(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
2017 年 10 月 5 日(05.10.2017)

WIPO | PCT

(10) 国際公開番号
WO 2017/168870 A1

(51) 国際特許分類:
G10L 13/10 (2013.01)

(21) 国際出願番号: PCT/JP2016/087316

(22) 国際出願日: 2016 年 12 月 14 日(14.12.2016)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2016-063784 2016 年 3 月 28 日(28.03.2016) JP

(71) 出願人: ソニー株式会社(SONY CORPORATION) [JP/JP]; 〒1080075 東京都港区港南１丁目７番１号 Tokyo (JP).

(72) 発明者: 井手　直紀(IDE, Naoki); 〒1080075 東京都港区港南１丁目７番１号　ソニー株式会社内 Tokyo (JP). 藤田　拓也(FUJITA, Takuya); 〒1080075 東京都港区港南１丁目７番１号　ソニー株式会社内 Tokyo (JP). 中村　章(NAKAMURA, Akira); 〒1080075 東京都港区港南１丁目７番１号　ソニー株式会社内 Tokyo (JP). 成平　拓也(NARIHIRA, Takuya); 〒1080075 東京都港区港南１丁目７番１号　ソニー株式会社内 Tokyo (JP). 中橋　亮(NAKAHASHI, Ryo); 〒1080075 東京都港区港南１丁目７番１号　ソニー株式会社内 Tokyo (JP).

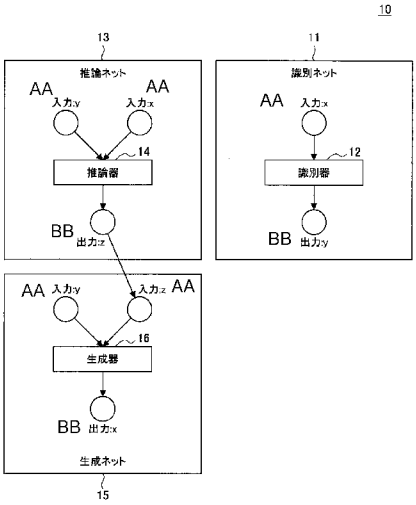
(74) 代理人: 亀谷　美明, 外(KAMEYA, Yoshiaki et al.); 〒1600004 東京都新宿区四谷３－１－３　第一冨澤ビル　はづき国際特許事務所　四谷オフィス Tokyo (JP).

(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE,

[続葉有]

(54) **Title**: INFORMATION PROCESSING DEVICE AND INFORMATION PROCESSING METHOD

(54) 発明の名称 : 情報処理装置及び情報処理方法



11 Identification network
12 Identification apparatus
13 Inference network
14 Inference apparatus
15 Generation network
16 Generation apparatus
AA Input
BB Output

(57) **Abstract**: [Problem] To provide a system that can appropriately characterize data. [Solution] An information processing device comprising: an acquisition unit that acquires first characteristic information, to which meaning is assigned, and second characteristic information, to which meaning is not assigned; and a generation unit that generates time series data having characteristics that indicate the first characteristic information and the second characteristic information acquired by the acquisition unit.

(57) 要約: 【課題】データの特徴付けを適切に行うことが可能な仕組みが提供する。 【解決手段】意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得する取得部と、前記取得部により取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを生成する生成部と、を備える情報処理装置。

WO 2017/168870 A1

SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告（条約第 21 条(3)）

明　細　書

発明の名称 ： 情報処理装置及び情報処理方法

技術分野

[0001] 本開示は、情報処理装置及び情報処理方法に関する。

背景技術

[0002] 近年、データを画一的に表現するのではなく、何らかの特徴を付けて表現する技術が注目されている。例えば、下記特許文献１では、テキストから音声データを生成する際に自然な発話音声を表現するために、イントネーションの特徴等を示す統計的表現を用いた学習を行う技術が開示されている。

先行技術文献

特許文献

[0003] 特許文献1：特表平８－５０８１２７号公報

発明の概要

発明が解決しようとする課題

[0004] しかし、データの特徴付けを適切に行うことは困難であった。一例として音声に関して言えば、個人の声紋、イントネーション、アクセント、訛り、間の取り方、会話前後のコンテキスト、又は今の話者の状態（特に喜怒哀楽の感情の状態）等を、数値等の特徴量として操作可能に設計することは困難であった。そのため、データの特徴付けを適切に行うことが可能な仕組みが提供されることが望ましい。

課題を解決するための手段

[0005] 本開示によれば、意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得する取得部と、前記取得部により取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを生成する生成部と、を備える情報処理装置が提供される。

[0006] また、本開示によれば、意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を他の装置に通知する通知部と、前記第１の特徴情報及

び前記第２の特徴情報が示す特徴を有する時系列データを前記他の装置から取得する取得部と、を備える情報処理装置が提供される。

[0007] また、本開示によれば、意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得することと、取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データをプロセッサにより生成することと、を含む情報処理方法が提供される。

[0008] また、本開示によれば、意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を他の装置に通知することと、前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを前記他の装置からプロセッサにより取得することと、を含む情報処理方法が提供される。

発明の効果

[0009] 以上説明したように本開示によれば、データの特徴付けを適切に行うことが可能な仕組みが提供される。なお、上記の効果は必ずしも限定的なものではなく、上記の効果とともに、または上記の効果に代えて、本明細書に示されたいずれかの効果、または本明細書から把握され得る他の効果が奏されてもよい。

