

(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
2025年1月9日(09.01.2025)



(10) 国際公開番号

WO 2025/009127 A1

(51) 国際特許分類:

G06V 10/774 (2022.01)

(21) 国際出願番号:

PCT/JP2023/025022

(22) 国際出願日:

2023年7月5日(05.07.2023)

(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: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 水野 沙希 (MIZUNO, Saki); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 増村 亮

(MASUMURA, Ryo); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 小橋川 哲 (KOBASHIKAWA, Satoshi); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 北条 伸克 (HOJO, Nobukatsu); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).

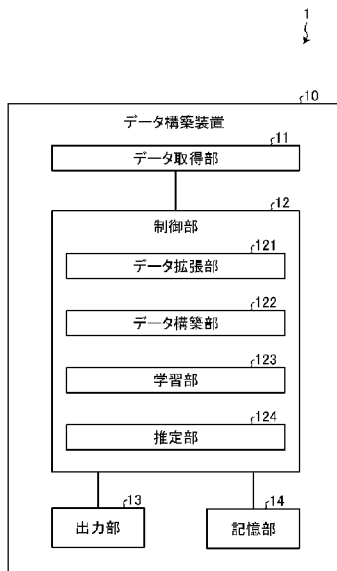
(74) 代理人: 弁理士法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎ノ門ダイビルイースト Tokyo (JP).

(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,

(54) Title: DATA CONSTRUCTION DEVICE, DATA CONSTRUCTION METHOD, AND DATA CONSTRUCTION PROGRAM

(54) 発明の名称: データ構築装置、データ構築方法、およびデータ構築プログラム

【図2】



10 Data construction device
11 Data acquisition unit
12 Control unit
13 Output unit
14 Storage unit
121 Data extension unit
122 Data construction unit
123 Training unit
124 Inference unit

(57) Abstract: A data construction device (10) comprises: a data acquisition unit (11) that acquires speech image data pertaining to speech by a first speaker and facial image data pertaining to the face of a second speaker; a data extension unit (121) that generates, on the basis of the speech image data and the facial image data acquired by the data acquisition unit (11), extended speech image data pertaining to speech by the second speaker; and a data construction unit (122) that constructs, as training data for training a learning model, a set of the extended speech image data generated by the data

[続葉有]

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, 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))

extension unit (121) and an impression label pertaining to an impression of the speech.

(57) 要約: データ構築装置 (10) は、第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得するデータ取得部 (11) と、データ取得部 (11) によって取得された発話画像データおよび顔画像データに基づいて、第二の話者による発話に関する拡張発話画像データを生成するデータ拡張部 (121) と、データ拡張部 (121) によって生成された拡張発話画像データと発話の印象に関する印象ラベルとの組を、学習モデルを学習させるための学習データとして構築するデータ構築部 (122) と、を有する。

明 細 書

発明の名称：

データ構築装置、データ構築方法、およびデータ構築プログラム

技術分野

[0001] 本発明は、データ構築装置、データ構築方法、およびデータ構築プログラムに関する。

背景技術

[0002] ニュートラル（平均的）な表情の顔画像データ、発話の音声データ、および感情の種類、強度を入力することによって、発話動画データを、顔画像データの個人性、発話の音声データに応じた口の動き、および感情の種類、強度に応じた表情の発話に関する発話画像データに変換する技術が知られている。このような技術には、例えば、複数の俳優による発話動画データを使用して、顔が変換された発話動画データを推定するための学習モデルを学習させる変換器等がある（非特許文献1参照）。

先行技術文献

非特許文献

[0003] 非特許文献1：Sanjana Sinha, Sandika Biswas, Ravindra Yadav, Brojeshwar Bhowmick, "Emotion-Controllable Generalized Talking Face Generation", arXiv:2205.01155v1 [cs.CV] 2 May 2022

発明の概要

発明が解決しようとする課題

[0004] しかし、従来技術では、自然な発話画像データへの変換を低コストですることができない場合がある。例えば、話者の感情に関する感情ラベル付きの多数の話者のそれぞれによる発話に関する発話動画データベースを用いる従来技術では、学習モデルを学習させる場合に、データベースの構築が高コストになる場合がある。一例として、従来技術では、発話の印象（例えば、「親しみやすい」、「配慮を感じる」等）のように、感情とは異なる情報に応

じた発話動画データへの変換を行うには、発話の印象に関する印象ラベルが付与された、多数の話者による発話に関する、印象ラベル付き多数話者動画データのデータベースである印象ラベル付き多数話者発話動画データベースを新たに構築する必要があるが、このようなデータベースの構築にはコストがかかる。また、コストを軽減するため、印象ラベルが付与された、少数の話者による発話に関する、印象ラベル付き少数話者動画データのデータベースである印象ラベル付き少数話者発話動画データベースを構築し、従来技術により学習モデルの学習を行う場合に、学習データに話者の顔の個人性と合ったデータが存在せず、不自然な表情の発話動画データが出力される場合がある。

課題を解決するための手段

[0005] 上述の課題を解決するため、本発明は、第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得するデータ取得部と、前記データ取得部によって取得された前記発話画像データおよび前記顔画像データに基づいて、前記第二の話者による前記発話に関する拡張発話画像データを生成するデータ拡張部と、前記データ拡張部によって生成された前記拡張発話画像データと前記発話の印象に関する印象ラベルとの組を、学習モデルを学習させるための学習データとして構築するデータ構築部と、を有する。

発明の効果

[0006] 本発明によれば、自然な発話画像データへの変換を低コストですることができる。

図面の簡単な説明

[0007] [図1]図1は、第一の実施形態に係るデータ構築システムの概要の一例について説明するための図である。

[図2]図2は、第一の実施形態に係るデータ構築システムの構成の一例を示す図である。

[図3]図3は、データ拡張の一例について説明するための図である。

[図4]図4は、データ拡張の一例について説明するための図である。

[図5]図5は、学習の一例について説明するための図である。

[図6]図6は、第一の実施形態に係るデータ構築システムによって実行される処理の流れの一例を示すフローチャートである。

[図7]図7は、第二の実施形態に係るデータ構築システムの構成の一例を示す図である。

[図8]図8は、算出の一例について説明するための図である。

[図9]図9は、算出の一例について説明するための図である。

[図10]図10は、算出の一例について説明するための図である。

[図11]図11は、学習の一例について説明するための図である。

[図12]図12は、第二の実施形態に係るデータ構築システムによって実行される処理の流れの一例を示すフローチャートである。

[図13]図13は、データ構築プログラムを実行するコンピュータの構成の一例を示す図である。

[図14]図14は、参考技術に係るデータ構築システムの概要の一例について説明するための図である。

[図15]図15は、参考技術に係るデータ構築システムの概要の一例について説明するための図である。

発明を実施するための形態

[0008] 以下、図面を用いて本発明の実施形態について説明するが、本発明は、以下の実施形態に限定されない。また、図面の記載は、同一部分には同一の符号を付しており、重複する説明は省略され、同様の機能を有する部材および同様の処理の説明も省略される。

[0009] (0. 参考技術)

(概要の一例)

本実施形態について説明する前に、図14および図15を用いて、参考技術に係るデータ構築システム1#の概要の一例について説明する。図14および図15は、参考技術に係るデータ構築システムの概要の一例について説

明するための図である。参考技術に係るデータ構築システム 1 # は、発話動画データと顔画像データとを合成し、顔画像データの個人性等に応じた表情の発話に関する発話画像データが出力される学習モデルを学習させる。ここで、発話画像データとは、第一の話者による発話に関する画像データのことをいい、静止画像データが含まれる。顔画像データとは、第二の話者に関する画像データのことをいう。

[0010] (学習の一例)

図 1 4 に示される例では、データ構築システム 1 # は、話者の感情に関する感情ラベルおよび入力発話動画データに基づいて、話者および発話内容が同一で、感情ラベルに応じて話者の顔の表情が変換された出力発話動画データを生成する。この場合、データ構築システム 1 # は、例えば、入力発話動画データおよび感情ラベルを入力とし、出力発話動画データにおける話者の顔の表情の特徴量を出力の正解データとして、教師あり学習により学習モデルを学習させる。

[0011] 図 1 5 に示される例では、データ構築システム 1 # は、感情ラベルが付与された多数の話者による発話動画データのデータベースである感情ラベル付き多数話者発話動画データベース中の感情ラベルが付与された入力発話動画データから、話者および発話内容が同一で、異なる感情ラベルに応じて話者の顔の表情が変換された出力発話動画データを生成する。ここで、多数とは、所定の数以上であれば特に限定されないが、例えば、数百人以上または数千人以上のことをいう。また、少数とは、所定の数未満であれば特に限定されないが、例えば、2、3 人のことをいう。

[0012] 例えば、データ構築システム 1 # は、入力発話動画データおよび感情ラベルを入力とし、出力発話動画データにおける話者の顔の表情の特徴量の時系列に関する顔表情特徴量系列を出力の正解データとして、教師あり学習により学習モデルを学習させる。一例として、データ構築システム 1 # における変換器（発話動画データにおける話者の表情を変換するもの）は、まず、入力発話動画データの映像データから映像特徴量を抽出し、入力発話動画デー

タの音声データから音声特徴量を抽出する。次に、変換器は、例えば、映像特徴量、音声特徴量および感情ラベルに基づいて、出力発話動画データにおける話者の顔表情特徴量系列を推定可能となるように学習モデルを学習させる。

[0013] (推定の一例)

図14に示される例では、データ構築システム1#は、学習モデルを用いて、感情ラベルおよび入力発話動画データに基づき、出力発話動画データにおける話者の顔の表情の特徴量を推定し、この特徴量の出力発話動画データを推定する。図15に示される例では、変換器は、学習モデルの学習時と同様に、入力発話動画データの映像データから映像特徴量を抽出し、音声データから音声特徴量を抽出する。例えば、変換器は、映像特徴量、音声特徴量および感情ラベルに基づいて顔表情特徴量系列を推定する。例えば、変換器は、学習モデルに入力された話者に関する入力話者画像データおよび顔表情特徴量系列に基づくrendering（レンダリング）により出力発話動画データを推定する。一例として、変換器は、顔表情特徴量系列に沿って入力話者画像データにおける話者の顔の表情が出力発話動画データの話者の表情に対応する表情となるように変換される出力発話動画データを推定する。変換器は、例えば、以下の参考文献のrendering技術を用いる。

参考文献: Aliaksandr Siarohin, Stephane Lathuiliere, Sergey Tulyakov, Elisa Ricci and Nicu Sebe, "First Order Motion Model for Image Animation", NeurIPS 2019

[0014] (課題の一例)

しかし、上述のような参考技術では、自然な発話画像データへの変換を低コストでできない場合がある。例えば、参考技術では、感情ラベルとは異なる発話の印象に関する印象ラベルに応じた発話動画データへの変換を行うために、印象ラベル付き多数話者発話動画データベースを新たに構築する必要があるが、この構築にはコストがかかる。本実施形態は、この課題を解決するためになされたものであり、自然な発話画像データへの変換を低コストで

することができる。本実施形態の詳細については後述する。

[0015] (1. 第一の実施形態)

(概要の一例)

図1を用いて、第一の実施形態に係るデータ構築システム1の概要の一例について説明する。図1は、第一の実施形態に係るデータ構築システムの概要の一例について説明するための図である。図1に示される例では、データ構築システム1は、入力発話画像データ（発話画像データ）および顔画像データを取得し、入力発話画像データおよび顔画像データに基づいて、拡張発話画像データを生成し、拡張発話画像データと印象ラベルとの組を学習データとして構築する。

