帰分析することにより、同一話者の 2 つの長時間音声データの類似度と、互いに異なる話者の 2 つの長時間音声データの類似度とを識別可能な第 3 閾値を算出する。

【0404】

次に、ステップ S 2 9 2 において、第 3 閾値算出部 4 2 7 は、算出した第 3 閾値を閾値記憶部 4 2 8 に記憶する。

【0405】

ステップ S 2 9 3 ～ステップ S 2 9 8 の処理は、図 3 0 及び図 3 1 に示すステップ S 2 3 7 ～ステップ S 2 4 2 の処理と同じであるので、説明を省略する。

【0406】

次に、ステップ S 2 9 9 において、第 4 話者識別部 4 3 6 は、第 4 話者識別モデル記憶部 4 3 5 から第 4 話者識別モデルを取得する。

【0407】

次に、ステップ S 3 0 0 において、第 4 話者識別部 4 3 6 は、特徴量抽出部 4 3 3 によって抽出された複数の短時間音声データの特徴量の中から、2 つの短時間音声データの特徴量を取得する。

【 0 4 0 8 】

次に、ステップ S 3 0 1 において、第 4 話者識別部 4 3 6 は、取得した 2 つの短時間音声データの特徴量を第 4 話者識別モデルに入力することにより、2 つの短時間音声データの類似度を算出する。なお、2 つの短時間音声データは、1 人の話者によって発話された 2 つの短時間音声データと、2 人の話者によって発話された 2 つの短時間音声データとのいずれかである。このとき、2 つの短時間音声データが 1 人の話者によって発話された短時間音声データである場合の類似度は、2 つの短時間音声データが 2 人の話者によって発話された短時間音声データである場合の類似度よりも高くなる。

10

【 0 4 0 9 】

次に、ステップ S 3 0 2 において、第 4 話者識別部 4 3 6 は、全ての短時間音声データの組み合わせの類似度が算出されたか否かを判定する。ここで、全ての短時間音声データの組み合わせの類似度が算出されていないと判定された場合（ステップ S 3 0 2 で N O）、ステップ S 3 0 0 に処理が戻る。そして、第 4 話者識別部 4 3 6 は、複数の短時間音声データの特徴量の中から、類似度を算出していない 2 つの短時間音声データの特徴量の特徴量抽出部 4 3 3 から取得する。

【 0 4 1 0 】

一方、全ての短時間音声データの組み合わせの類似度が算出されたと判定された場合（ステップ S 3 0 2 で Y E S）、ステップ S 3 0 3 において、第 4 閾値算出部 4 3 7 は、第 4 話者識別部 4 3 6 によって算出された複数の類似度を回帰分析することにより、同一話者の 2 つの短時間音声データの類似度と、互いに異なる話者の 2 つの短時間音声データの類似度とを識別可能な第 4 閾値を算出する。

20

【 0 4 1 1 】

次に、ステップ S 3 0 4 において、第 4 閾値算出部 4 3 7 は、算出した第 4 閾値を閾値記憶部 4 2 8 に記憶する。

【 0 4 1 2 】

続いて、本実施の形態 6 における話者識別装置 2 6 の話者識別処理の動作について説明する。

30

【 0 4 1 3 】

図 4 1 は、本実施の形態 6 における話者識別装置の話者識別処理の動作について説明するための第 1 のフローチャートであり、図 4 2 は、本実施の形態 6 における話者識別装置の話者識別処理の動作について説明するための第 2 のフローチャートである。

【 0 4 1 4 】

ステップ S 3 1 1 ～ステップ S 3 2 1 の処理は、図 3 4 及び図 3 5 に示すステップ S 2 6 1 ～ステップ S 2 7 1 の処理と同じであるので、説明を省略する。

【 0 4 1 5 】

次に、ステップ S 3 2 2 において、類似度補正部 2 3 3 6 は、閾値記憶部 2 2 3 6 から第 3 閾値又は第 4 閾値を取得する。このとき、モデル選択部 2 1 7 4 によって第 3 話者識別モデルが選択された場合、類似度補正部 2 3 3 6 は、閾値記憶部 2 2 3 6 から第 3 閾値を取得する。また、モデル選択部 2 1 7 4 によって第 4 話者識別モデルが選択された場合、類似度補正部 2 3 3 6 は、閾値記憶部 2 2 3 6 から第 4 閾値を取得する。

40

【 0 4 1 6 】

次に、ステップ S 3 2 3 において、類似度補正部 2 3 3 6 は、取得した第 3 閾値又は第 4 閾値を用いて、類似度算出部 2 3 1 5 によって算出された類似度を補正する。このとき、類似度補正部 2 3 3 6 は、類似度算出部 2 3 1 5 によって算出された類似度から、第 3 閾値又は第 4 閾値を減算する。

【 0 4 1 7 】

次に、ステップ S 3 2 4 において、類似度判定部 2 3 2 6 は、類似度補正部 2 3 3 6 に

50

よって補正された類似度が第5閾値より大きいかな否かを判定する。なお、第5閾値は、例えば0である。類似度判定部2326は、補正された類似度が0より大きい場合、識別対象音声データが予め登録されている登録音声データと一致すると判定し、補正された類似度が0以下である場合、識別対象音声データが予め登録されている登録音声データと一致しないと判定する。

【0418】

ここで、類似度補正部2336によって補正された類似度が第5閾値より大きいと判定された場合（ステップS324でYES）、ステップS325において、類似度判定部2326は、登録音声データの話者を識別対象音声データの話者として識別する。

【0419】

一方、類似度補正部2336によって補正された類似度が第5閾値以下であると判定された場合（ステップS324でNO）、ステップS326において、類似度判定部2326は、識別対象音声データの話者が登録音声データの話者ではないと判定する。

【0420】

ステップS327の処理は、図35に示すステップS275の処理と同じであるので、説明を省略する。

【0421】

2つの異なる第3話者識別モデルと第4話者識別モデルとが用いられる場合、第3話者識別モデルと第4話者識別モデルとの出力値の範囲が異なるおそれがある。そこで、本実施の形態6では、登録時において、第3話者識別モデル及び第4話者識別モデルのそれぞれに対し、同一の話者を識別することが可能な第3閾値及び第4閾値が算出される。また、話者識別時において、算出された識別対象音声データと登録音声データとの類似度から第3閾値又は第4閾値が減算されることにより類似度が補正される。そして、補正された類似度が第3話者識別モデル及び第4話者識別モデルに共通の第5閾値と比較されることにより、識別対象音声データの話者をより高い精度で識別することができる。

【0422】

なお、上記各実施の形態において、各構成要素は、専用のハードウェアで構成されるか、各構成要素に適したソフトウェアプログラムを実行することによって実現されてもよい。各構成要素は、CPUまたはプロセッサなどのプログラム実行部が、ハードディスクまたは半導体メモリなどの記録媒体に記録されたソフトウェアプログラムを読み出して実行することによって実現されてもよい。

【0423】

本開示の実施の形態に係る装置の機能の一部又は全ては典型的には集積回路であるLSI (Large Scale Integration) として実現される。これらは個別に1チップ化されてもよいし、一部又は全てを含むように1チップ化されてもよい。また、集積回路化はLSIに限るものではなく、専用回路又は汎用プロセッサで実現してもよい。LSI製造後にプログラムすることが可能なFPGA (Field Programmable Gate Array)、又はLSI内部の回路セルの接続や設定を再構成可能なリプログラマブル・プロセッサを利用してもよい。

【0424】

また、本開示の実施の形態に係る装置の機能の一部又は全てを、CPU等のプロセッサがプログラムを実行することにより実現してもよい。

【0425】

また、上記で用いた数字は、全て本開示を具体的に説明するために例示するものであり、本開示は例示された数字に制限されない。

【0426】

また、上記フローチャートに示す各ステップが実行される順序は、本開示を具体的に説明するために例示するためのものであり、同様の効果が得られる範囲で上記以外の順序であってもよい。また、上記ステップの一部が、他のステップと同時（並列）に実行されてもよい。