図面の簡単な説明

[0010] [図1]本開示の一実施形態に係るシステムの概略的な構成の一例を示す図である。

[図2]本実施形態に係る処理装置の論理的な構成の一例を示すブロック図である。

[図3]本実施形態に係る端末装置の論理的な構成の一例を示すブロック図である。

[図4]本実施形態に係るオートエンコーダの概要を説明するための図である。

[図5]本実施形態に係る処理装置において実行される学習処理の流れの一例を示すフローチャートである。

[図6]本実施形態に係るシステムにおいて実行されるインタラクション処理の流れの一例を示すシーケンス図である。

[図7]本実施形態に係るシステムにおいて実行されるインタラクション処理の流れの一例を示すシーケンス図である。

[図8]第１の実施形態に係る識別ネットの構成の一例を示す図である。

[図9]同実施形態に係る推論ネットの構成の一例を示す図である。

[図10]同実施形態に係る生成ネットの構成の一例を示す図である。

[図11]同実施形態に係る推論処理の一例を説明するための図である。

[図12]同実施形態に係るユーザインタフェースの一例を示す図である。

[図13]同実施形態に係る音声特徴情報ＤＢの一例を示す図である。

[図14]同実施形態に係るユーザインタフェースの一例を示す図である。

[図15]同実施形態に係る翻訳場面における音声データ生成処理を説明するための図である。

[図16]第２の実施形態に係る識別ネットの構成の一例を示す図である。

[図17]同実施形態に係る推論ネットの構成の一例を示す図である。

[図18]同実施形態に係る生成ネットの構成の一例を示す図である。

[図19]同実施形態に係る処理装置において実行される学習処理の流れの一例を示すフローチャートである。

[図20]同実施形態に係るユーザインタフェースの一例を示す図である。

[図21]同実施形態に係る生成ネット１５の構成の一例を示す図である。

[図22]同実施形態に係る処理装置において実行される学習処理の流れの一例を示すフローチャートである。

[図23]同実施形態に係る制御信号生成処理の一例を説明するための図である。

[図24]同実施形態に係る制御信号生成処理の一例を説明するための図である。

[図25]同実施形態に係る制御信号生成処理の一例を説明するための図である。

[図26]本実施形態に係る情報処理装置のハードウェア構成の一例を示すブロック図である。

発明を実施するための形態

[0011] 以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

[0012] なお、説明は以下の順序で行うものとする。

１．はじめに
１．１．オートエンコーダの概要
１．２．オートエンコーダの詳細
１．３．ＶＡＥの詳細
２．構成例
２．１．システムの構成例
２．２．処理装置の構成例
２．３．端末装置の構成例
３．基本的な技術的特徴
４．第１の実施形態
４．１．各種データ
４．２．ニューラルネットの構成
４．３．学習
４．４．音声データの生成
４．５．補足
５．第２の実施形態
５．１．各種データ
５．２．ニューラルネットの構成
５．３．学習
５．４．センサデータの生成
５．５．制御信号の生成
６．補足

７．ハードウェア構成例

８．まとめ

[0013] ＜＜１．はじめに＞＞

＜１．１．オートエンコーダの概要＞

オートエンコーダ（Auto-Encoder）とは、自己符号化器とも称される、ニューラルネットの技術である。本明細書では、オートエンコーダの一例として、変分オートエンコーダ（ＶＡＥ：Variational Auto-Encoder）を用いた技術を説明する。もちろん、変分オートエンコーダ以外の、例えば生成アドバーサリアルネットワーク（ＧＡＮ：Generative Adversarial Network）が用いられてもよい。

[0014] なお、オートエンコーダは、「岡谷貴之、“深層学習”、講談社、２０１５年４月８日」に詳しく記載されている。また、ＶＡＥは、「Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013).」、及び「Kingma, Diederik P., et al. "Semi-supervised learning with deep generative models." Advances in Neural Information Processing Systems. 2014.」に詳しく記載されている。また、ＧＡＮは、「Goodfellow, Ian, et al. "Generative adversarial nets." Advances in Neural Information Processing Systems. 2014.」に詳しく記載されている。

[0015] また、ＶＡＥの時系列拡張版として、再帰ニューラルネット（ＲＮＮ：Recurrent Neural Network）と結合したオートエンコーダが用いられてもよい。ＲＮＮについては、「岡谷貴之、“深層学習”、講談社、２０１５年４月８日」に詳しく記載されている。また、ＲＮＮとＶＡＥとが結合したオートエンコーダに関しては、「Fabius, Otto, and Joost R. van Amersfoort. "Variational recurrent auto-encoders." arXiv preprint arXiv:1412.6581 (2014).」、「Bayer, Justin, and Christian Osendorfer. "Learning stochastic recurrent networks." arXiv preprint arXiv:1411.7610 (2014).」、及び「Chung, Junyoung, et al. "A recurre

nt latent variable model for sequential data." Advances in neural information processing systems. 2015.」に詳しく記載されている。

[0016] ＜１．２．オートエンコーダの詳細＞

オートエンコーダは、ニューラルネットで構成される関数である。オートエンコーダでは、まず、次式のように、データｘは、ニューラルネットで規定された第１の射影関数に入力されて、一旦、中間変数ｚに変換される。

[0017] [数1]

$$z = f_{\phi}(x) \quad \cdots (1)$$

[0018] ただし、φは、第１のニューラルネットのすべての重みパラメータ（バイアスも含む）である。

[0019] そして、次式のように、この中間変数ｚは、ニューラルネットで規定された第２の射影関数に入力されて、再構成データｘ’が出力される。

[0020] [数2]

$$x' = g_{\theta}(z) \quad \cdots (2)$$

[0021] ただし、θは、第２のニューラルネットのすべての重みパラメータ（バイアスも含む）である。

[0022] ここで、第１の射影関数はエンコーダとも称され、第２の射影関数はデコーダとも称され得る。なお、エンコーダは、後述する推論ネットに相当し、デコーダは、後述する生成ネットに相当する。

[0023] オートエンコーダでは、再構成データｘ’が、データｘに近くなるように、第１の射影関数と第２の射影関数とが学習される。学習のための目的関数は、学習データセットをＸ＝｛x_n｝とすると例えば次式で表現される。

[0024] [数3]

$$L(\phi, \theta) = \sum_{n=1}^{N} \left(x_n - g_{\theta}\left(f_{\phi}(x_n) \right) \right)^2 \quad \cdots (3)$$

[0025] 例えば、上記数式（３）に示した目的関数Ｌを最小化するφ及びθが学習される。例えば、勾配法などは、目的関数Ｌをφ及びθの各々で微分することで得られた勾配に従って、Ｌが小さくなるようφ及びθを更新することで、学習が行われる。