[0016] 発話画像データとは、第一の話者による発話に関する画像データのことをいい、静止画像データが含まれる。顔画像データとは、第二の話者に関する画像データのことをいう。拡張発話画像データとは、第二の話者の表情が変換された発話に関する広義の発話画像データのことをいい、静止画像データが含まれる。学習データとは、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の顔の特徴量を出力とする学習モデル等の学習に用いられるデータのことをいう。

[0017] 例えば、データ構築システム1は、入力発話画像データが含まれる印象ラベル付き発話動画データにおける第一の話者の顔の表情に沿いつつ、第一の話者の顔が、顔画像データにおける第二の話者の顔となるように、顔の個人性の特徴が変換された拡張発話画像データが含まれる拡張発話動画データを生成し、拡張発話動画データと印象ラベルとの組を学習データとして構築する。

[0018] (データ構築システムの構成の一例)

図2を用いて第一の実施形態に係るデータ構築システム1の構成の一例について説明する。図2は、第一の実施形態に係るデータ構築システムの構成の一例を示す図である。図2に示される例では、データ構築システム1は、データ構築装置10を有する。

[0019] (データ構築装置)

データ構築装置 10 は、発話画像データおよび顔画像データを取得し、発話画像データおよび顔画像データに基づいて、拡張発話画像データを生成し、拡張発話画像データと印象ラベルとの組を学習データとして構築する。図 2 に示される例では、データ構築装置 10 は、データ取得部 11 と、制御部 12 と、出力部 13 と、記憶部 14 とを有する。

[0020] (データ取得部)

データ取得部 11 は、第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得する。例えば、データ取得部 11 は、発話画像データが含まれる発話動画データおよび顔画像データを取得してよい。一例として、データ取得部 11 は、所定の数未満の少数話者のそれぞれによる発話に関する少数話者発話画像データベースから発話画像データを取得し、所定の数以上の多数の話者のそれぞれの顔画像に関する多数顔画像データベースから顔画像データを取得してよい。少数話者発話画像データベースは、発話動画データを含む場合には、少数話者発話動画データベースともいう。例えば、データ取得部 11 は、少数話者発話動画データベースから発話動画データ (i) を取得し、多数顔画像データベースから顔画像データ (k) を取得する。少数話者発話画像データベースおよび多数顔画像データは低コストで構築可能なものであるため、データ取得部 11 は、これらのデータベースから画像データを取得することによって、コストを低減することができる。

[0021] 例えば、データ取得部 11 は、拡張発話画像データにおける第二の話者の顔の特徴量をさらに取得してよい。一例として、データ取得部 11 は、拡張発話動画データ (i, k) から拡張発話動画データ (i, k) の話者の顔の特徴量 (Landmark) を抽出することで拡張発話動画データ (i, k) の話者の顔の Landmark を取得する。i は、発話動画データの話者の番号に対応する数である。k は、顔画像データの話者の番号に対応する数である。

[0022] (制御部)

制御部 12 は、データ構築装置 10 全体を制御する。例えば、制御部 12 は、各種処理手順が規定されたプログラムや、制御情報が格納された内部メモリを有する一以上のプロセッサによって構成され、プロセッサは、プログラムや内部メモリを用いて各処理を実行する。制御部 12 は、例えば、CPU (Central Processing Unit)、MPU (Micro Processing Unit)、GPU (Graphics Processing Unit) 等の電子回路やASIC (Application Specific Integrated Circuit)、FPGA (Field Programmable Gate Array) 等の集積回路により実現される。図 2 に示される例では、制御部 12 は、データ拡張部 121 と、データ構築部 122 と、学習部 123 と、推定部 124 とを有する。

[0023] (データ拡張部)

データ拡張部 121 は、データ取得部 11 によって取得された発話画像データおよび顔画像データに基づいて、第二の話者による発話に関する拡張発話画像データを生成する。例えば、データ拡張部 121 は、データ取得部 11 によって取得された発話動画データおよび顔画像データに基づいて、拡張発話画像データが含まれる拡張発話動画データを生成してよい。データ拡張部 121 は、例えば、話者の顔の表情の特徴量を変換する変換器等によって実現される。以下、図 3 および図 4 を用いて、データ拡張部 121 によるデータ拡張の一例について説明する。図 3 および図 4 は、データ拡張の一例について説明するための図である。

[0024] 図 3 に示される例では、データ拡張部 121 は、印象ラベル付き発話動画データベース中の印象ラベル付き発話動画データ (i) および顔画像データ (k) に基づいて、拡張発話動画データ (i, k) を生成する。一例として、データ拡張部 121 は、発話動画データ (i) における話者の顔が顔画像データ (k) における話者の顔に変換された表情となるように顔の表情の特徴量を変換された、顔画像データ (k) における話者による発話に関する拡張発話動画データ (i, k) を生成する。

[0025] 図 4 に示される例では、データ拡張部 121 は、少数の話者による発話動

画データ（ i ）が含まれる少数話者発話動画データベースと、多数の顔画像データ（ k ）が含まれる多数顔画像データベースとを用いて、発話動画データのデータ拡張を行う。例えば、データ拡張部 121 は、少数話者発話動画データベース中の発話動画データ（ i ）と、多人数顔画像データベース中の顔画像データ（ k ）とに基づいて、発話動画データ（ i ）における話者の表情および顔画像データ（ k ）における話者の個人性の特徴を持つ拡張発話動画データ（ i, k ）を生成する。一例として、データ拡張部 121 は、印象ラベル付き少数話者の発話動画データベース中の発話動画データ（ i ）の各フレーム画像と、多数顔画像データベース中の各顔画像データ（ k ）との全ての組み合わせに対応する拡張発話動画データ（ i, k ）の各フレーム画像を生成し、発話動画データ（ i ）における話者の表情および顔画像データ（ k ）における話者の個人性の特徴を持つ拡張発話動画データ（ i, k ）を生成する。例えば、データ拡張部 121 は、以下の参考文献 2 のような顔表情変換技術を使用して拡張発話動画データ（ i, k ）を生成する。また、データ拡張部 121 は各 i, k について上述の処理を実施してよい。

参考文献 2 : Sirui Zhao, Shukang Yin, Huaying Tang, Rijin Jin, Yifan Xu, Tong Xu, Enhong Chen “Fine-grained Micro-Expression Generation based on Thin-Plate Spline and Relative AU Constraint” ,MM '22: Proceedings of the 30th ACM International Conference on Multimedia, October 2022, Pages 7150–7154

[0026] （データ構築部）

データ構築部 122 は、データ拡張部 121 によって生成された拡張発話画像データと発話の印象に関する印象ラベルとの組を、学習モデルを学習させるための学習データとして構築する。データ構築部 122 は、多数の話者の自然な表情の発話動画データを学習データとして低コストで構築できるため、自然な発話画像データへの変換を低コストですることができる。例えば、データ構築部 122 は、データ拡張部 121 によって生成された拡張発話動画データと印象ラベルとの組を学習データとして構築してよい。

[0027] 例えば、図4に示されるように、データ構築部122は、印象ラベル付き拡張発話動画データ(i, k)が含まれる印象ラベル付き拡張発話動画データベースを学習データ用のデータベースとして構築する。一例として、データ構築部122は、拡張発話動画データ(i, k)と発話の印象に関する印象ラベルとの組(印象ラベル付き拡張発話動画データ)を印象ラベル付き拡張発話動画データベースに追加する。これにより、データベースは、多数の話者の自然な表情の発話動画データを含むことになる。データ構築部122は、多数の話者を含むデータによって学習モデルを学習させることにより、入力話者の顔の個人性と合った自然な表情となるように拡張発話動画データにおける話者の表情を変換することができる。少数話者発話動画データベースおよび多数顔画像データは低コストで構築可能であるため、データ構築部122は、自然な発話画像データへの変換を低コストですることができる。

[0028] (学習部)

学習部123は、データ構築部122によって構築された学習データ、発話画像データおよび顔画像データによって、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の顔の特徴量を出力とする学習モデルを学習させる。このように、学習部123は、印象ラベル付き拡張発話動画データを学習データとして用いて学習モデルを学習させる点を除き、参考技術と同様に学習モデルを学習させる。これにより、学習部123は、学習モデルに入力される顔画像データの話者(入力話者)の顔の個人性と合った自然な表情となるように拡張発話画像データにおける話者の表情の変換を実現し、自然な発話画像データへの変換を低コストですることができる。

[0029] 以下、図5を用いて、学習部123による学習の一例について説明する。図5は、学習の一例を示す図である。例えば、図5に示されるように、学習部123は、データ構築部122によって構築された印象ラベル付き拡張発話動画データ、発話動画データおよび顔画像データにより印象ラベル、発話動画データおよび顔画像データを入力とし、拡張発話動画データにおける第

二の話者の顔の特徴量を出力とする学習モデルを学習させてよい。

[0030] 一例として、学習部123は、印象ラベル付き拡張発話動画データ(i, k)、発話動画データ(i)および顔画像データ(k)によって、印象ラベル付き発話動画データ(i)および顔画像データ(k)を入力とし、拡張発話動画データ(i, k)における話者の顔のLandmarkを出力とする学習モデルを学習させる。これにより、学習部123は、入力話者の顔の個人性と合った自然な表情(印象変換用顔表情特徴量)に話者の表情(顔表情特徴量)が変換された自然な発話動画データへの変換を低コストで可能なように学習モデルを学習させることができる。

[0031] (推定部124)

推定部124は、学習部123によって学習された学習モデルに基づいて、拡張発話画像データにおける第二の話者の顔の特徴量を推定する。例えば、図5に示されるように、推定部124は、学習モデルに基づいて印象ラベル付き拡張発話動画データ(i, k)の各時刻における話者の顔のLandmarkの座標の座標(変換後の顔表情特徴量)を推定し、各時刻における話者の顔のLandmark(印象変換用顔表情特徴量)を推定する。また、例えば、推定部124は、rendering(レンダリング)により、印象変換用顔表情特徴量が反映された印象ラベル付き拡張発話動画データ(i, k)に対応する出力発話動画データを推定する。このように、図5に示される例では、推定部124は、拡張発話動画データにおける第二の話者の顔の特徴量および拡張発話動画データに対応する出力発話動画データを推定対象とする点を除き、参考技術と同様に推定する。推定部124は、例えば、話者の顔の表情の特徴量を変換する変換器等によって実現される。

[0032] (出力部)

出力部13は、各種データを出力する。例えば、出力部13は、推定部124によって推定された印象ラベル付き拡張発話動画データ(i, k)に対応する出力発話動画データを出力する。出力部13は、例えば、ディスプレイといった表示部等により実現される。

[0033] (記憶部)

記憶部 14 は、各種データを記憶する。例えば、記憶部 14 は、印象ラベル付き少数話者発話動画データベース、多数顔画像データベース、学習モデル、および、データ構築装置 10 によって実行される OS (Operating System) や各種プログラム等を記憶する。記憶部 14 は、例えば、HDD (Hard Disk Drive)、SSD (Solid State Drive)、光ディスク等の記憶装置や RAM (Random Access Memory)、フラッシュメモリ、NVS RAM (Non Volatile Static Random Access Memory) 等の情報を書き換え可能な半導体メモリ等によって実現される。

[0034] (処理の流れの一例)

以下、図 6 を用いて第一の実施形態に係るデータ構築システム 1 のデータ構築装置 10 によって実行される処理の流れの一例について説明する。図 6 は、第一の実施形態に係るデータ構築システムによって実行される処理の流れの一例を示すフローチャートである。

[0035] ステップ S 11 において、データ取得部 11 は、第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得する。例えば、データ取得部 11 は、学習モデルを学習させる場合に、少数話者発話動画データベースから発話動画データ (i) を取得し、多数顔画像データベースから顔画像データ (k) を取得する。