**【産業上の利用可能性】****【0427】**

本開示に係る技術は、識別対象の話者が予め登録されている話者であるか否かを識別する精度を向上させることができるので、話者を識別する技術として有用である。

10

20

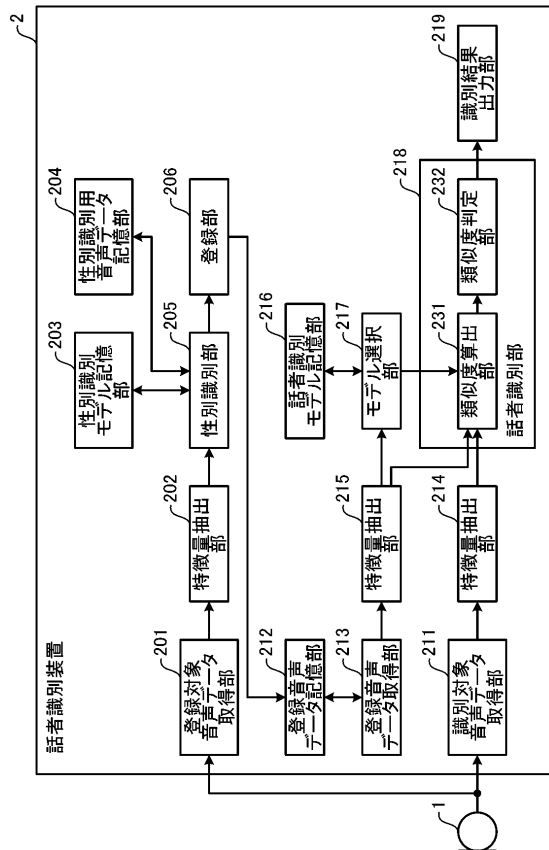
30

40

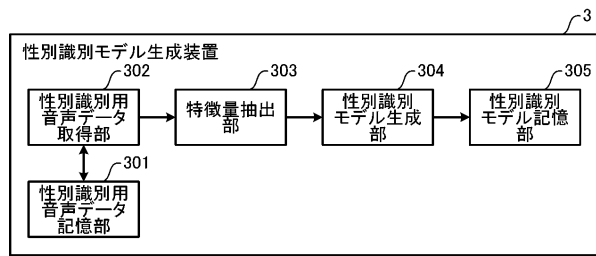
50

【図面】

【図 1】



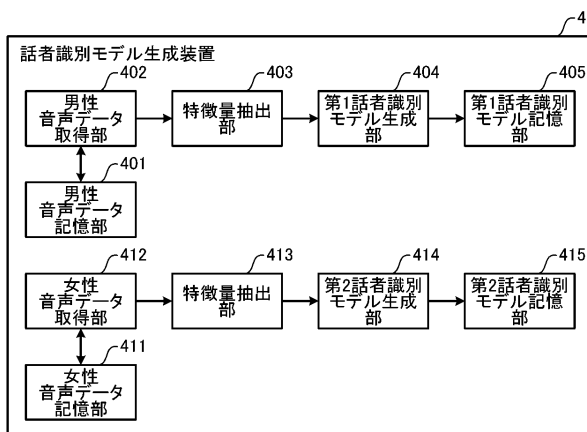
【図 2】



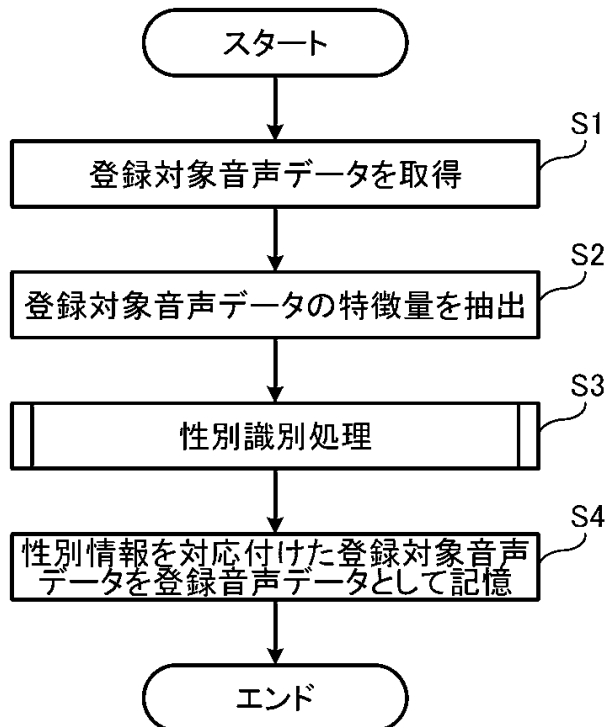
10

20

【圖 3】



【図 4】

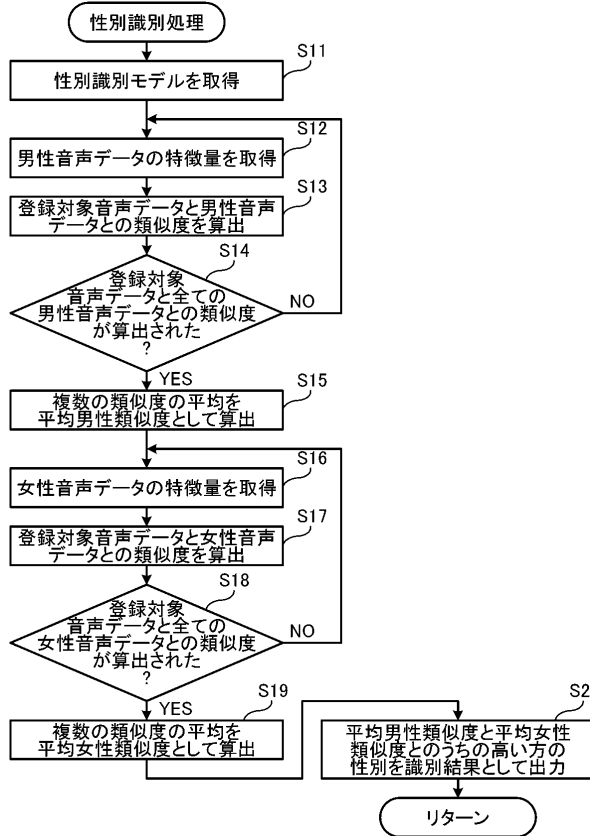


30

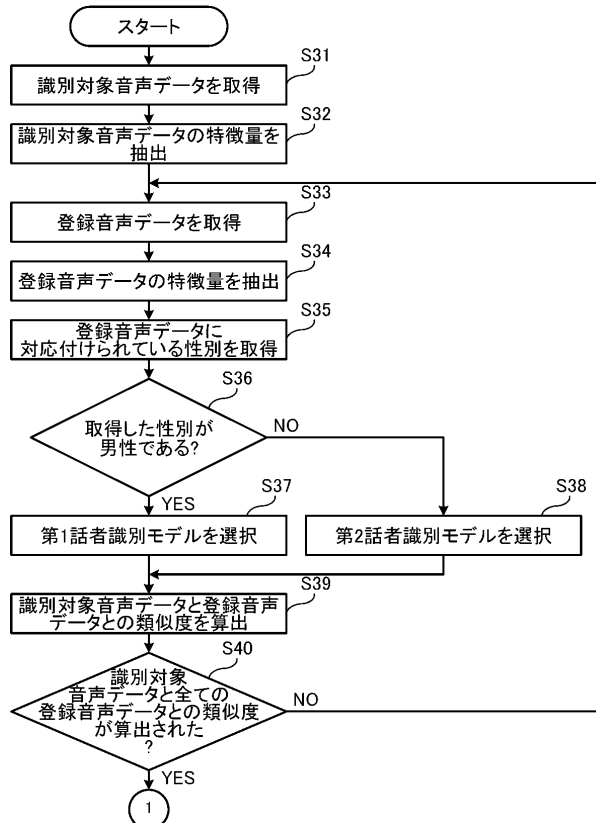
40

50

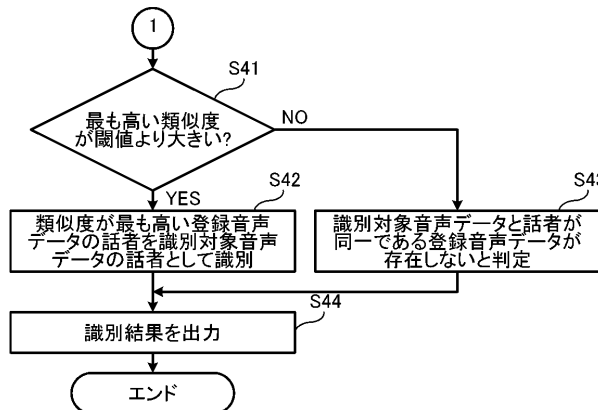
【図 5】



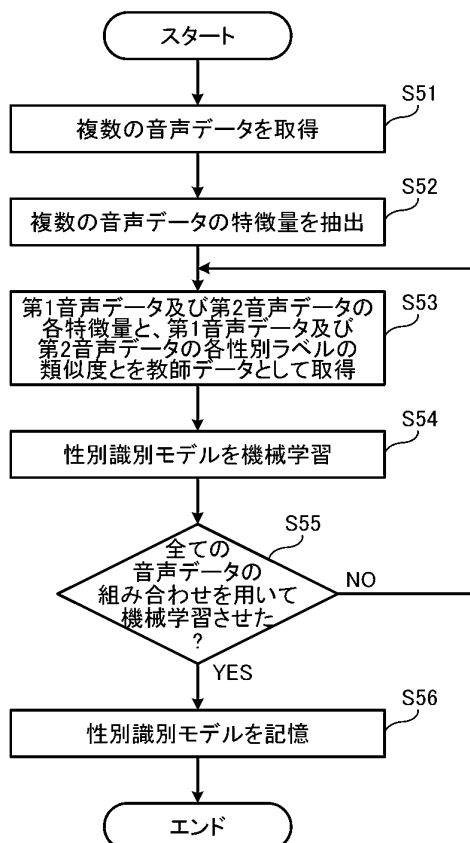
【図 6】



【図 7】



【図 8】



10

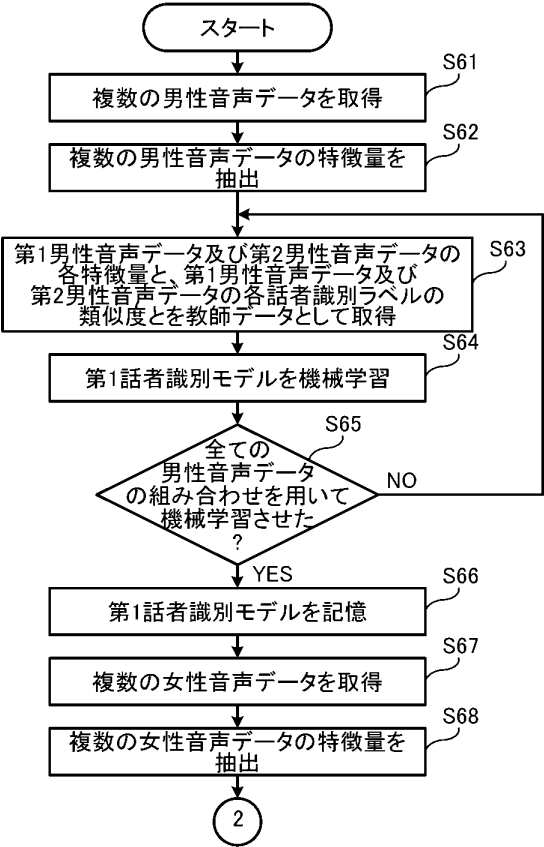
20

30

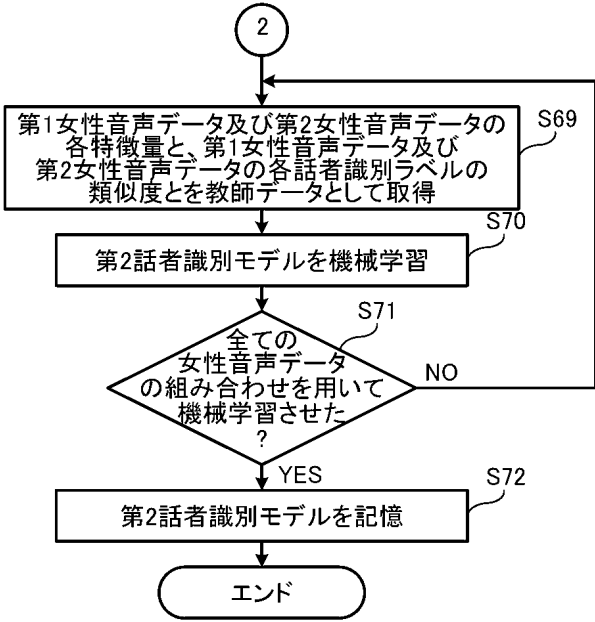
40

50

【図 9】



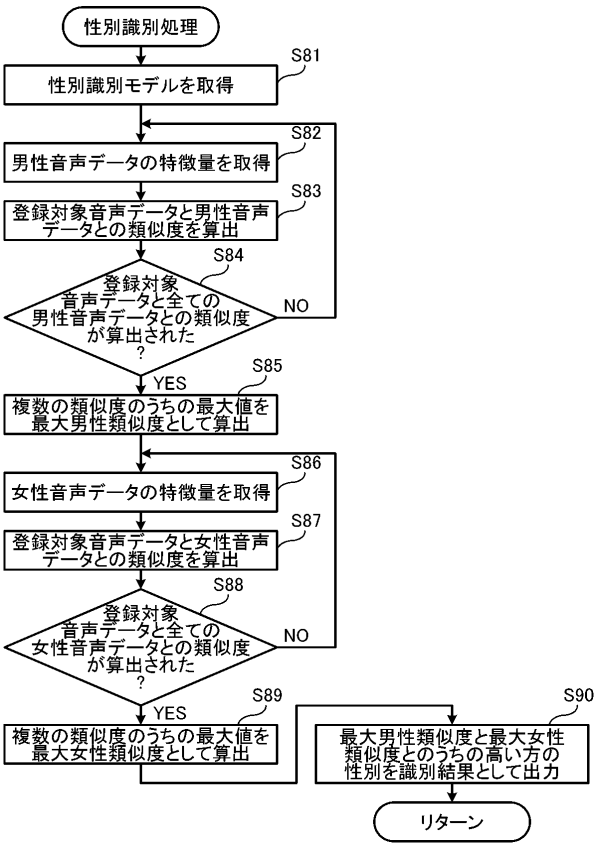
【図 10】



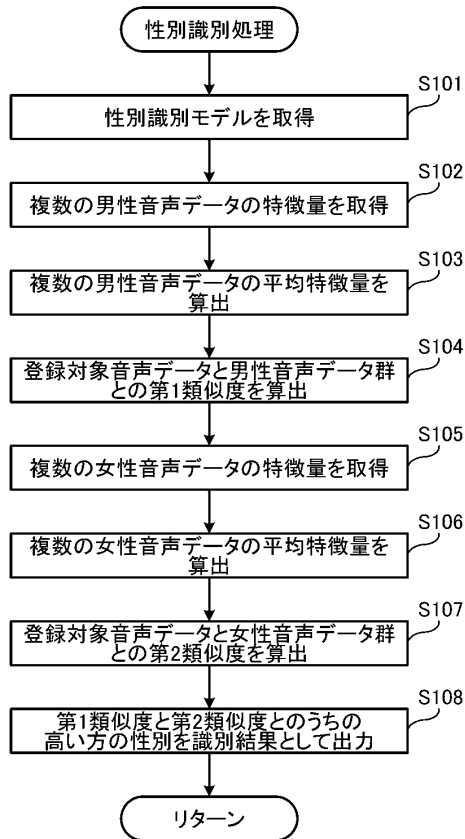
【図 11】

|                         | SRE19 evaluation dataset |         |        |  |
|-------------------------|--------------------------|---------|--------|--|
|                         | actC                     | minC    | EER(%) |  |
|                         | 0.32738                  | 0.32098 | 3.7037 |  |
| SRE19 progress deataset | actC                     | minC    | EER(%) |  |
|                         | 0.31763                  | 0.31065 | 3.7302 |  |
|                         | 0.29795                  | 0.29214 | 3.1762 |  |
| 従来の話者識別装置               |                          |         |        |  |
| 実施例1の話者識別装置             |                          |         |        |  |
| 従来の話者識別装置               |                          |         |        |  |