[0026] 学習の際にはエンコーダとデコーダとが組み合わされて使用される一方、学習以外の場面ではエンコーダとデコーダとが別々に使用されてもよい。エンコーダがデータｘから変数ｚを計算する過程は推論とも称され得る。変数ｚは特徴量とも称され得る。また、デコーダが変数ｚから再構成データｘ’を計算する過程は生成とも称され得る。

[0027] ＜１．３．ＶＡＥの詳細＞

ＶＡＥは、オートエンコーダに確率の概念をもちこんだモデルである。

[0028] ＶＡＥでは、データｘ及び変数ｚが確率変数として扱われる。つまり、第１及び第２の射影関数が、決定論的ではなく、確率分布ｐ（ｚ｜ｘ）、ｐ（ｘ｜ｚ）からのサンプリングを含む確率的な射影となる。なお、事後確率ｐ（ｚ｜ｘ）としては、変分法で近似された確率分布ｑ（ｚ｜ｘ）が用いられる。さらに、これらの確率分布は、ガウス分布、ベルヌーイ分布、又は多項分布のような限られた個数のパラメータで決まる分布で近似される。このようにパラメタライズされた確率分布が用いられる場合、上記確率分布は次式のように表される。

[0029] [数4]

$$q(z|x) = q\big(z|\phi(x)\big)$$
$$p(x|z) = p(x|\theta(z)) \qquad \cdots（４）$$

[0030] ここで、φ（ｘ）、θ（ｚ）は、入力（ｘ，ｚ）に対して、確率分布のパラメータφ及びθの各々が出力となる射影関数である

[0031] 数式（１）及び数式（２）の変数を確率変数と考えて確率過程も含むものとすれば、ＶＡＥの数式も数式（１）及び数式（２）で表現可能である。なお、ＶＡＥでは、数式（３）に示した目的関数Ｌを最小化する方法とは異なり、モデルエビデンス（尤度のようなもの）を最大化することで学習が行わ

れる。実際には、モデルエビデンスの下界を最大化するパラメータが求められる。詳しい数式については、上記文献に詳しく記載されているので説明を省略する。

[0032] ＜＜２．構成例＞＞

＜２．１．システムの構成例＞

図１は、本開示の一実施形態に係るシステムの概略的な構成の一例を示す図である。図１に示すように、システム１は、処理装置１００及び端末装置２００を含む。

[0033] 処理装置１００及び端末装置２００は、ネットワーク３００により接続される。ネットワーク３００は、ネットワーク３００により接続されている装置から送信される情報の有線又は無線の伝送路である。ネットワーク３００は、例えばセルラーネットワーク、有線ＬＡＮ（Local Area Network）、又は無線ＬＡＮ等を含み得る。

[0034] 処理装置１００は、各種処理を行う情報処理装置である。端末装置２００は、ユーザとのインタフェースとして機能する情報処理装置である。典型的には、システム１は、処理装置１００及び端末装置２００の協働により、ユーザとのインタラクションを行う。

[0035] 続いて、図２及び図３を参照して、各装置の構成例を説明する。

[0036] ＜２．２．処理装置の構成例＞

図２は、本実施形態に係る処理装置１００の論理的な構成の一例を示すブロック図である。図２に示すように、処理装置１００は、通信部１１０、記憶部１２０及び制御部１３０を含む。

[0037] （１）通信部１１０

通信部１１０は、情報を送受信する機能を有する。例えば、通信部１１０は、端末装置２００からの情報を受信し、端末装置２００への情報を送信する。

[0038] （２）記憶部１２０

記憶部１２０は、処理装置１００の動作のためのプログラム及び様々なデ

ータを一時的に又は恒久的に記憶する。

[0039] （３）制御部１３０

制御部１３０は、処理装置１００の様々な機能を提供する。制御部１３０は、取得部１３１、演算部１３３、学習部１３５及び通知部１３７を含む。なお、制御部１３０は、これらの構成要素以外の他の構成要素をさらに含み得る。即ち、制御部１３０は、これらの構成要素の動作以外の動作も行い得る。

[0040] 各構成要素の動作を簡易に説明する。取得部１３１は、情報を取得する。演算部１３３は、後述するオートエンコーダを用いて各種演算を行う。学習部１３５は、後述するオートエンコーダに関する学習を行う。通知部１３７は、演算部１３３による演算結果を示す情報を端末装置２００に通知する。その他の詳細な動作は、後に詳しく説明する。

[0041] ＜２．３．端末装置の構成例＞

図３は、本実施形態に係る端末装置２００の論理的な構成の一例を示すブロック図である。図３に示すように、端末装置２００は、入力部２１０、出力部２２０、通信部２３０、記憶部２４０及び制御部２５０を含む。

[0042] （１）入力部２１０

入力部２１０は、情報の入力を受け付ける機能を有する。例えば、入力部２１０は、ユーザからの情報の入力を受け付ける。入力部２１０は、例えばキーボード又はタッチパネル等による文字入力を受け付けてもよいし、音声入力を受け付けてもよいし、ジェスチャ入力を受け付けてもよい。その他、入力部２１０は、フラッシュメモリ等の記憶媒体からのデータ入力を受け付けてもよい。

[0043] （２）出力部２２０

出力部２２０は、情報の出力を行う機能を有する。例えば、出力部２２０は、画像、音声、振動、又は発光等により情報を出力する。

[0044] （３）通信部２３０

通信部２３０は、情報を送受信する機能を有する。例えば、通信部２３０

は、処理装置１００からの情報を受信し、処理装置１００への情報を送信する。

[0045] （４）記憶部２４０

記憶部２４０は、端末装置２００の動作のためのプログラム及び様々なデータを一時的に又は恒久的に記憶する。

[0046] （５）制御部２５０

制御部２５０は、端末装置２００の様々な機能を提供する。制御部２５０は、通知部２５１及び取得部２５３を含む。なお、制御部２５０は、これらの構成要素以外の他の構成要素をさらに含み得る。即ち、制御部２５０は、これらの構成要素の動作以外の動作も行い得る。