[0036] ステップ S 12 において、データ拡張部 121 は、データ取得部 11 によって取得された発話画像データおよび顔画像データに基づいて、第二の話者による発話に関する拡張発話画像データを生成する。例えば、データ拡張部 121 は、発話動画データ (i) における話者の顔の表情が顔画像データ (k) における話者の顔の表情に変換された、顔画像データの話者による発話に関する拡張発話動画データ (i, k) を生成する。

[0037] ステップ S 13 において、データ構築部 122 は、データ拡張部 121 によって生成された拡張発話画像データと発話の印象に関する印象ラベルとの組を、学習モデルを学習させるための学習データとして構築する。例えば、

データ構築部 122 は、印象ラベル付き拡張発話動画データ (i, k) が含まれる印象ラベル付き拡張発話動画データベースを学習データ用のデータベースとして構築する。

[0038] ステップ S14 において、学習部 123 は、データ構築部 122 によって構築された学習データ、発話画像データおよび顔画像データによって、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の顔の特徴量を出力とする学習モデルを学習させる。例えば、学習部 123 は、印象ラベル付き拡張発話動画データ (i, k)、発話動画データ (i) および顔画像データ (k) によって、印象ラベル付き発話動画データ (i) および顔画像データ (k) を入力とし、拡張発話動画データ (i, k) における話者の顔のLandmarkを出力とする学習モデルを学習させる。

[0039] ステップ S15 において、推定部 124 は、学習部 123 によって学習された学習モデルに基づいて、拡張発話画像データにおける第二の話者の顔の特徴量を推定する。例えば、推定部 124 は、学習モデルに基づいて印象ラベル付き拡張発話動画データ (i, k) の各時刻における話者の顔のLandmarkの座標を推定し、各時刻における当該話者の顔のLandmarkが反映された印象ラベル付き拡張発話動画データ (i, k) に対応する出力発話動画データを推定する。

[0040] (2. 第二の実施形態)

(概要の一例)

データ構築システム 1 は、例えば、学習モデルに基づいて各時刻における話者の顔のLandmarkの座標を推定し、各時刻における当該話者の顔のLandmarkが反映された印象ラベル付き拡張発話動画データ (i, k) に対応する出力発話動画データを推定する。このLandmarkの位置には、顔の個人性（平均的なLandmarkの位置）および表情の変化（平均的なLandmarkの位置からの差分）という 2 種類の情報が含まれるため、上述の例では、データ構築システム 1 は、2 種類の異なる情報を同時に推定している。

[0041] これに対し、第二の実施形態に係るデータ構築システム 1 X は、学習データ、印象ラベル、発話画像データおよび顔画像データによって、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の表情の変化の特徴量を出力とする学習モデルを学習させる。例えば、データ構築システム 1 X は、学習モデルの学習時に、話者毎に話者の顔のLandmarkの平均値を 0、分散を 1 とする正規化により、顔の個人性の情報を削除し、表情の変化のみが表現されたベクトルを作成する。上述の例では、データ構築システム 1 X は、表情の変化のみが表現されたベクトル等の拡張発話画像データにおける第二の話者の表情の変化の特徴量を推定する。このため、データ構築システム 1 X は、データ構築システム 1 と比べて、学習モデルに対する入出力の関係を単純化し、学習モデルの学習を容易とし、変換精度を改善し、少ない学習データで個人性に合った顔の表情となるように、発話動画データと顔画像データとの合成を可能とする。

[0042] (データ構築システムの構成の一例)

図 7 を用いて第二の実施形態に係るデータ構築システム 1 X の構成の一例について説明する。図 7 は、第二の実施形態に係るデータ構築システムの構成の一例を示す図である。図 7 に示される例では、データ構築システム 1 X は、データ構築装置 10 X を有する。

[0043] (データ構築装置)

データ構築装置 10 X は、学習データ、印象ラベル、発話画像データおよび顔画像データによって、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の表情の変化の特徴量を出力とする学習モデルを学習させる。データ構築装置 10 X は、第一の実施形態における制御部 12 の代わりに、制御部 12 X を有する。この点を除き、データ構築装置 10 X は、第一の実施形態に係るデータ構築装置 10 と同様である。

[0044] (制御部)

制御部 12 X は、算出部 125 をさらに有し、第一の実施形態における学

習部 1 2 3 と、推定部 1 2 4 との代わりに、学習部 1 2 3 X と、推定部 1 2 4 X とを有する。この点を除き、制御部 1 2 X は、第一の実施形態における制御部 1 2 と同様である。

[0045] (算出部)

算出部 1 2 5 は、データ構築部 1 2 2 によって構築された学習データに基づいて、拡張発話画像データにおける第二の話者の表情の変化の特徴量を算出する。以下、図 8～図 10 を用いて、算出部 1 2 5 による算出の一例について説明する。図 8～図 10 は、算出の一例について説明するための図である。

[0046] 図 8 に示される例では、算出部 1 2 5 は印象ラベル付き拡張発話動画データベースに含まれる各話者の拡張発話動画データから各話者の顔の Landmark を抽出する。この印象ラベル付き拡張発話動画データベースには各話者の印象ラベル付き発話動画データベースが含まれ、算出部 1 2 5 は、例えば、各話者の印象ラベル付き拡張発話動画データベースにおける各話者の印象ラベル付き拡張発話動画データから各話者の顔の Landmark を抽出する。

[0047] 次に、図 8 に示される例では、算出部 1 2 5 は、各話者に対応する全データに基づいて、各話者の顔の Landmark の座標の平均値および分散に基づいて、主に各話者の個人性の情報が含まれる各話者の顔の Landmark 統計量を算出する。例えば、算出部 1 2 5 は、各話者の様々な表情が含まれるデータに基づいて各話者の顔の Landmark 統計量を算出するため、各話者の拡張発話動画データにおける各話者の顔の特徴量の時系列に関する各話者の顔の Landmark 系列に含まれる情報のうち、表情の情報が平均化され、主に個人性の情報が含まれる各話者の顔の Landmark を算出することができる。

[0048] 図 9 に示される例では、算出部 1 2 5 は、印象ラベル付き拡張発話動画データ (i, k) から、 $t = 1 \sim t = T$ までの k 番目の話者の Landmark $(t = 1)$ 、Landmark $(t = 2)$ 、 $\dots$ 、Landmark $(t = T - 1)$ および Landmark $(t = T)$ を、 k 番目の話者の顔の特徴量として抽出する。 t は、印象ラベル付き拡張発話動画データ (i, k) の各時刻に対応するフレーム番号である。 T

は、印象ラベル付き拡張発話動画データ（ i, k ）の最後のフレーム番号である。次に、図9に示される例では、算出部125は、 k 番目の話者の顔のLandmarkが含まれる各話者に対応する全データである、各時刻に対応するフレーム番号のLandmark特徴量に基づいて、 k 番目の話者の顔のLandmarkの座標の平均値および分散を算出する。次に、図9に示される例では、算出部125は、 k 番目の話者の顔のLandmarkの平均値および分散を k 番目の話者の顔のLandmark統計量として算出する。例えば、算出部125は、 k 番目の話者の顔のLandmarkの座標の平均値および分散を k 番目の話者の顔のLandmark統計量として算出する。算出部125は、例えば、上述の例と同様に、各話者の顔のLandmark統計量を算出する。

[0049] 次に、図10に示されるように、算出部125は、各話者の顔のLandmarkの平均値および分散に基づいて各話者の顔のLandmarkを正規化する。例えば、算出部125は、各話者の顔のLandmarkの平均値を0、分散を1とするように各話者のLandmark統計量を標準正規化し、正規化された値を各話者の表情変化Landmarkとして算出する。

[0050] （学習部）

学習部123Xは、データ構築部122によって構築された学習データ、印象ラベル、発話画像データおよび顔画像データによって、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の表情の変化の特徴量を出力とする学習モデルを学習させる。このように、学習部123Xは、表情変化Landmarkを推定対象とすることにより、学習モデルに対する入出力の関係を単純化し、学習モデルを容易に学習させ、変換精度を改善できる。また、学習部123Xは、少ない学習データで個人性に合った顔の表情となるように発話動画データと顔画像データとを合成できる。

[0051] 例えば、学習部123Xは、第二の話者の顔の特徴量と第二の話者の平均的な顔の特徴量との差分を第二の話者の表情の変化の特徴量として出力する学習モデルを学習させてよい。一例として、学習部123Xは、算出部12

5によって標準正規化された各話者の顔のLandmarkの平均値および分散を、各話者の表情変化Landmark系列として学習モデルを学習させる。これにより、学習部123Xは、顔の個人性の情報が削除され、表情の変化のみが表現された顔の特徴量を、拡張発話画像データにおける第二の話者の表情の変化の特徴量として推定可能な学習モデルとなるように、学習モデルを学習させることができる。

[0052] 以下、図11を用いて、学習部123Xによる学習の一例について説明する。図11は、学習の一例について説明するための図である。図11に示される例では、学習部123Xは、各話者の顔の表情の特徴量に関する表情変化Landmarkや各話者の顔の表情の特徴量の時系列に関する表情変化Landmark系列を推定対象とする学習モデルを学習させる点を除き、第一の実施形態における学習部123と同様に学習モデルを学習させる。

[0053] 図11に示される例では、学習部123Xは、算出部125によって算出された表情変化Landmark系列、印象ラベル、入力発話動画データ、入力顔画像データおよび出力発話動画データによって、印象ラベル、入力発話動画データおよび入力顔画像データを入力とし、出力発話動画データにおける話者の表情変化Landmark系列を出力とする学習モデルを学習させる。これにより、学習部123Xは、各話者の顔のLandmark系列の持つ情報のうち、主に表情の情報が含まれる各話者の表情変化Landmark系列を推定可能な学習モデルとなるように学習モデルを学習させる。

[0054] (推定部124X)

推定部124Xは、学習部123Xによって学習された学習モデルに基づいて拡張発話画像データにおける第二の話者の表情の変化の特徴量を推定する。以下、図11を用いて、推定部124Xによる推定の一例について説明する。

[0055] 図11に示される例では、推定部124Xは、表情変化Landmarkや表情変化Landmark系列を推定対象とする点を除き、第一の実施形態に係るデータ構築システム1と同様に推定する。例えば、推定部124Xは、学習モデルを

用いて印象ラベル、入力発話動画データの映像データから抽出された映像特徴量および入力発話動画データの音声データから抽出された音声特徴量に基づき、表情変化Landmarkおよび表情変化Landmark系列を推定する。

[0056] 図11に示される例では、推定部124Xは、入力発話動画データの映像データにおける入力話者の顔のLandmarkの抽出および入力話者の顔のLandmark統計量の算出を行い、入力話者の顔のLandmark統計量を得る。次に、推定部124Xは、例えば、表情変化Landmark系列および入力話者画像データにおける入力話者の顔のLandmark統計量に基づいてデコード（個人性の反映）を行い、表情変化Landmark系列に入力話者の顔のLandmark統計量が加算されたLandmark系列を得る。次に、推定部124Xは、例えば、Landmark系列および入力話者画像に基づいてrenderingを行い、入力話者画像における話者の顔の個人性（平均的なLandmarkの位置）および表情の変化（平均的なLandmarkの位置からの差分）が反映された出力発話動画データを得る。

[0057] （処理の流れの一例）

以下、図12を用いて、第二の実施形態に係るデータ構築システム1Xのデータ構築装置10Xによって実行される処理の流れの一例について説明する。図12は、第二の実施形態に係るデータ構築システムによって実行される処理の流れの一例を示すフローチャートである。ステップS21～ステップS23は、第一の実施形態におけるステップS11～ステップS13と同様である。