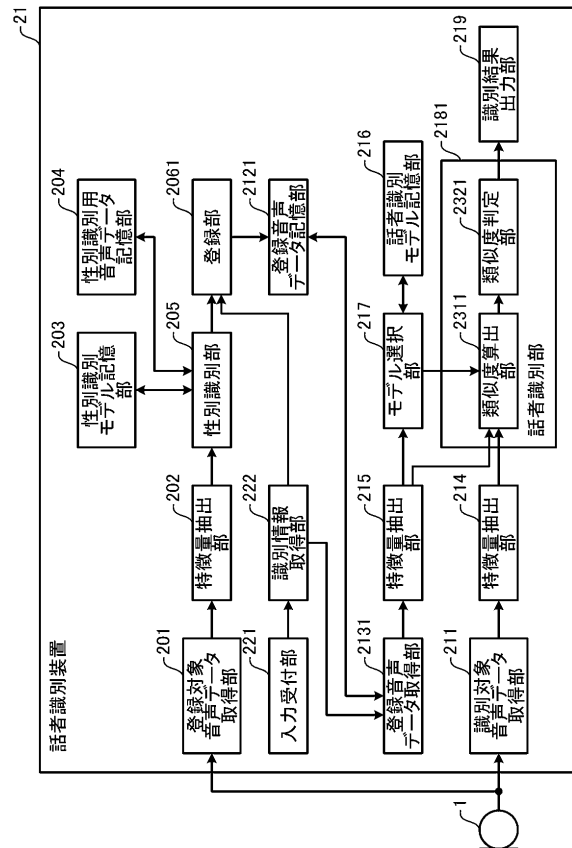
【図 12】



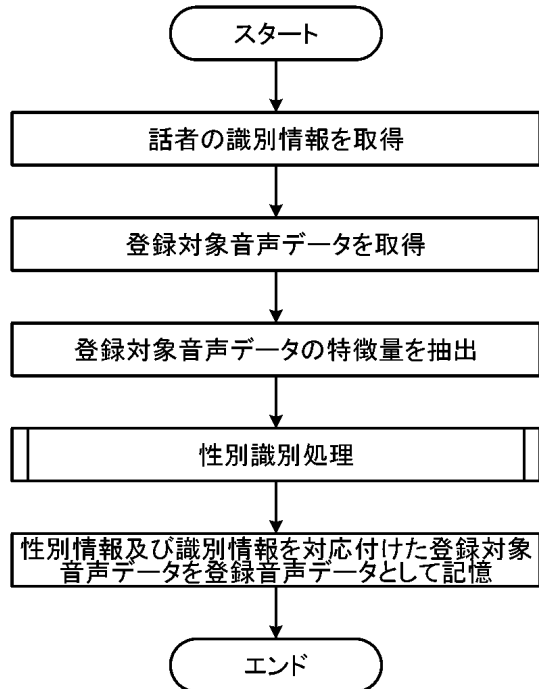
【図 1 3】



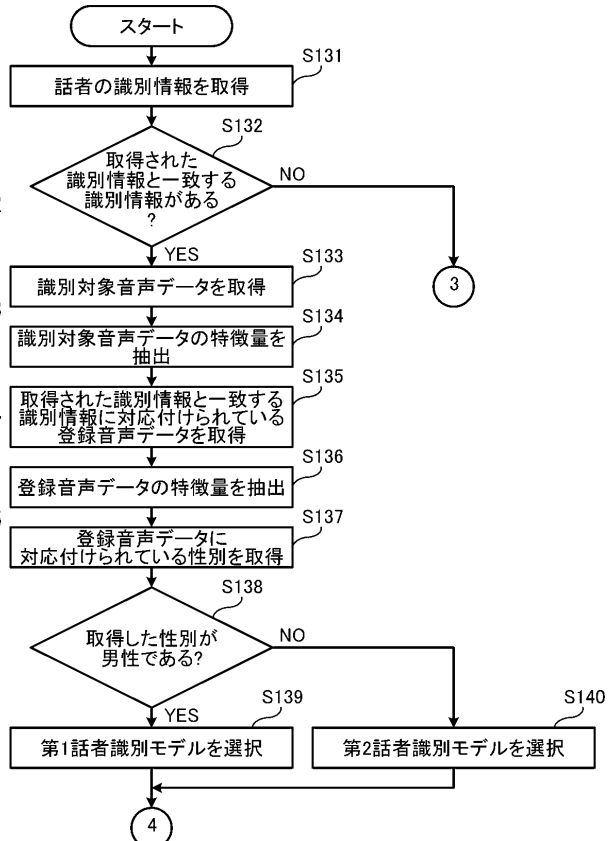
【図 1 4】



【図 1 5】



【図 1 6】



10

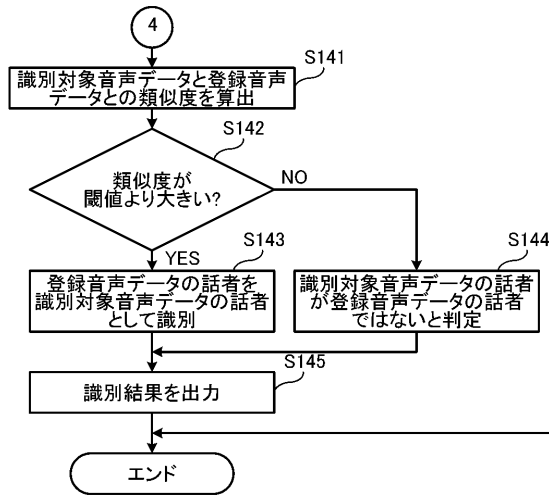
20

30

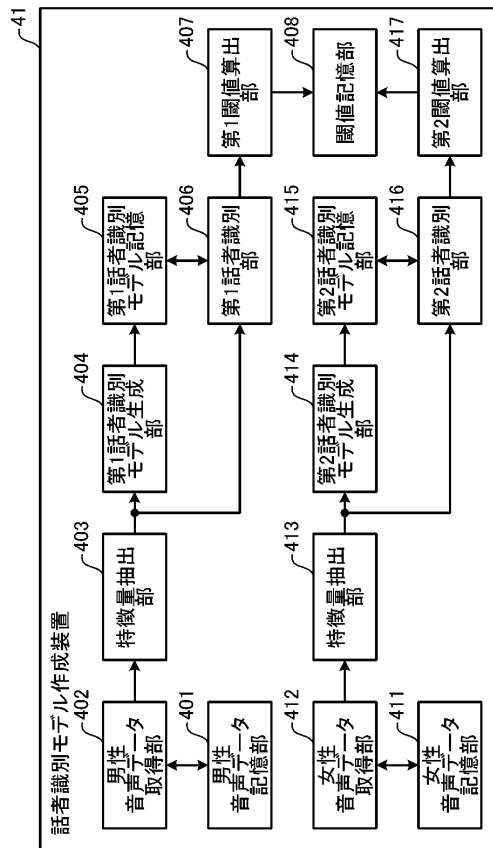
40

50

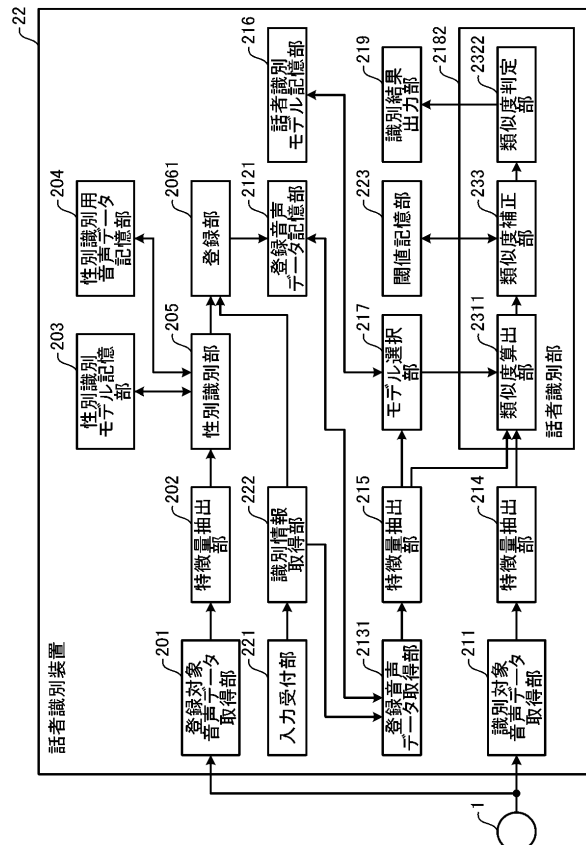
【図 17】



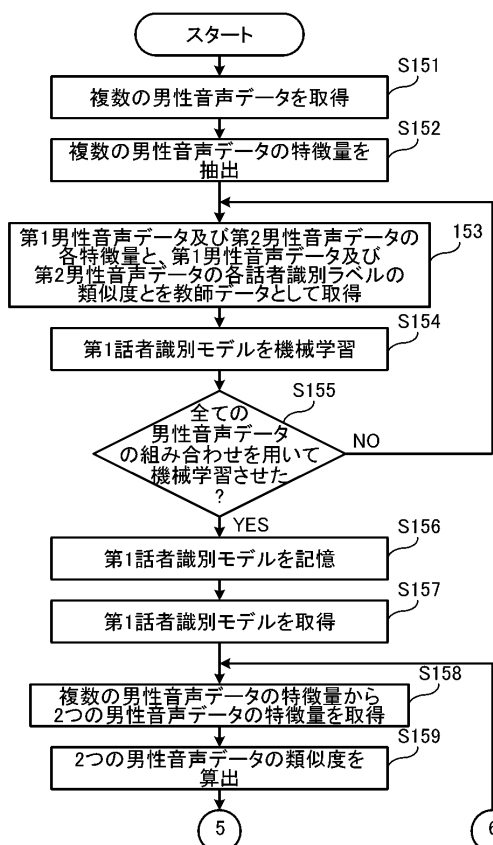
【図 18】



【図 19】



【図 20】



10

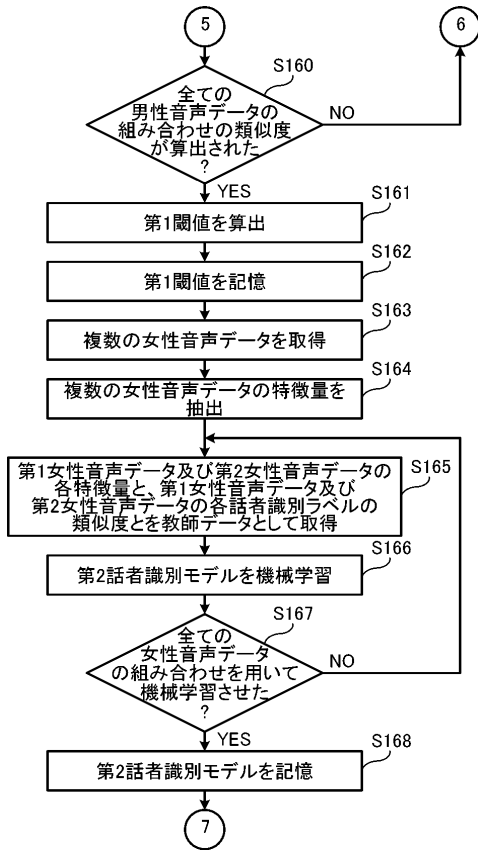