[0047] 各構成要素の動作を簡易に説明する。通知部２５１は、入力部２１０に入力されたユーザ入力を示す情報を処理装置１００に通知する。取得部２５３は、処理装置１００による演算結果を示す情報を取得して、出力部２２０により出力させる。その他の詳細な動作は、後に詳しく説明する。

[0048] ＜＜３．基本的な技術的特徴＞＞

続いて、本実施形態に係るシステム１の基本的な技術的特徴を説明する。

[0049] （１）オートエンコーダ

オートエンコーダにおける中間変数ｚは、データｘを表現するための特徴量とも捉えられる。中間変数ｚを特徴量と捉える方式は、データｘを表現する特徴量が設計困難な場合に有効である。設計困難とは、例えば、データとして取得することが困難である、数値化することが困難である、複数の概念が融合され且つその組み合わせ方が不明な概念である、又はそもそも概念として表現し切れない等を指す。このような場合、特徴量をニューラルネットにより表現して、特徴量の取り扱いをニューラルネットに任せてしまうことが望ましい。

[0050] 一方で、データｘを表現するための情報として、設計容易な特徴量も考えられる。設計容易とは、例えば概念として容易に表現でき且つ当該概念に対応する特徴量の数値化も容易であることを指す。設計容易な特徴量は、対応

する概念（以下、ラベルとも称する）が有る特徴量であるから、以下ではラベル有り特徴情報（意味付けされた第１の特徴情報に相当）とも称する。一方で、中間変数ｚにより表現される設計困難な特徴量は、対応するラベルが無い特徴量であるから、以下ではラベル無し特徴情報（意味付けされない第２の特徴情報に相当）とも称する。

[0051] ラベル有り特徴情報は、ラベル無し特徴情報とは別に、ニューラルネットに入力されてもよい。この場合、生成ネットを示す数式（２）は、次式のように変更される。

[0052] [数5]

$$x' = g_\theta(y, z) \quad \cdots (5)$$

[0053] ここで、ｙはラベル有り特徴情報であり、ｚはラベル無し特徴情報である。

[0054] 同様に、推論ネットを示す数式（１）は、次式のように変更される。

[0055] [数6]

$$\boldsymbol{z} = \boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) \quad \cdots (6)$$

[0056] さらに、データｘから、ラベル有り特徴情報ｙを推定するモデルも考えられる。これを識別ネットとすると、識別ネットは、以下の関数で表現される。

[0057] [数7]

$$\boldsymbol{y} = \boldsymbol{h}_{\boldsymbol{\phi}}(\boldsymbol{x}) \quad \cdots (7)$$

[0058] 以上説明した、識別ネット、推論ネット、及び生成ネットを組み合わせたオートエンコーダを、図４に示した。

[0059] 図４は、本実施形態に係るオートエンコーダの概要を説明するための図である。図４に示すように、オートエンコーダ１０は、識別ネット１１、推論ネット１３、及び生成ネット１５を含む。

[0060] 識別ネット１１は、データｘからラベル有り特徴情報ｙを識別する識別部

である、上記数式（７）に示した関数h_{ϕ}（・）に対応する識別器１２を含む。

[0061] 推論ネット１３は、データｘ及びラベル有り特徴情報ｙからラベル無し特徴情報ｚを推論する推論部である、上記数式（６）に示した関数f_{ϕ}（・）に対応する推論器１４を含む。推論ネット１３は、データｘから、ラベル有り特徴情報ｙを除去して、ラベル無し特徴情報ｚを抽出する、とも捉えることができる。

[0062] 生成ネット１５は、ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚから、ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚが示す特徴を有するデータｘを生成する生成部である、上記数式（５）に示した関数g_{θ}（・）に対応する生成器１６を含む。例えば、ラベル有り特徴情報ｙは１０次元の特徴量であり、ラベル無し特徴情報ｚは５０次元の特徴量であり得る。この場合、各々の特徴量が直列的に連結された、６０次元の特徴量が生成器１６へ入力されることとなる。

[0063] （２）各種情報

・データｘ

データｘは、時系列データである。

[0064] ・ラベル有り特徴情報ｙ

ラベル有り特徴情報ｙは、時系列データｘの特徴を示す多次元の特徴量である。

[0065] 例えば、ラベル有り特徴情報ｙは、時系列データｘの枠組みを示す情報を含み得る。枠組みを示す情報とは、対象の時系列データｘの基礎的な構造を示す情報である。枠組みを示す情報を、以下では枠組み情報とも称する。

[0066] また、ラベル有り特徴情報ｙは、概念として容易に表現できる特徴として、時系列データｘのコンテキストを示す情報を含み得る。コンテキスト情報とは、対象の時系列データｘが発生する状況を示す情報である。コンテキストを示す情報を、以下ではコンテキスト情報とも称する。

[0067] ・ラベル無し特徴情報ｚ

ラベル無し特徴情報ｚは、時系列データｘの特徴を示す多次元の特徴量である。

[0068] （３）学習

システム１（例えば、学習部１４５）は、ニューラルネット（即ち、推論器、識別器、及び生成器の各々）のパラメータを学習する。例えば、システム１は、所定の目的関数を最適化するように学習する。

[0069] ここで、学習データセットＸは、$X_l = \{x_n, y_n\}$のように、時系列データｘにラベル有り特徴情報ｙが対応付けられていてもよいし、$X_u = \{x_n\}$のようにラベル有り特徴情報ｙが対応付けられていなくてもよい。後者の場合、システム１は、識別ネット１１によりラベル有り特徴情報ｙを推定して、時系列データｘに対応付けてもよい。その他、ラベル有り特徴情報ｙと時系列データｘとの対応付けは、ユーザにより行われてもよい。

[0070] 他にも、システム１は、ラベル無し特徴情報ｚを学習してもよい。この場合の学習とは、上記パラメータ学習後の推論ネット１３により、時系列データｘに対応するラベル無し特徴情報ｚを推論することを指す。ラベル有り特徴情報ｙに関しても同様である。