[0058] ステップS24において、算出部125は、データ構築部122によって構築された学習データに基づいて、拡張発話画像データにおける第二の話者の表情の変化の特徴量を算出する。例えば、算出部125は、各印象ラベル付き拡張発話動画データにおける各話者の顔のLandmarkの座標の平均値および分散を算出する。次に、算出部125は、例えば、各話者の顔のLandmarkの座標の平均値および分散に基づいて、各話者の顔のLandmark統計量および各話者の表情変化Landmark系列を算出する。

[0059] ステップS25において、学習部123Xは、データ構築部122によっ

て構築された学習データ、印象ラベル、発話画像データおよび顔画像データによって、印象ラベル、発話画像データおよび顔画像データを入力とし、拡張発話画像データにおける第二の話者の表情の変化の特徴量を出力とする学習モデルを学習させる。例えば、学習部 1 2 3 X は、学習部 1 2 3 X は、算出部 1 2 5 によって算出された表情変化Landmark系列、印象ラベル、入力発話動画データ、入力顔画像データおよび出力発話動画データによって、印象ラベル、入力発話動画データおよび入力顔画像データを入力とし、出力発話動画データにおける話者の表情変化Landmark系列を出力とする学習モデルを学習させる。

[0060] ステップ S 2 6 において、推定部 1 2 4 X は、学習部 1 2 3 X によって学習された学習モデルに基づいて拡張発話画像データにおける第二の話者の表情の変化の特徴量を推定する。例えば、推定部 1 2 4 X は、学習モデルを用いて、印象ラベル、入力発話動画データにおける映像データから抽出された映像特徴量および入力発話動画データにおける音声データから抽出された音声特徴量に基づき、入力話者の表情変化Landmarkを推定する。

[0061] (3. プログラム)

データ構築装置 1 0 およびデータ構築装置 1 0 X は、パッケージソフトウェアやオンラインソフトウェアとしてデータ構築プログラムをコンピュータにインストールさせることにより実装できる。例えば、コンピュータがデータ構築プログラムを実行することにより、データ構築装置 1 0 およびデータ構築装置 1 0 X として機能させることができる。

[0062] 図 1 3 は、データ構築プログラムを実行するコンピュータの構成の一例を示す図である。図 1 3 に示される例では、コンピュータ 1 0 0 0 は、メモリ 1 0 1 0 と、CPU 1 0 2 0 と、ハードディスクドライバインタフェース 1 0 3 0 と、ディスクドライバインタフェース 1 0 4 0 と、シリアルポートインタフェース 1 0 5 0 と、ビデオアダプタ 1 0 6 0 と、ネットワークインタフェース 1 0 7 0 とを有する。これらの各部は、バス 1 0 8 0 によって接続される。

- [0063] メモリ 1010 は、ROM (Read Only Memory) 1011 および RAM 1012 を含む。ROM 1011 は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブインタフェース 1030 は、ハードディスクドライブ 1090 に接続される。ディスクドライブインタフェース 1040 は、ディスクドライブ 1100 に接続される。例えば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ 1100 に挿入される。シリアルポートインタフェース 1050 は、例えば、マウス 1110、キーボード 1120 に接続される。ビデオアダプタ 1060 は、例えば、ディスプレイ 1130 に接続される。
- [0064] ハードディスクドライブ 1090 は、例えば、OS 1091、アプリケーションプログラム 1092、プログラムモジュール 1093 およびプログラムデータ 1094 を記憶する。すなわち、データ構築装置 10 およびデータ構築装置 10X によって実行される各処理を規定するプログラムは、コンピュータにより実行可能なコードが記述されたプログラムモジュール 1093 として実装される。プログラムモジュール 1093 は、例えば、ハードディスクドライブ 1090 に記憶される。例えば、データ構築装置 10 およびデータ構築装置 10X における機能構成と同様の処理を実行するためのプログラムモジュール 1093 が、ハードディスクドライブ 1090 に記憶される。なお、ハードディスクドライブ 1090 は、SSD によって代替されてよい。
- [0065] また、上述の実施形態の処理において用いられるデータは、プログラムデータ 1094 として、例えば、メモリ 1010 やハードディスクドライブ 1090 に記憶される。そして、CPU 1020 が、メモリ 1010 やハードディスクドライブ 1090 に記憶されたプログラムモジュール 1093 やプログラムデータ 1094 を、必要に応じて RAM 1012 に読み出して実行する。
- [0066] なお、プログラムモジュール 1093 やプログラムデータ 1094 は、ハードディスクドライブ 1090 に記憶される場合に限定されず、例えば、着

脱可能な記憶媒体に記憶され、ディスクドライブ 1100等を介してCPU 1020によって読み出されてもよい。あるいは、プログラムモジュール 1093およびプログラムデータ 1094は、ネットワーク（LAN（Local Area Network）、WAN（Wide Area Network）等）を介して接続される他のコンピュータに記憶されてもよい。そして、プログラムモジュール 1093およびプログラムデータ 1094は、他のコンピュータから、ネットワークインタフェース 1070を介してCPU 1020によって読み出されてもよい。

[0067] （４．変形例）

図示された各部の各構成要素は機能概念的なものであり、必ずしも物理的に図示のように構成されていることを要しない。すなわち、装置の分散、統合の具体的形態は図示のものに限定されず、その全部または一部を、各種の負荷や使用状況等に応じて、任意の単位にて機能的または物理的に分散、統合させて構成することができる。さらに、装置にて行われる各処理機能は、その全部または任意の一部が、CPUおよび当該CPUにて実行されるプログラムにて実現されるか、または、ワイヤードロジックによるハードウェアとして実現されうる。例えば、図2に示される学習部 123および推定部 124は、それぞれ、データ構築装置 10とは独立した学習装置および推定装置として実現されうる。

[0068] また、上述の各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともできれば、手動的に行われるものとして説明した処理の全部または一部を公知の方法によって自動的に行うこともできる。この他、上述の文書中や図面中において示した処理手順、具体的名称、各種情報やパラメータ等については、特記されている場合を除いて任意に変更することができる。

符号の説明

- [0069] 1 データ構築システム
 10 データ構築装置

- 1 1 データ取得部
- 1 2 制御部
- 1 3 出力部
- 1 4 記憶部
- 1 2 1 データ拡張部
- 1 2 2 データ構築部
- 1 2 3 学習部
- 1 2 4 推定部

請求の範囲

- [請求項1] 第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得するデータ取得部と、
- 前記データ取得部によって取得された前記発話画像データおよび前記顔画像データに基づいて、前記第二の話者による前記発話に関する拡張発話画像データを生成するデータ拡張部と、
- 前記データ拡張部によって生成された前記拡張発話画像データと前記発話の印象に関する印象ラベルとの組を、学習モデルを学習させるための学習データとして構築するデータ構築部と、
- を有することを特徴とするデータ構築装置。
- [請求項2] 前記データ取得部は、所定の数未満の少数話者のそれぞれによる発話に関する少数話者発話画像データベースから前記発話画像データを取得し、所定の数以上の多数の話者のそれぞれの顔画像データに関する多数顔画像データベースから前記顔画像データを取得することを特徴とする請求項1に記載のデータ構築装置。
- [請求項3] 前記データ取得部は、前記発話画像データが含まれる発話動画データおよび前記顔画像データを取得し、
- 前記データ拡張部は、前記データ取得部によって取得された前記発話動画データおよび前記顔画像データに基づいて、前記拡張発話画像データが含まれる拡張発話動画データを生成し、
- 前記データ構築部は、前記データ拡張部によって生成された前記拡張発話動画データと前記印象ラベルとの組を前記学習データとして構築することを特徴とする請求項1に記載のデータ構築装置。
- [請求項4] 前記データ構築部によって構築された前記学習データ、前記発話画像データおよび前記顔画像データによって、前記印象ラベル、前記発話画像データおよび前記顔画像データを入力とし、前記拡張発話画像データにおける前記第二の話者の顔の特徴量を出力とする前記学習モデルを学習させる学習部をさらに有することを特徴とする請求項1～

3のいずれか1項に記載のデータ構築装置。

[請求項5] 前記学習部は、前記データ構築部によって構築された前記学習データ、前記印象ラベル、前記発話画像データおよび前記顔画像データによって、前記印象ラベル、前記発話画像データおよび前記顔画像データを入力とし、前記拡張発話画像データにおける前記第二の話者の表情の変化の特徴量を出力とする前記学習モデルを学習させることを特徴とする請求項4に記載のデータ構築装置。

[請求項6] 前記学習部は、前記第二の話者の顔の特徴量と当該第二の話者の平均的な顔の特徴量との差分を前記第二の話者の表情の変化の特徴量として出力する前記学習モデルを学習させることを特徴とする請求項5に記載のデータ構築装置。

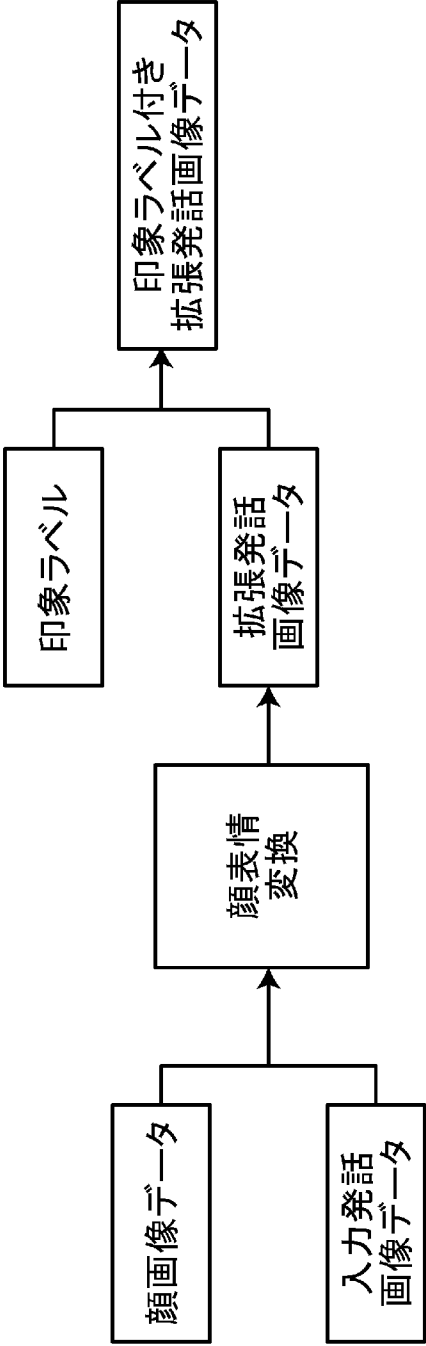
[請求項7] データ構築装置によって実行されるデータ構築方法であって、
第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得するデータ取得工程と、
前記データ取得工程によって取得された前記発話画像データおよび前記顔画像データに基づいて、前記第二の話者による前記発話に関する拡張発話画像データを生成するデータ拡張工程と、
前記データ拡張工程によって生成された前記拡張発話画像データと前記発話の印象に関する印象ラベルとの組を、学習モデルを学習させるための学習データとして構築するデータ構築工程と、
を含むことを特徴とするデータ構築方法。

[請求項8] 第一の話者による発話に関する発話画像データおよび第二の話者の顔に関する顔画像データを取得するデータ取得ステップと、
前記データ取得ステップによって取得された前記発話画像データおよび前記顔画像データに基づいて、前記第二の話者による前記発話に関する拡張発話画像データを生成するデータ拡張ステップと、
前記データ拡張ステップによって生成された前記拡張発話画像データと前記発話の印象に関する印象ラベルとの組を、学習モデルを学習