20

30

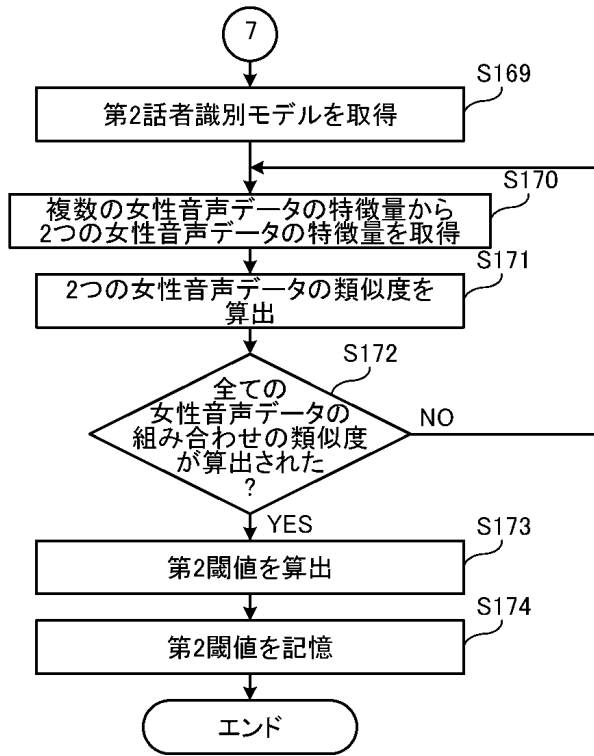
40

50

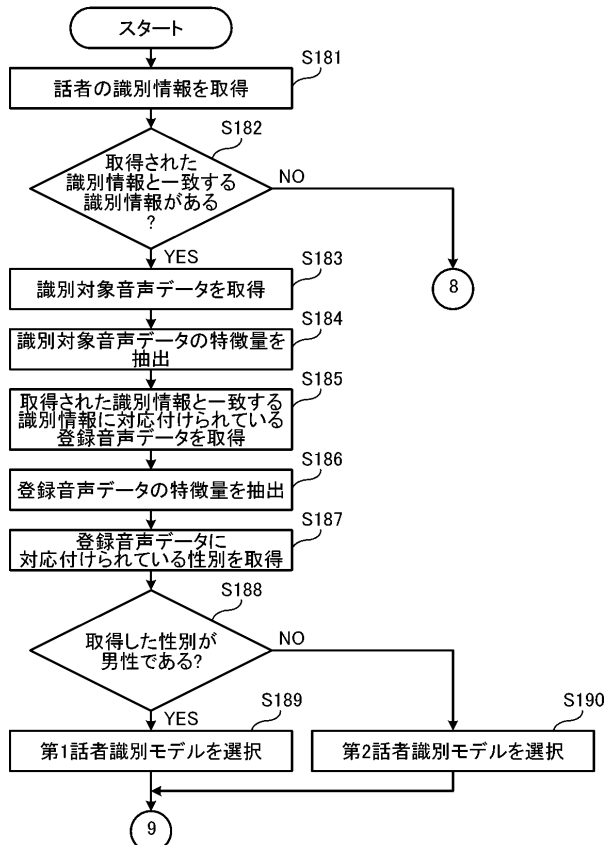
【図 2 1】



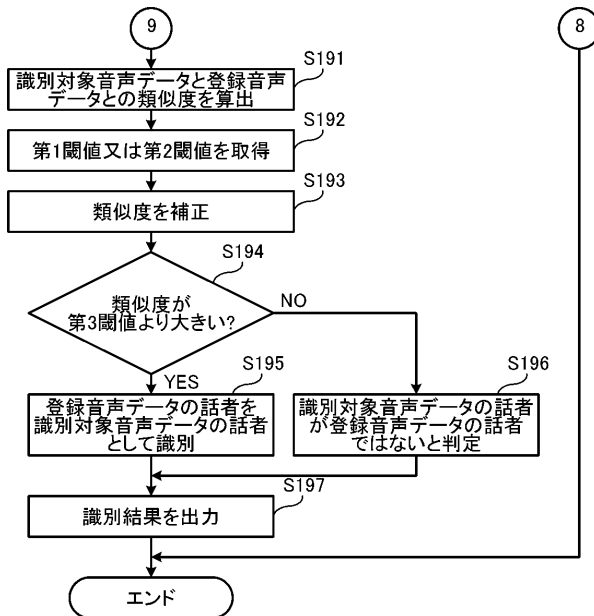
【図 2 2】



【図 2 3】



【図 2 4】



10

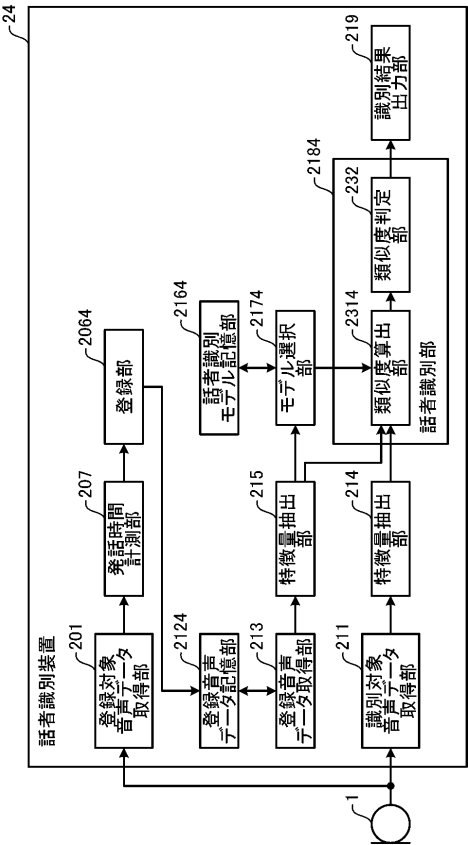
20

30

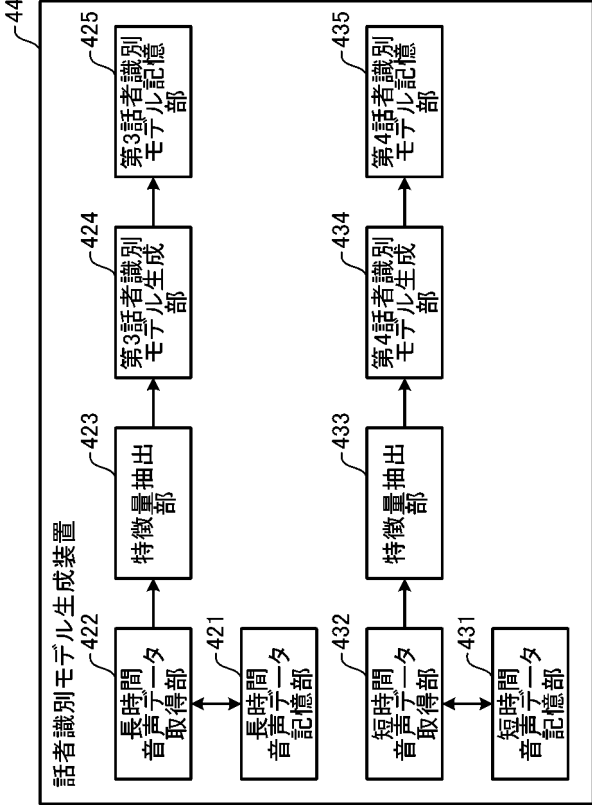
40

50

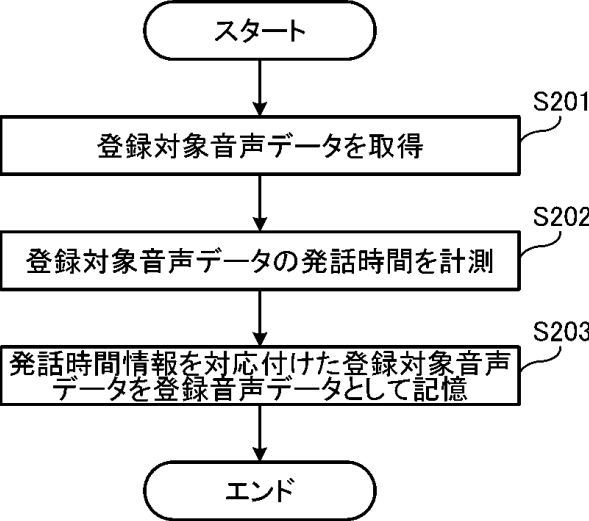
【図 25】



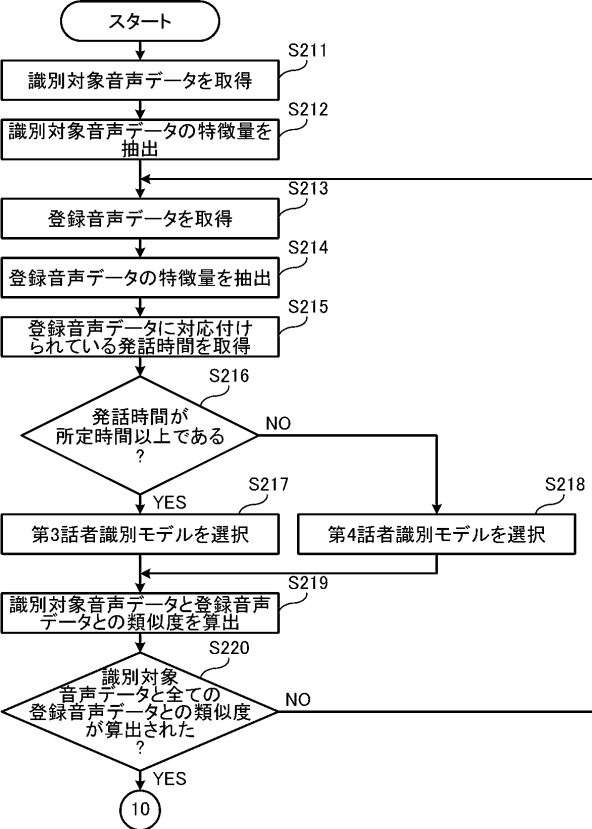
【図 26】



【図 27】



【図 28】



10