[0071] 学習結果は、例えば記憶部１２０に記憶され得る。

[0072] （４）時系列データの生成

システム１（例えば、演算部１３３）は、上記学習後のニューラルネットを用いて、時系列データｘを生成する。詳しくは、システム１は、ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚを生成器１６に入力することで、ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚが示す特徴を有する時系列データｘを生成する。これにより、システム１は、設計困難な特徴量及び設計容易な特徴量の双方を反映した時系列データｘを生成することが可能となる。

[0073] システム１（例えば、取得部１３１）は、生成に用いられるラベル有り特徴情報ｙ及びラベル無し特徴情報ｚを取得してもよい。取得元は、他の装置（例えば、端末装置２００）、又は記憶部１２０等、多様に考えられる。ま

た、生成に用いられるラベル有り特徴情報ｙは、識別器１２から出力されたものであってもよいし、生成に用いられるラベル無し特徴情報ｚは、推論器１４から出力されたものであってもよい。

[0074] （５）処理の流れ

続いて、図５～図７を参照して、システム１により実行される各種処理の流れの一例を説明する。

[0075] 図５は、本実施形態に係る処理装置１００において実行される学習処理の流れの一例を示すフローチャートである。図５に示すように、まず、取得部１３１は、学習データセットを取得する（ステップＳ１０２）。次いで、学習部１３５は、識別器１２により、学習データセットに含まれる時系列データからラベル有り特徴情報ｙを識別する（ステップＳ１０４）。次に、学習部１３５は、推論器１４により、学習データセットに含まれる時系列データ、及び学習データセットに含まれる又は識別器１２により識別されたラベル有り特徴情報ｙからラベル無し特徴情報ｚを推論する（ステップＳ１０６）。次いで、学習部１３５は、生成器１６により、学習データセットに含まれる又は識別器１２により識別されたラベル有り特徴情報ｙ、及び推論されたラベル無し特徴情報ｚから時系列データｘを生成する（ステップＳ１０８）。次に、学習部１３５は、学習を終了するか否かを判定する（ステップＳ１１０）。例えば、学習部１３５は、学習データセットに含まれる時系列データｘと、生成された時系列データｘとの差異が所定の閾値以下である場合に終了すると判定し、超える場合は終了しないと判定する。ラベル有り特徴情報ｙに関しても同様である。終了しないと判定された場合、学習部１３５は、ニューラルネット（識別器１２、推論器１４及び／又は生成器１６）のパラメータ（例えば、重みパラメータ及びバイアス）を更新する（ステップＳ１１２）。その後、処理は再度ステップＳ１０４に戻る。終了すると判定された場合、記憶部１２０は、学習結果を示す情報を記憶する（ステップＳ１１４）。

[0076] 図６は、本実施形態に係るシステム１において実行されるインタラクショ

ン処理の流れの一例を示すシーケンス図である。本シーケンスには、処理装置１００及び端末装置２００が関与する。図６に示すように、端末装置２００（例えば、入力部２１０）は、ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚのユーザ入力を受け付ける（ステップＳ２０２）。次いで、端末装置２００（例えば、通知部２５１）は、ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚを処理装置１００へ通知し、処理装置１００（例えば、取得部１３１）が取得する（ステップＳ２０４）。次に、処理装置１００（例えば、演算部１３３）は、通知されたラベル有り特徴情報ｙ及びラベル無し特徴情報ｚを用いて時系列データｘを生成する（ステップＳ２０６）。次いで、処理装置１００（例えば、通知部１３７）は、生成した時系列データｘを端末装置２００へ通知し、端末装置２００（例えば取得部２５３）が取得する（ステップＳ２０８）。そして、端末装置２００（例えば、出力部２２０）は、取得した時系列データｘを出力する（ステップＳ２１０）。

[0077] 図７は、本実施形態に係るシステム１において実行されるインタラクション処理の流れの一例を示すシーケンス図である。本シーケンスには、処理装置１００及び端末装置２００が関与する。図７に示すように、端末装置２００（例えば、入力部２１０）は、時系列データｘのユーザ入力を受け付ける（ステップＳ３０２）。次いで、端末装置２００（例えば、通知部２５１）は、時系列データｘを処理装置１００へ通知し、処理装置１００（例えば、取得部１３１）が取得する（ステップＳ３０４）。次に、処理装置１００（例えば、演算部１３３）は、時系列データｘからラベル有り特徴情報ｙを識別し、ラベル無し特徴情報ｚを推論する（ステップＳ３０６）。次いで、処理装置１００（例えば、演算部１３３）は、例えばユーザ入力への応答のためにラベル有り特徴情報ｙを変更して（ステップＳ３０８）、ラベル無し特徴情報ｚ及び変更後のラベル有り特徴情報ｙを用いて時系列データｘを生成する（ステップＳ３１０）。次に、処理装置１００（例えば、通知部１３７）は、生成した時系列データｘを端末装置２００へ通知し、端末装置２００（例えば取得部２５３）が取得する（ステップＳ３１２）。そして、端末装

置２００（例えば、出力部２２０）は、取得した時系列データｘを出力する（ステップＳ３１４）。

[0078] 以上、本実施形態に係る処理装置１００の基本的な技術的特徴を説明した。続いて、処理装置１００の多様な実施形態を具体的に説明する。

[0079] ＜＜４．第１の実施形態＞＞

本実施形態は、時系列データｘが話声を含む音声データである形態である。

[0080] ＜４．１．各種データ＞

・時系列データｘ

時系列データｘは、テキスト情報を読み上げた音声データである。

[0081] ・ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚ

ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚは、音声データｘの特徴を示す情報である。これらの特徴情報をまとめて、音声特徴情報とも称する。例えば、音声特徴情報は、声紋に代表される発話者の個性に関する特徴、訛り又はイントネーション等の発話者のバックグランドに関する特徴、及び感情等の特徴を表現する。