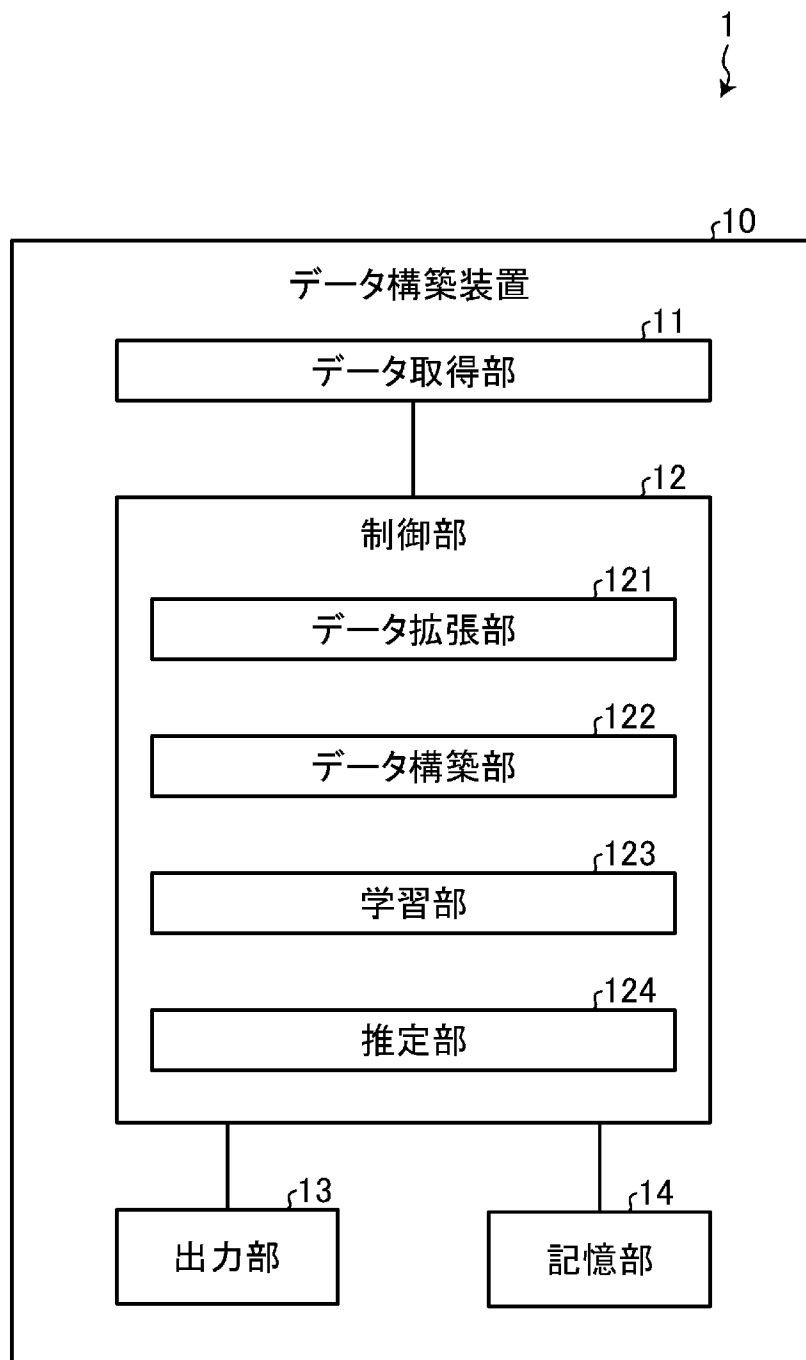
させるための学習データとして構築するデータ構築ステップと、
をコンピュータに実行させることを特徴とするデータ構築プログラム。
ム。

[図1]

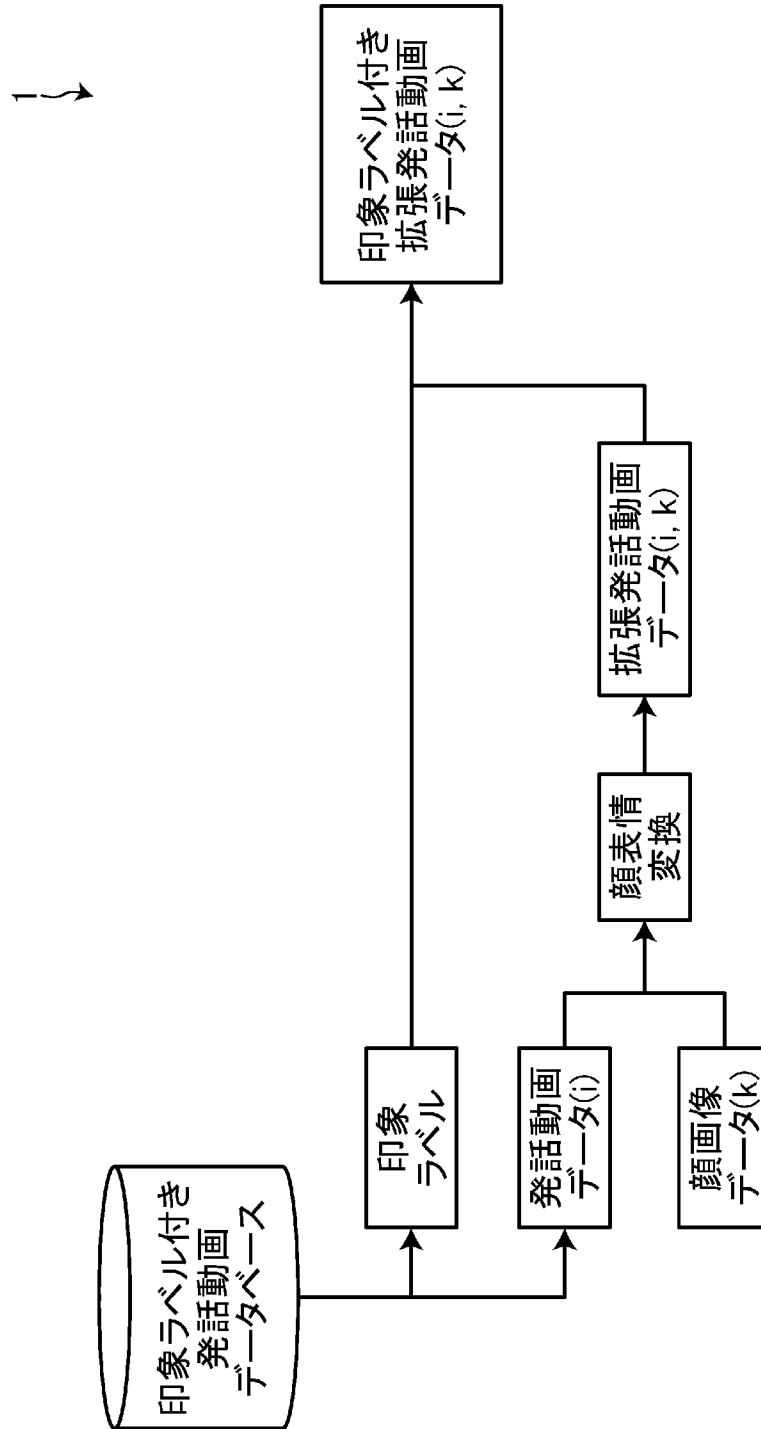
1



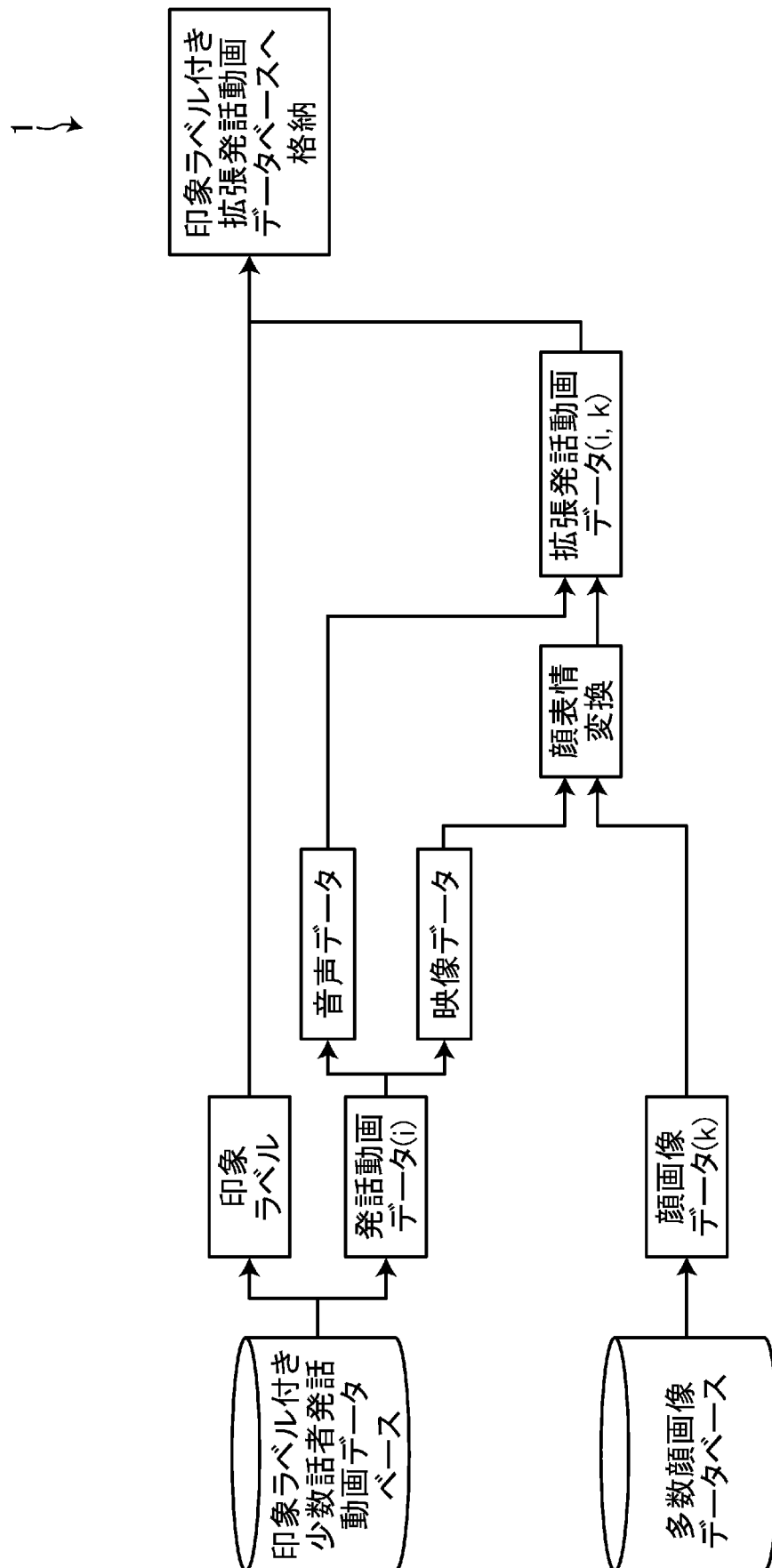
[図2]