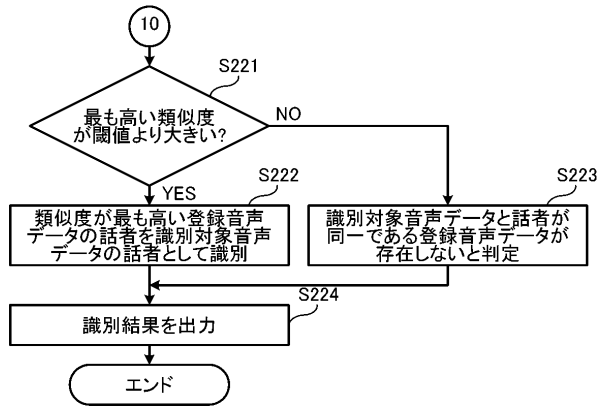
20

30

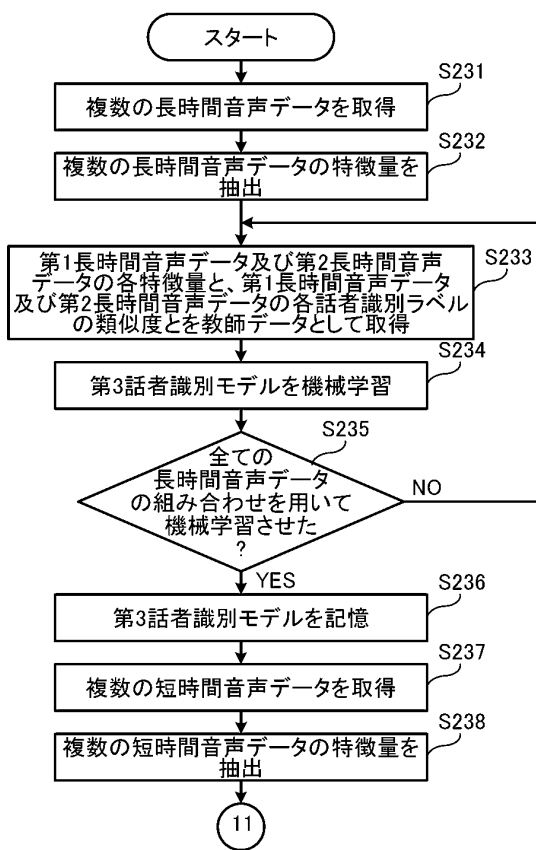
40

50

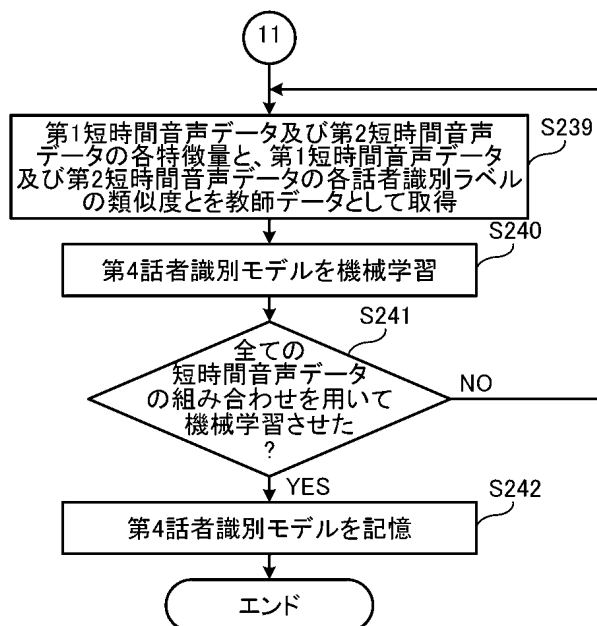
【図 29】



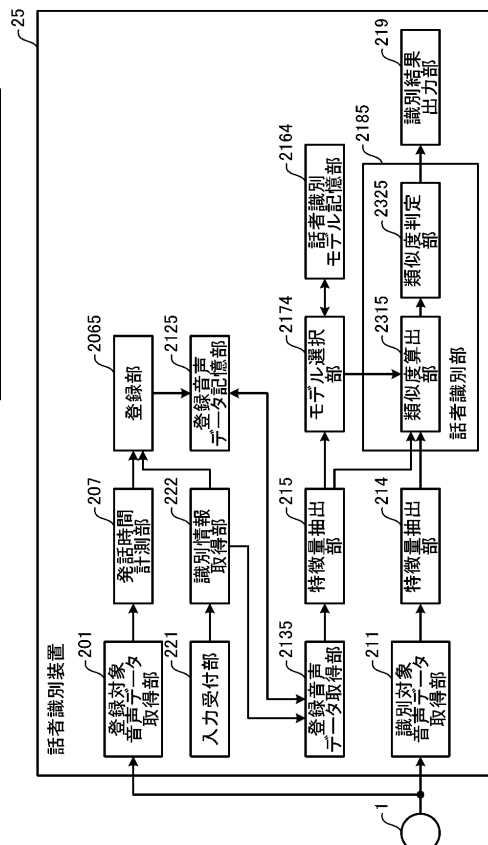
【図 30】



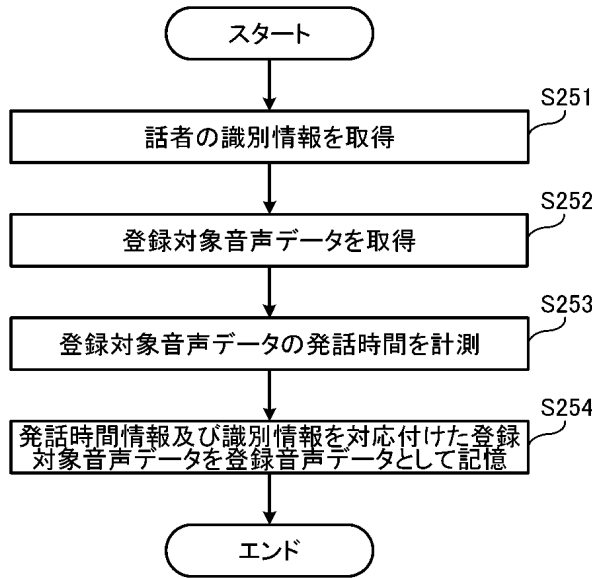
【図 31】



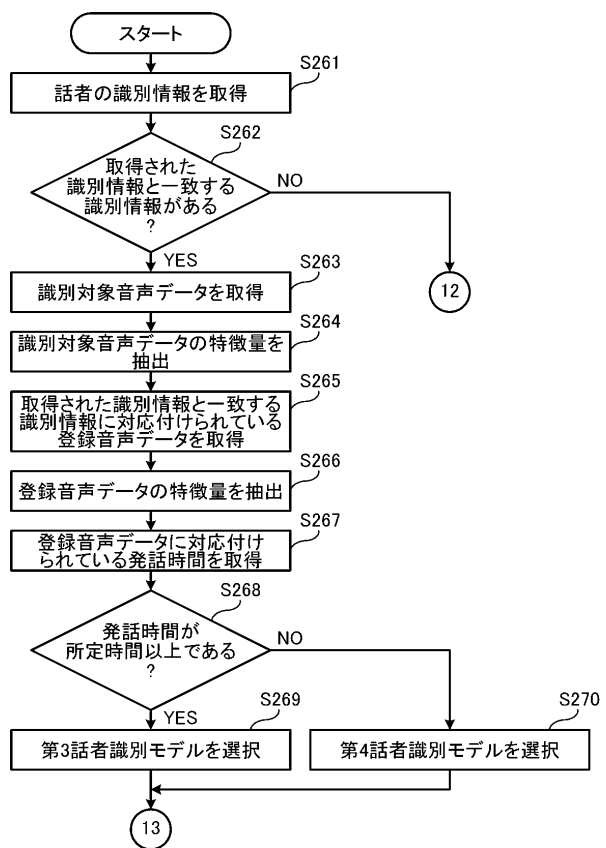
【図 32】



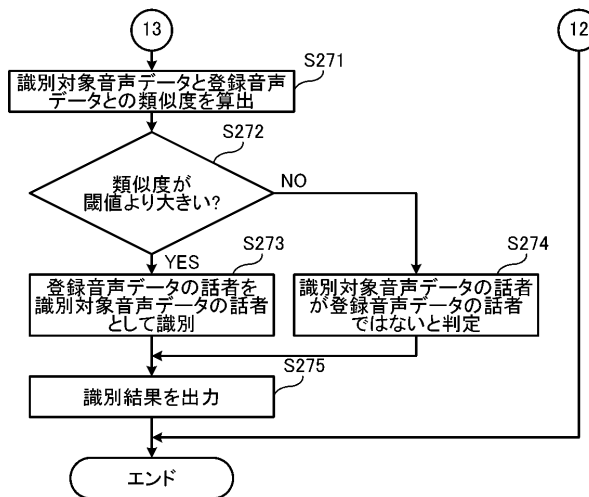
【図 3 3】



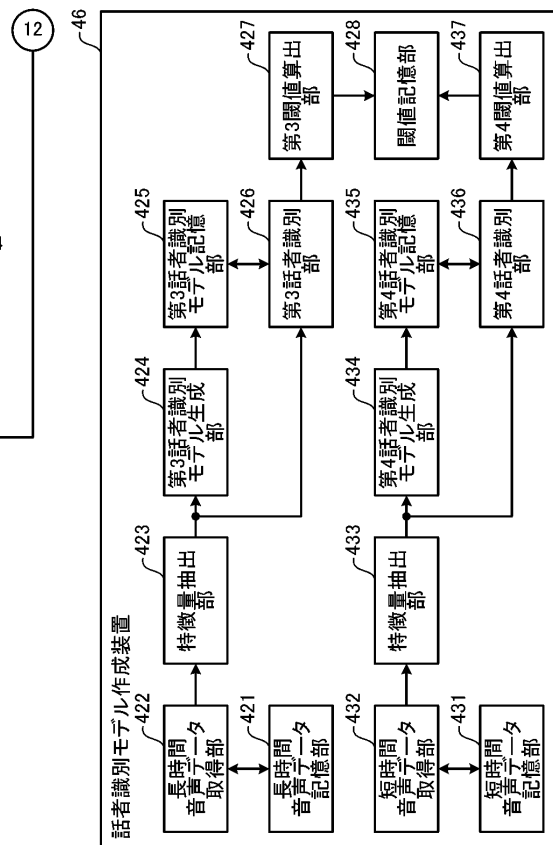
【図 3 4】



【図 3 5】



【図 3 6】



10