[0082] ラベル有り特徴情報ｙのうちコンテキスト情報は、話者のコンテキストを示す情報を含み得る。コンテキスト情報は、例えば個人を識別するラベル、世代（例えば、子供、大人）、男女、訛り（例えば、話者の出身地域）、状態（例えば、喜怒哀楽等の感情）、職業（例えば、アナウンサー、スチュワーデス、エレベータ案内係等）等の話者自身に関する情報であってもよい。また、コンテキスト情報は、例えば話し相手（例えば、話者の子供、夫等）、話している地域又は場所（例えば、職場、学校、自宅等）等の話者が話す環境に関する情報であってもよい。また、コンテキスト情報は、話の前後のコンテキスト、トピック（例えば、世間話、深刻な相談等）等の、話者が話す話題に関する情報であってもよい。

[0083] また、ラベル有り特徴情報ｙのうち枠組み情報は、例えばスピーチの原稿又は映画のセリフ等のテキスト情報である。この場合、枠組み情報は、例え

ばアルファベット２６文字に対応する２６次元の情報となり得る。

[0084] ラベル無し特徴情報ｚは、音声特徴情報のうち、ラベル有り特徴情報ｙ以外の特徴を示す。本実施形態では、ラベル無し特徴情報ｚを発話スタイル情報とも称する

[0085] ＜４．２．ニューラルネットの構成＞

本実施形態に係るニューラルネットの構成を、図８～図１０を参照して説明する。

[0086] 図８は、本実施形態に係る識別ネット１１の構成の一例を示す図である。図８に示すように、識別器１２は、音声データからテキスト情報及びコンテキスト情報を識別する。

[0087] 図９は、本実施形態に係る推論ネット１３の構成の一例を示す図である。図９に示すように、推論器１４は、テキスト情報、コンテキスト情報及び音声データから発話スタイル情報を推論する。

[0088] 図１０は、本実施形態に係る生成ネット１５の構成の一例を示す図である。図９に示すように、生成器１６は、テキスト情報、コンテキスト情報及び発話スタイル情報から音声データを生成する。

[0089] ＜４．３．学習＞

システム１（例えば、学習部１４５）は、ニューラルネット（即ち、推論器、識別器、及び生成器の各々）のパラメータを学習する。例えば、システム１は、所定の目的関数を最適化するように学習する。

[0090] また、システム１は、音声特徴情報を学習してもよい。この場合の学習とは、上記パラメータ学習後の推論器１４により、音声データに対応する発話スタイル情報を推論することを指す。他にも、図１１に示すように、システム１は、識別ネット１１と生成ネット１５とを連結して発話スタイル情報を推論することで、学習を行ってもよい。なお、図１１は、本実施形態に係る推論処理の一例を説明するための図である。ここで注目すべきは、識別器１２から出力されたテキスト情報及びコンテキスト情報が推論器１４に入力される点と、それに伴い入力が音声データのみである点である。これにより、

ユーザの操作負荷が軽減される。システム１は、学習した音声特徴情報を識別情報と対応付けて記憶してもよい。これにより、後述する話者の選択による音声特徴情報の編集が可能となる。

[0091] ＜４．４．音声データの生成＞

（１）ＴＴＳ

システム１（例えば、演算部１３３）は、テキスト情報から音声データを生成する。テキスト情報から音声データの生成には、例えば、テキストＴｏスピーチ（ＴＴＳ：Ｔｅｘｔ　Ｔｏ　Ｓｐｅｅｃｈ）変換技術が用いられる。ＴＴＳは、テキストから音声データを自動的に生成する技術であり、声優等に発話させて録音する場合と比較して安価に音声データを生成可能である。一般的なＴＴＳエンジンでは、機械的な音声が生成されるのみであって、例えば訛りが反映された音声は生成されない。そこで、システム１は、テキスト情報に加えて音声特徴情報を入力する生成器１６をＴＴＳエンジンに代えて用いることで、例えば訛りが反映された音声データを生成する。

[0092] ユーザは、音声データ化を所望するテキスト情報、及び所望する音声特徴情報をシステム１に入力することで、所望する特徴を有する音声でテキスト情報が読み上げられた音声データを生成させることが可能となる。

[0093] （２）音声特徴情報の編集

ユーザは、音声特徴情報を編集可能である。例えば、ユーザは、端末装置２００を介して音声特徴情報を編集する。音声特徴情報の編集は、関数又はコマンドラインインタフェースにより行われてもよいし、グラフィカルインタフェース等により行われてもよい。一例として、グラフィカルインタフェースの例を図１２に示す。

[0094] －ＧＵＩ

図１２は、本実施形態に係るユーザインタフェースの一例を示す図である。図１２に示す音声特徴情報編集画面２０は、例えば端末装置２００により表示され、ユーザによる音声特徴情報の編集操作を受け付ける。図１２に示すように、音声特徴情報編集画面２０は、発話スタイル情報である多次元の

特徴量z_0～z_{49}の調節部２１、及びコンテキスト情報である多次元の特徴量の調節部２２を含む。コンテキスト情報である多次元の特徴量には、それぞれ「喜」「怒」及び「哀」、並びに「地域Ａ」「地域Ｂ」及び「地域Ｃ」といったラベルが対応付けられる。ラベルは、「喜」「怒」及び「哀」が「感情」に、「地域Ａ」「地域Ｂ」及び「地域Ｃ」が「訛り」に、それぞれカテゴリ分けされてもよい。ユーザは、各々の特徴量の値を示すつまみ２３を上下させることで、各々の特徴量を調節可能である。

[0095] 図１２に示すように、ユーザは、調節部２２を操作することで、感情又は訛りといった概念に対応する特徴量を、所望する特徴を有する音声データが生成されるよう意図的に調節することが可能である。また、ユーザは、調節部２１を操作することで、所望する特徴を有する音声データが生成されるよう試行錯誤することが可能となる。このようにして、ユーザは、テキスト情報以外の音声データの特徴を簡単に制御することが可能である。別の観点から言えば、ユーザは、有限個の音声特徴情報という圧縮された情報で音声データの特徴を制御することが可能となるので、操作負荷が軽減される。