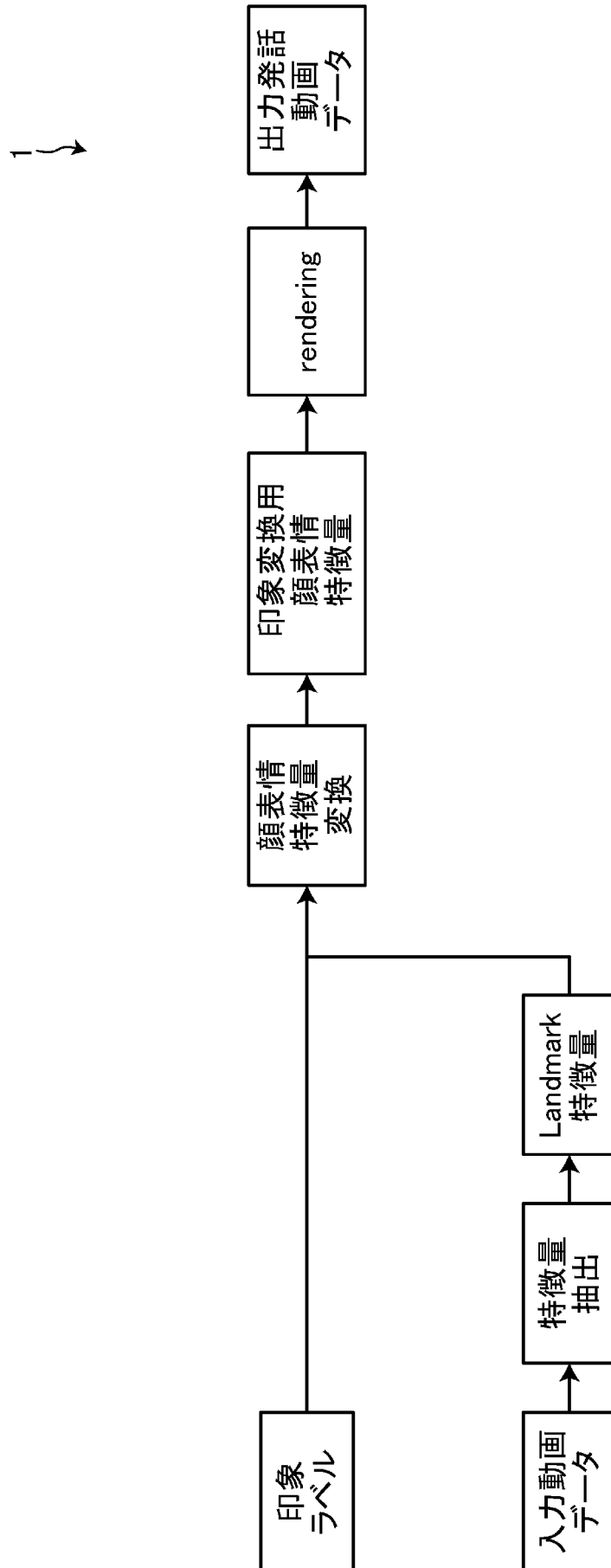
[図3]



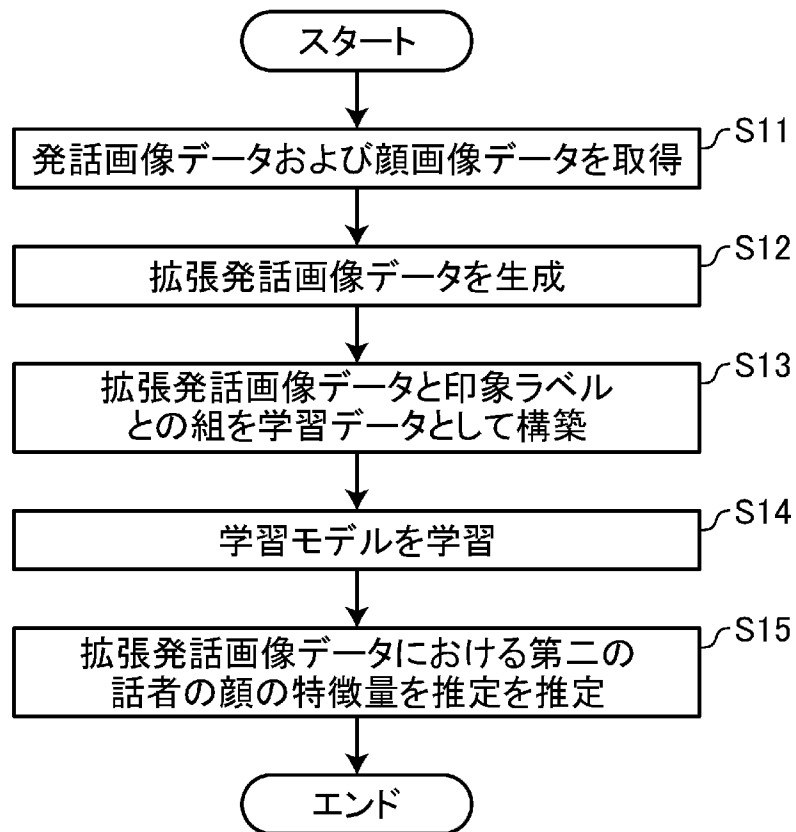
[図4]



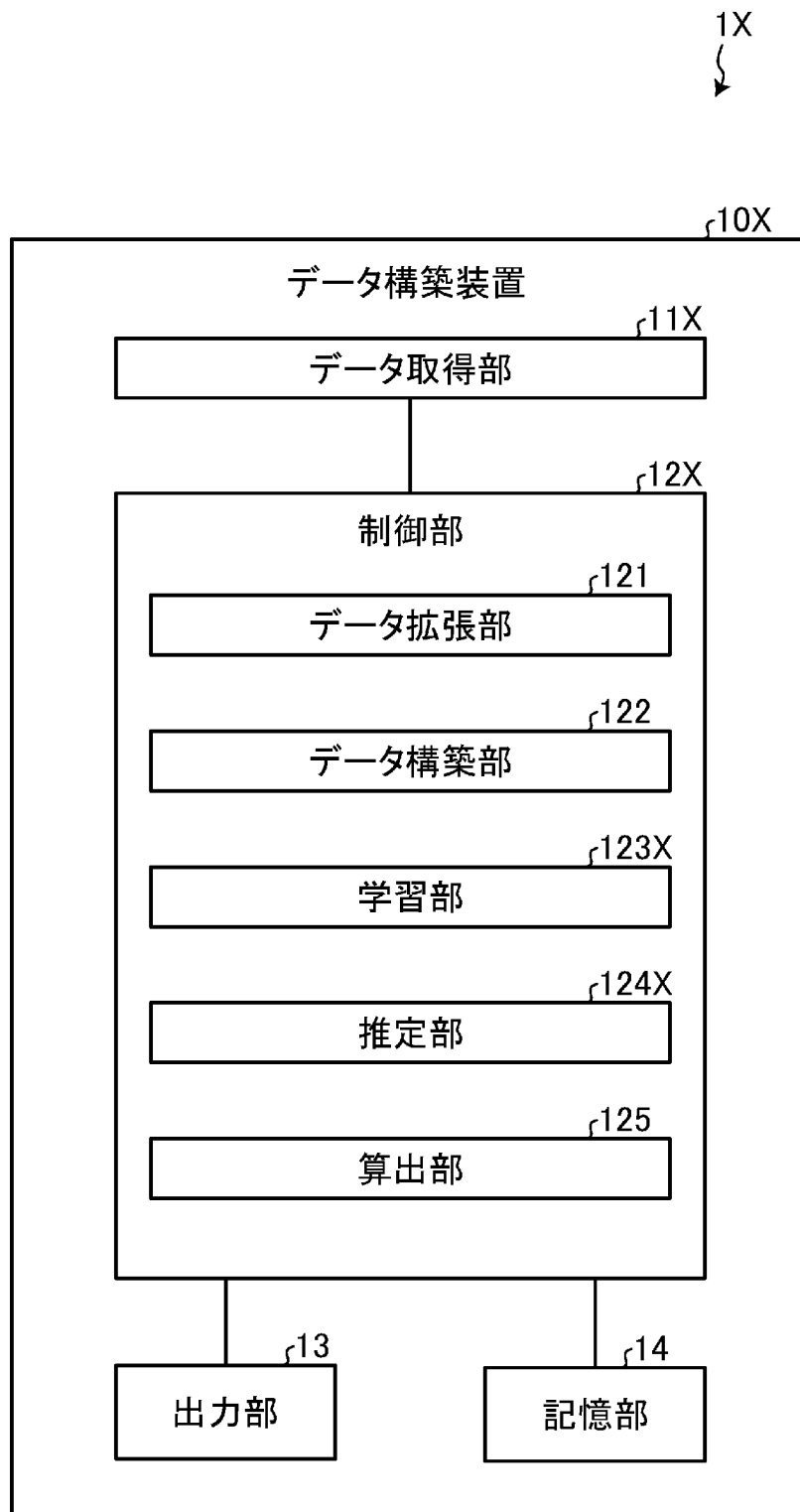
[図5]



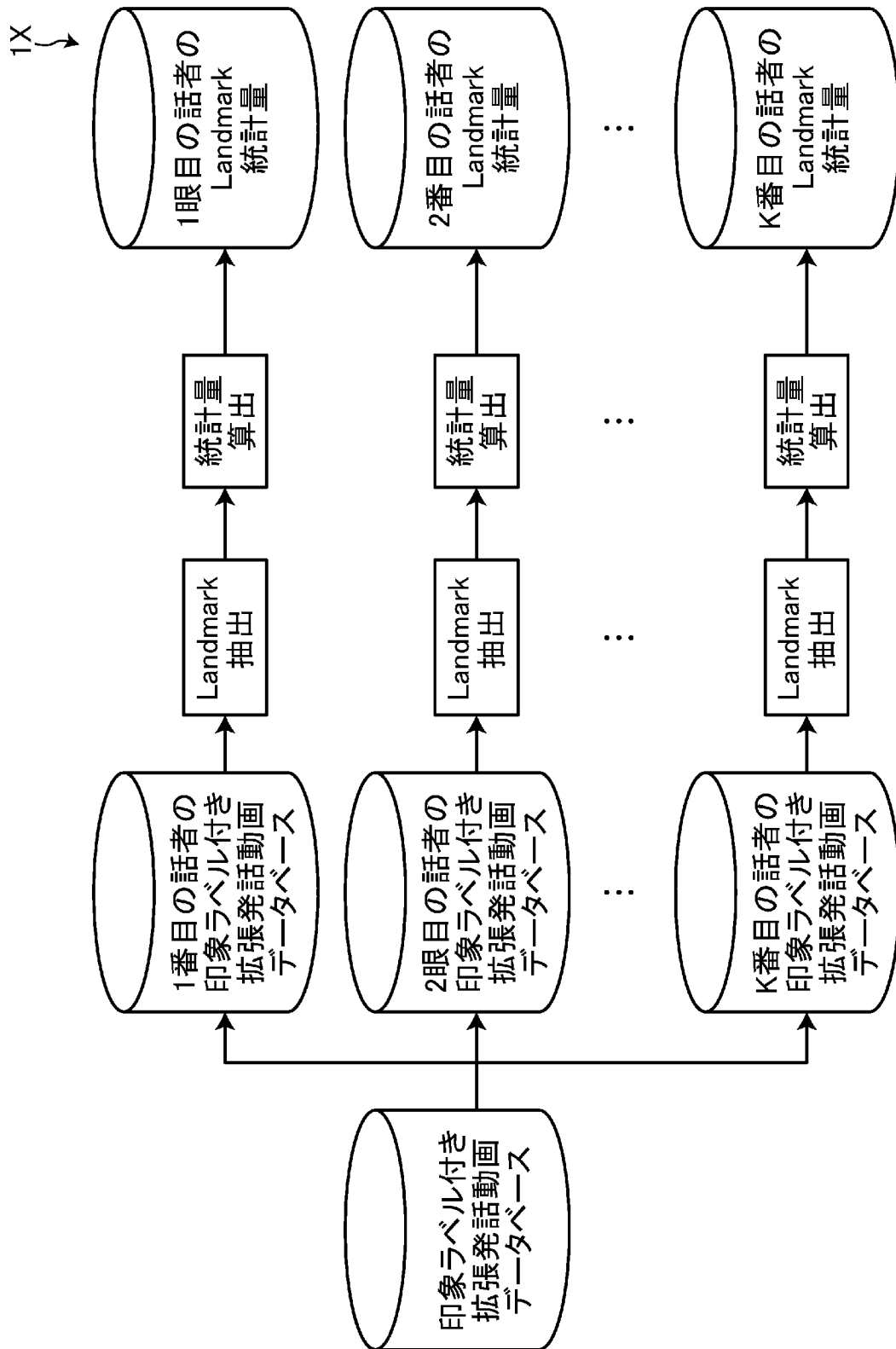
[図6]