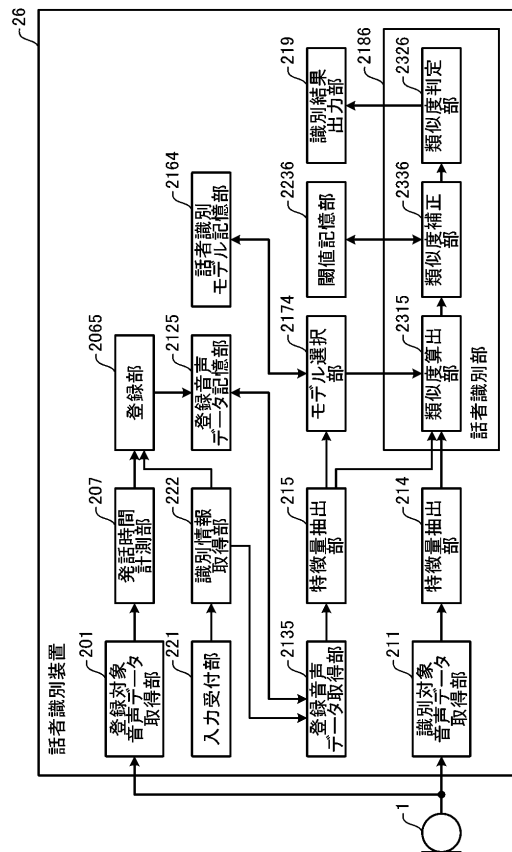
20

30

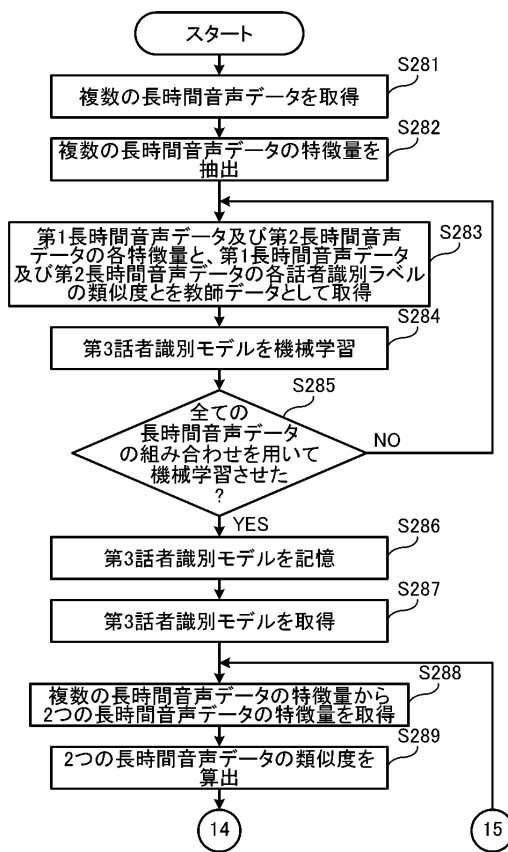
40

50

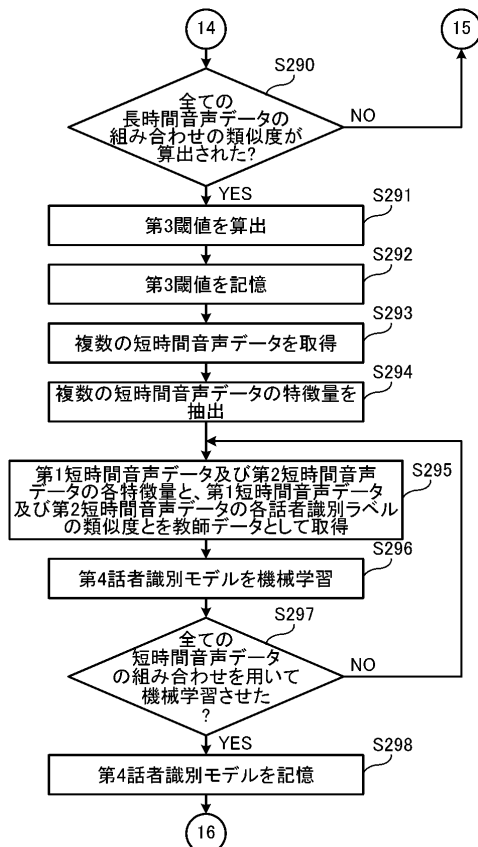
【図 37】



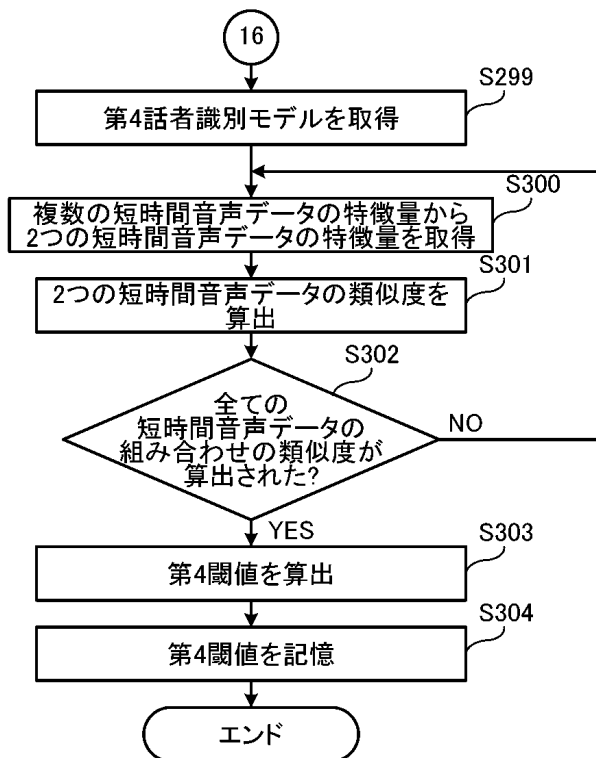
【図 38】



【図 39】



【図 40】



10

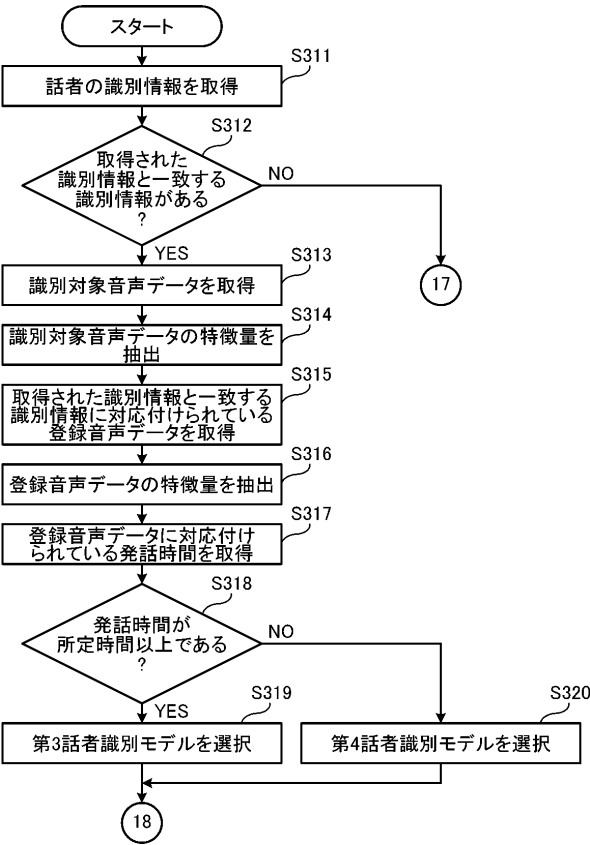
20

30

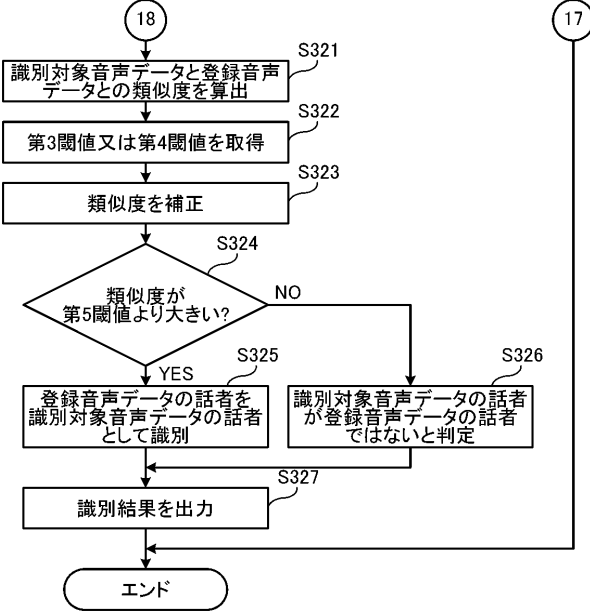
40

50

【図 4 1】



【図 4 2】



10

20

30

40

50

## フロントページの続き

ナソニックホールディングス株式会社内

審査官 中村 天真

(56)参考文献 特表 2018-508799 (JP, A)

特開 2014-048534 (JP, A)

ROHDIN, Johan et al., Robust 0-1 loss training for PLDA in speaker verification, 日本音響学会講演論文集, 2014年03月, p.101-102

RANJAN, Shivesh et al., An i-Vector PLDA based gender identification approach for severely distorted and multilingual DARPA RATS Data, IEEE Workshop on Automatic Speech Recognition and Understanding, 2015年12月, p.331-337

(58)調査した分野 (Int.Cl., DB名)

G10L 13/00-25/93

IEEE Xplore