[0096] 上記説明したＧＵＩによる編集は、細やかな編集が可能である。そのため、ＧＵＩによる編集は、プロフェッショナルなユーザによる音声コンテンツの作成に有用である。適用場面の一例として、例えばアニメーション又はＣＧ（Computer Graphics）作成過程における、シーン及びセリフの作成、並びに動画コンテンツの作成後の、音声コンテンツの作成が挙げられる。

[0097] －話者の選択

音声特徴情報の編集操作は、ユーザによる話者の選択により行われてもよい。この場合、ユーザの操作負荷がさらに軽減される。

[0098] 例えば、システム１（例えば、記憶部１２０）は、予め話者の識別情報と音声特徴情報との組み合わせを、音声特徴情報ＤＢとして記憶しておく。図１３は、本実施形態に係る音声特徴情報ＤＢの一例を示す図である。図１３に示すように、システム１は、各々の話者の識別情報と、音声特徴情報とを対応付けて記憶する。なお、話者の識別情報は、音声特徴情報の識別情報と

も捉えられる。

[0099] 図１４は、本実施形態に係るユーザインタフェースの一例を示す図である。図１４に示す話者選択画面３０は、例えば端末装置２００により表示され、ユーザによる話者の選択操作を受け付ける。例えば、ユーザは、図１４に示した話者選択画面３０において、「話者Ｘ」を選択したものとする。すると、処理装置１００（例えば、取得部１４１）は、「話者Ｘ」の識別情報（即ち、音声特徴情報の識別情報）を取得する。そして、処理装置１００（例えば、演算部１３３）は、当該識別情報に対応する音声特徴情報が示す特徴を有する音声データを生成する。

[0100] なお、話者の選択肢は、個人の名称であってもよいし、グループの名称であってもよい。前者の一例として、例えば話者の選択肢は、現実の存在する人物であってもよい。この場合、現実に存在する人物があたかもテキストを読み上げたかのような音声データが生成される。後者の一例として、話者の選択肢は、「アナウンサー」、「バスガイド」、「ヤクザ」、「教師」、「医者」、「歌舞伎役者」、「落語家」、「執事」等の、職業名であってもよい。その場合、各々の職業の人の代表的な話し方でテキストが読み上げられた音声データが生成される。同様に、話者の選択肢は、映画又はアニメ作品等の登場人物であってもよい。また、話者の選択肢は、「妖精」、「妖怪」、「モンスター」等のファンタジー作品における種族、「魔法使い」等のファンタジー作品における職業等であってもよい。

[0101] なお、図１４に示した話者選択画面３０により話者が選択された後、図１２に示した音声特徴情報編集画面２０により、音声特徴情報が調整されてもよい。例えば、音声特徴情報編集画面２０において訛りの音声特徴情報が編集された場合、選択した人物が、本人とは異なる訛りでテキストを読み上げたかのような音声データが生成され得る。

[0102] 上記説明した話者選択による編集は、ＧＵＩによる編集と比較して簡易である。そのため、話者選択による編集は、デフォルトの機械的な音声では満足できないものの、ある程度のカスタマイズで満足するライトユーザにとっ

て有用である。適用場面の一例として、音声で情報を提供する音声アシストエージェントの音声の切り替えが挙げられる。

[0103] －ユーザ音声での指示

音声特徴情報の編集操作は、ユーザ音声の認識結果に基づいて行われてもよい。この場合、ユーザの操作負荷がさらに軽減される。

[0104] 例えば、コンテキスト情報は、ユーザ音声による指示に応じて編集され得る。具体的には、ユーザは、コンテキスト情報の調節指示を発話することで、コンテキスト情報を編集する。具体的には、ユーザが「もっと嬉しそうにしゃべって」と発話した場合、システム１（例えば、演算部１３３）は、「喜」に対応する特徴量の値を上昇させる。また、システム１は、ユーザが指示した編集操作を支援するために、ユーザが指示した特徴量に反する効果を有する特徴量をも制御してもよい。例えば、ユーザが「もっと嬉しそうにしゃべって」と発話した場合、システム１は、「怒」「哀」「楽」に対応する特徴量の値を下降させてもよい。

[0105] また、ユーザは、音声特徴情報の良し悪しをフィードバックすることで、音声特徴情報を編集してもよい。具体的には、システム１は、発話毎等の適切な粒度で音声特徴情報を変更しつつ音声データを出力し、ユーザは「良い」又は「悪い」とフィードバックする。すると、システム１は、「良い」とフィードバックされた音声特徴情報に近く、且つ「悪い」とフィードバックされた音声特徴情報から遠くなるよう、各々の特徴量を制御する。

[0106] －ユーザ音声に応じた自動編集

音声特徴情報の編集は、ユーザにより明示的に行われずとも、ユーザ音声に応じて自動的に行われてもよい。例えば、音声データの生成に用いられる音声特徴情報は、ユーザ音声が有する特徴に対応する特徴を示す特徴情報であってもよい。この場合、ユーザが音声特徴情報を編集せずとも好みの音声データが生成されるので、ユーザの操作負荷がさらに軽減される。

[0107] 例えば、システム１（例えば、演算部１３３）は、ユーザ音声の特徴を示す音声特徴情報をそのまま用いて音声データを生成してもよい。そのために

は、システム１は、識別器１２及び推論器１４を用いてユーザ音声から音声特徴情報（例えば、コンテキスト情報及び発話スタイル情報）を識別又は推論し、新たに取得された応答のためのテキスト情報に基づいて音声データを生成する。このような自動編集により、システム１は、例えば、ユーザの訛りと同様の訛りでユーザ音声に応答したり、ユーザの話す速度と同一の速度でユーザ音声に応答したりすることが可能となる。