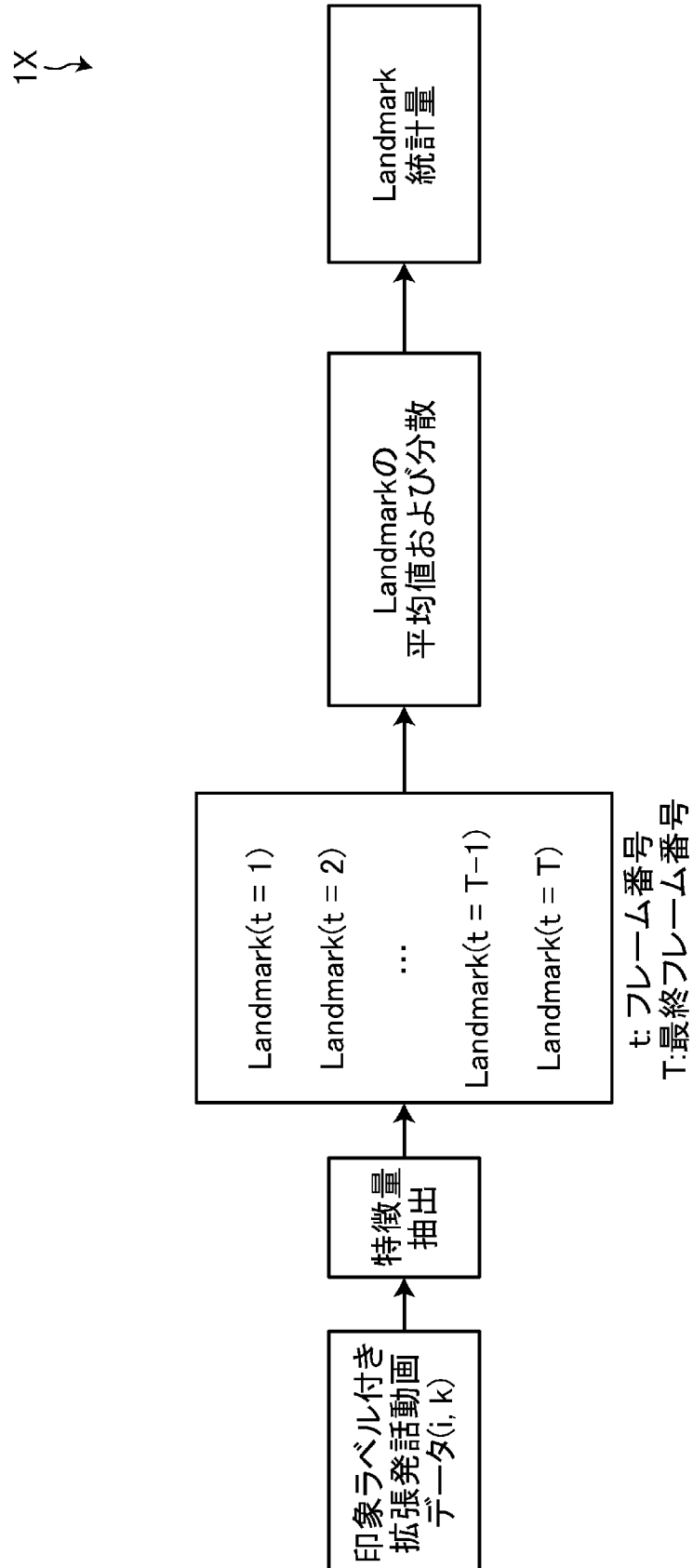
[図7]



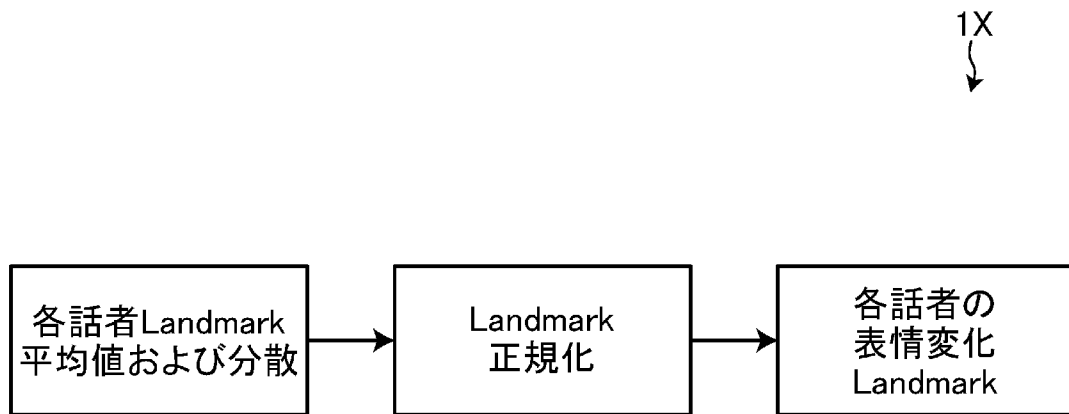
[図8]



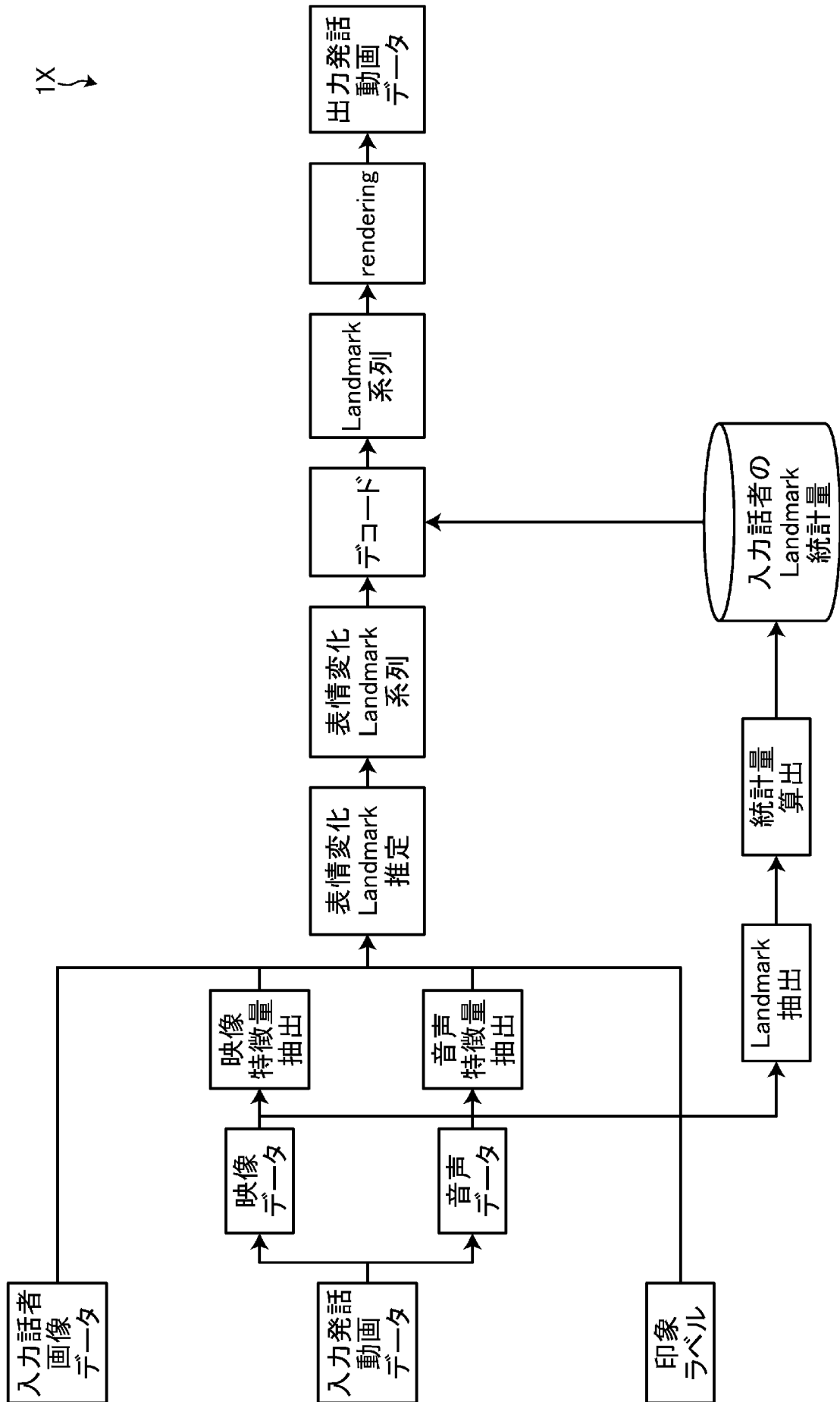
[図9]



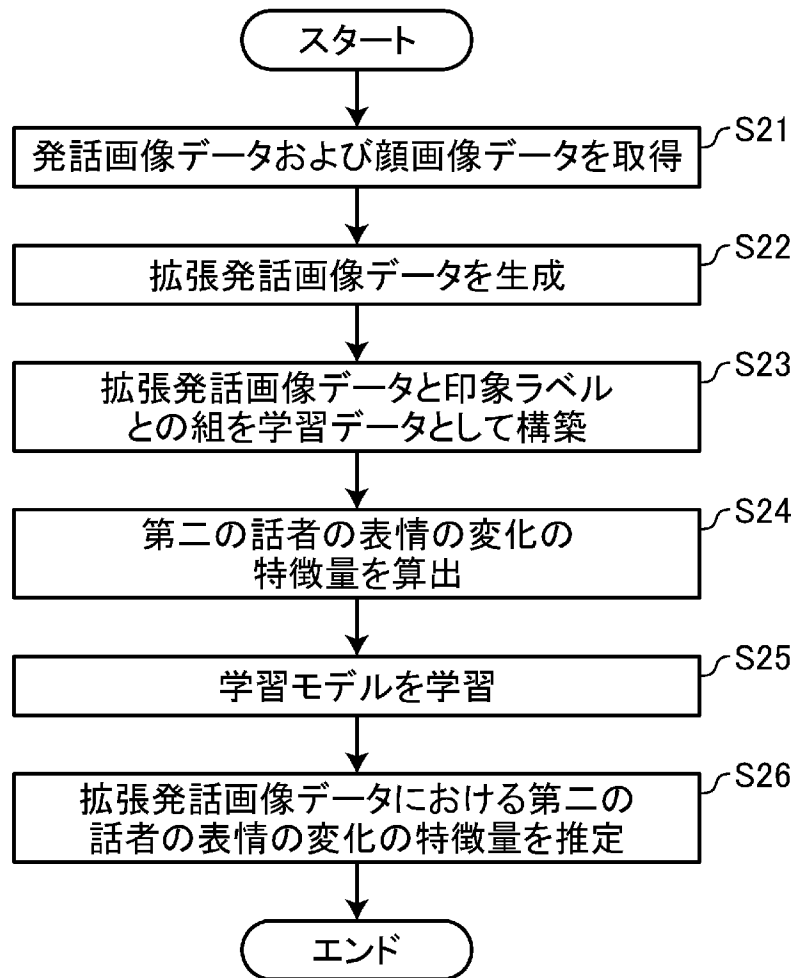
[図10]



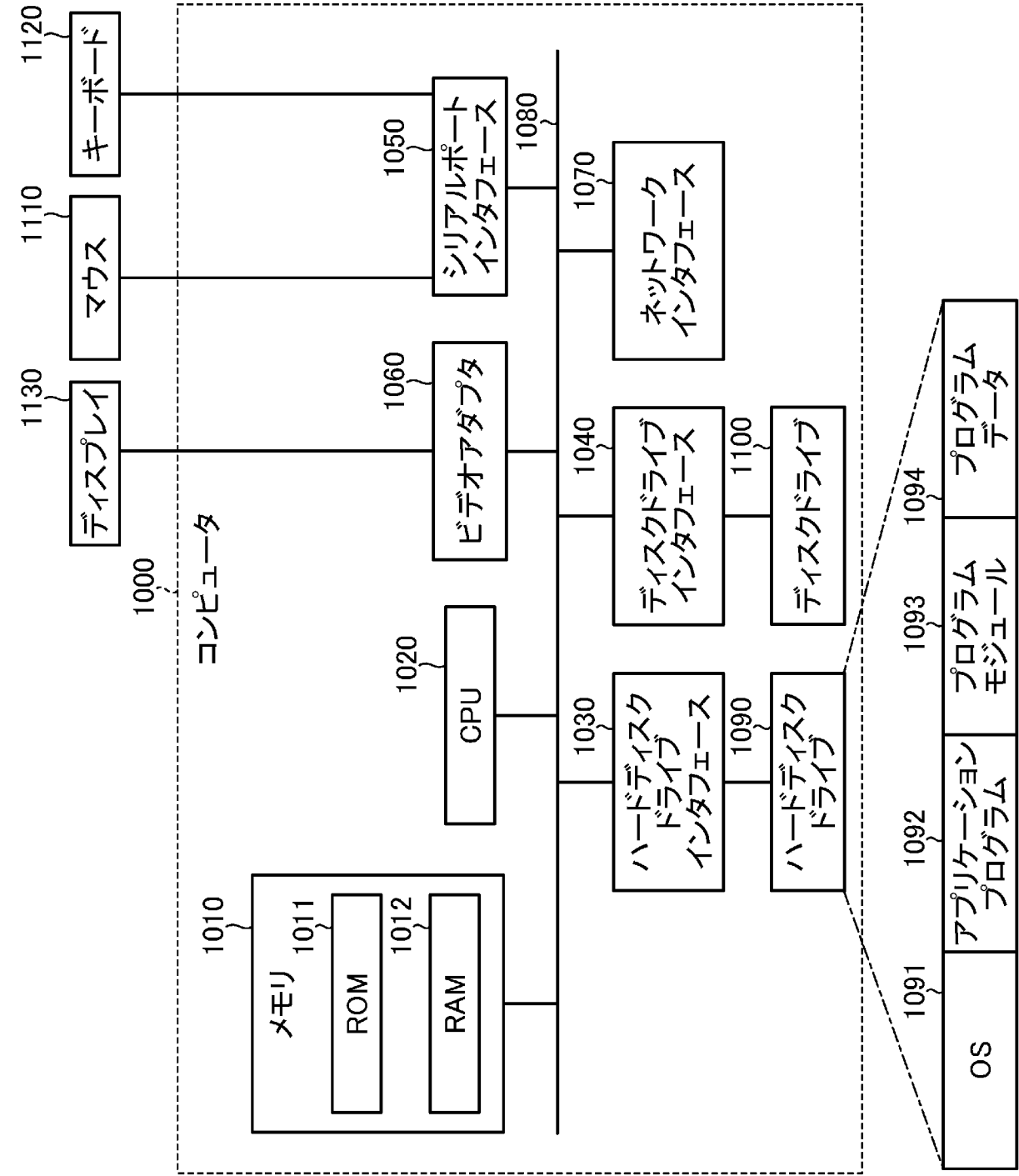
[図11]



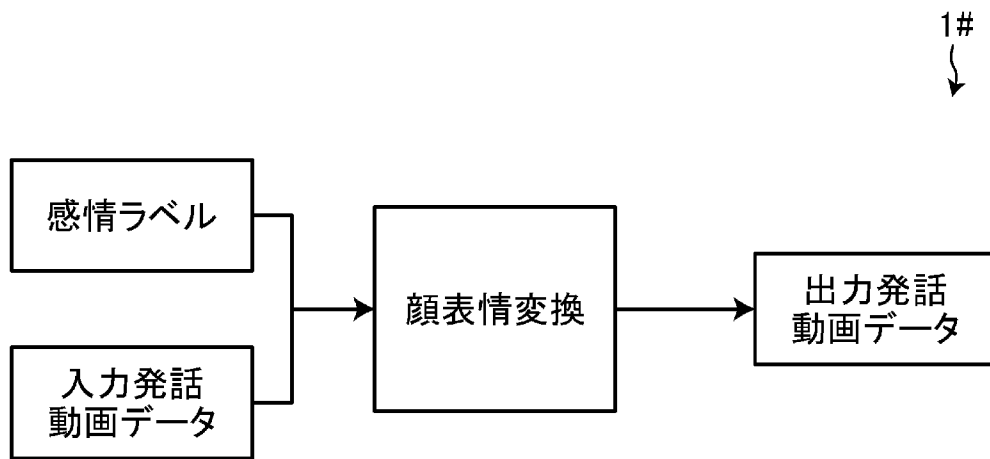
[図12]



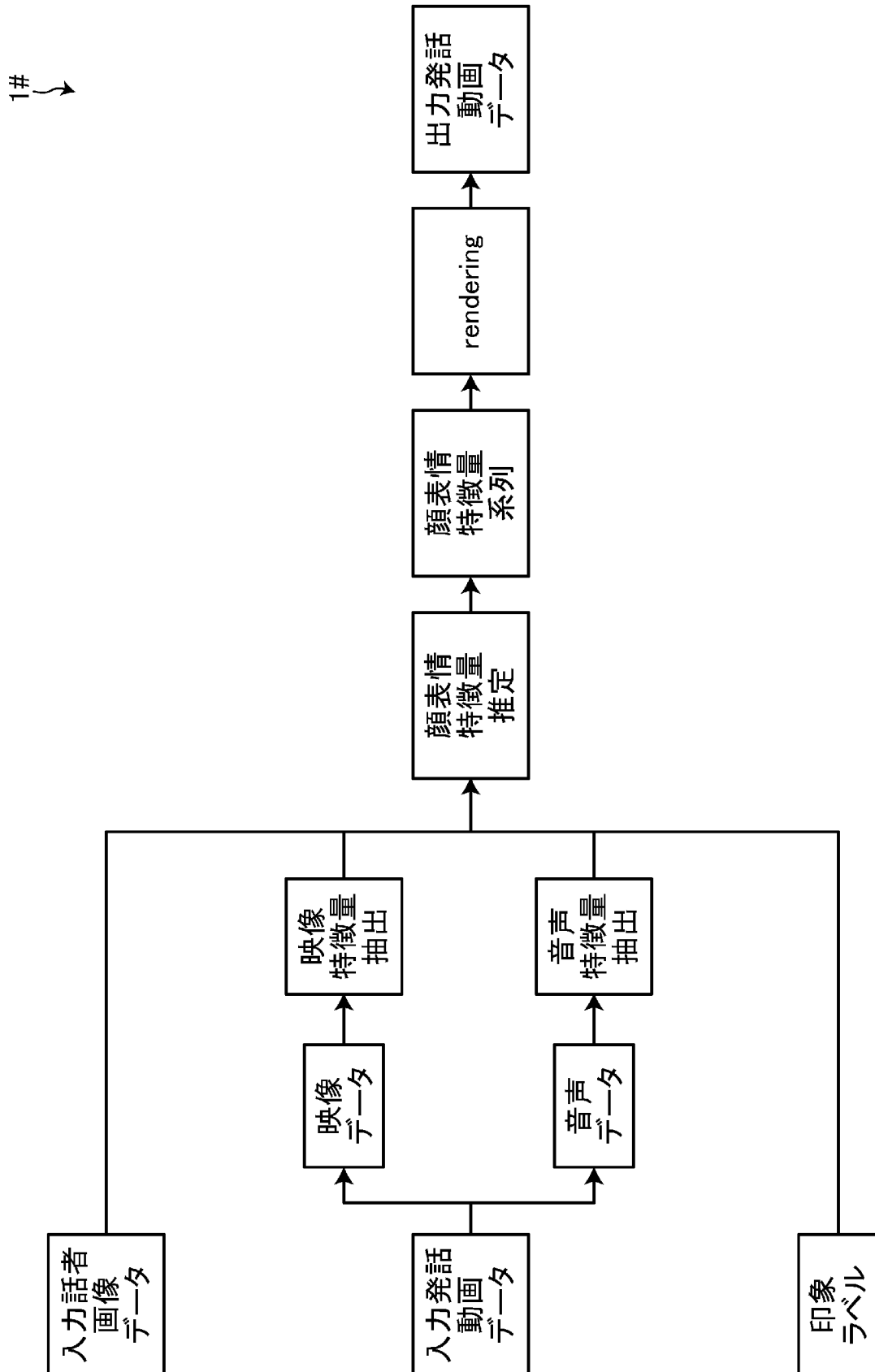
[図13]



[図14]



[図15]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/025022

A. CLASSIFICATION OF SUBJECT MATTER**G06V 10/774**(2022.01)i

FI: G06V10/774

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)

G06V10/774

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.
Y	JP 2022-175923 A (AIINFLUENCER INC.) 25 November 2022 (2022-11-25) paragraphs [0030], [0080], [0085]-[0086]	1-8
Y	JP 7207539 B2 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 18 January 2023 (2023-01-18) paragraph [0003]	1-8
Y	WO 2019/146315 A1 (HITACHI, LTD.) 01 August 2019 (2019-08-01) paragraphs [0059]-[0062]	1-8
Y	CN 113807265 A (SHANDONG UNIV OF FINANCE AND ECONOMICS) 17 December 2021 (2021-12-17) paragraphs [0055]-[0066]	4-6



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

07 September 2023

Date of mailing of the international search report

19 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/JP2023/025022

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2022-175923	A	25 November 2022	(Family: none)	
JP	7207539	B2	18 January 2023	US 2022/0383106 A1 paragraph [0003]	
WO	2019/146315	A1	01 August 2019	(Family: none)	
CN	113807265	A	17 December 2021	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06V 10/774(2022.01)i FI: G06V10/774		
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）） G06V10/774		
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
Y	JP 2022-175923 A (A I インフルエンサー株式会社) 25.11.2022 (2022 - 11 - 25) 段落 [0030] , [0080] , [0085] - [0086]	1-8
Y	JP 7207539 B2 (日本電信電話株式会社) 18.01.2023 (2023 - 01 - 18) 段落 [0003]	1-8
Y	WO 2019/146315 A1 (株式会社日立製作所) 01.08.2019 (2019 - 08 - 01) 段落 [0059] - [0062]	1-8
Y	CN 113807265 A (SHANDONG UNIVERSITY OF FINANCE AND ECONOMICS) 17.12.2021 (2021 - 12 - 17) 段落 [0055] - [0066]	4-6
☐ 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 07.09.2023	国際調査報告の発送日 19.09.2023	
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） 山田 辰美 5H 8383 電話番号 03-3581-1101 内線 3531	

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/JP2023/025022

引用文献			公表日	パテントファミリー文献	公表日
JP	2022-175923	A	25.11.2022	(ファミリーなし)	
JP	7207539	B2	18.01.2023	US 2022/0383106 A1	
				段落 [0003]	
WO	2019/146315	A1	01.08.2019	(ファミリーなし)	
CN	113807265	A	17.12.2021	(ファミリーなし)	