[0108] また、システム１は、ユーザの属性又は状態等に応じて音声特徴情報を制御してもよい。例えば、システム１は、ユーザの「性別」「年齢」「誰であるか」「喜怒哀楽」「出身地」「人種」「システムへの集中度合い」等に応じた音声特徴情報を用いて音声データを生成してもよい。なお、これらの属性又は状態は、予め用意されたユーザデータベース、又は画像認識若しくは音声認識等の結果により取得され得る。このような自動編集により、システム１は、例えば子供に対してはゆっくり応答し、お年寄りには大きな音量で応答することが可能となる。

[0109] ユーザ音声に応じた自動編集の適用場面の一例として、翻訳場面が挙げられる。

[0110] 例えば、外国映画の吹き替え又はスピーチの同時翻訳等は、通常、本来の話者とは異なる話者により行われる。そのため、異なる話者の音声からは、本来の話者の音声に含まれていた声紋及び喜怒哀楽等のニュアンスが失われる場合が多い。この点、ユーザ音声に応じた自動編集が適用されることで、システム１は、ユーザ音声に含まれるテキストを翻訳しつつ、ユーザ音声と同様の音声特徴情報を反映させた、音声データを生成することが可能である。よって、翻訳された音声に、本来の音声に含まれていた声紋及び喜怒哀楽等のニュアンスを残すことが可能となる。これにより、映画又はスピーチ等の翻訳音声を、本来の話者の声色で聞いてみたい、又は話者の感情等を音声だけからリアルに感じ取りたい、というユーザの潜在的な欲求を満たすことが可能となる。以下、図１５を参照して、翻訳場面における処理装置１００による処理を詳しく説明する。

[0111] 図１５は、本実施形態に係る翻訳場面における音声データ生成処理を説明するための図である。図１５に示すように、システム１は、識別器１２、推論器１４、生成器１６及び翻訳器４２を用いて、音声データから翻訳された音声データを生成する。まず、システム１（例えば、取得部１４１）は、翻訳対象の音声データｘを適切な区間に区切って取得する。次いで、システム１（例えば、演算部１３３）は、翻訳対象の音声データｘを識別器１２に入力して、テキスト情報ｙ及びコンテキスト情報ｙを推定する。なお、テキスト情報ｙ及びコンテキスト情報ｙの推定は、識別器１２以外にも、人により行われてもよいし、音声認識器等により行われてもよい。そして、システム１は、音声データｘ、テキスト情報ｙ及びコンテキスト情報ｙを推論器１４に入力して、発話スタイル情報ｚを推定する。他方、システム１は、テキスト情報ｙを翻訳器４２に入力して、テキスト情報ｙが所望の言語に翻訳されたテキスト情報ｙ´を推定する。なお、翻訳は、翻訳器４２以外にも人により行われてもよい。そして、システム１は、翻訳されたテキスト情報ｙ´及び発話スタイル情報ｚを生成器１６に入力して、翻訳された音声データｘ´を生成する。このとき、もちろんシステム１は、コンテキスト情報ｙも生成器１６に入力してもよい。

[0112] なお、翻訳場面では、国ごとに訛りの傾向は異なるので、例えば英語の訛りを日本語の訛りに適切に変換することは困難な場合がある。そのため、処理装置１００は、生成器１６に入力する音声特徴情報から訛りの特徴量を取り除いてもよい。

[0113] ＜４．５．補足＞

上記では、テキスト情報を枠組み情報として音声データを生成する例を説明したが、本技術はかかる例に限定されない。

[0114] 例えば、システム１は、楽譜を枠組み情報として、楽譜に従った演奏された音楽データを生成してもよい。この場合、識別器１２は、音楽データから楽譜情報を識別する、自動採譜機能を有する。また、推論器１４は、音楽データ及び楽譜情報から、例えば、演奏者の特徴、楽器の特徴、演奏会場の音

響の特徴等を示す音楽特徴情報を推論する。なお、これらの音楽特徴情報の少なくとも一部は、識別器１２により識別されてもよい。また、生成器１６は、楽譜情報及び音楽特徴情報から音楽データを生成する。これにより、システム１は、例えば楽譜、演奏者、楽器又は演奏会場等を変更した音楽データを生成することが可能となる。

[0115] 具体的には、音響状態の良くない場所で録音された音楽データに関し、音楽特徴情報を音響状態の良い場所のものに変更することで、音楽データの音響状態を改善することが可能である。同様に、演奏スタイル、曲の入れ替え、歌手又は歌詞の入れ替え等も可能となる。あるいは、作曲スタイルを取り込んだ楽曲の自動作成も可能となる。

[0116] ＜＜５．第２の実施形態＞＞

本実施形態は、時系列データｘが、動物体をセンサによりセンシングしたセンサデータである形態である。

[0117] ＜５．１．各種データ＞

・時系列データｘ

時系列データｘは、動物体をセンシングしたセンサデータである。センシング対象の動物体は、人、動物、車、昆虫等多様に考えられる。センシングするセンサは、例えば加速度センサ、又はジャイロセンサ等の任意のセンサであってもよい。その他、センサはイメージセンサであり、時系列データｘは動画データであってもよい。

[0118] ・ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚ

ラベル有り特徴情報ｙ及びラベル無し特徴情報ｚは、センシング対象の動物体の動作の特徴を示す情報である。これらの特徴情報をまとめて、動作特徴情報とも称する。例えば、動作特徴情報は、動物体の動作の種別、動作の大きさ、及び周波数等の特徴を表現する。

[0119] ラベル有り特徴情報ｙのうちコンテキスト情報は、例えば動物体の動作種別を示す情報である。動作種別としては、動物体が人である場合、例えば「立つ」「座る」「歩く」「走る」等が挙げられる。他にも、動作種別は、動

物体が昆虫である場合の「飛ぶ」「仲間を呼ぶ」等が挙げられる。以下では、このような動作種別を示す情報を、動作種別情報とも称する。

[0120] ラベル有り特徴情報ｙのうち枠組み情報は、例えば動物体の行動の構成を示す情報である。行動の構成を示す情報とは、例えば１０秒歩いて１０秒走る、といった情報である。なお、本実施形態では、説明を簡略化するため、枠組み情報に関しては説明を省略する。

[0121] ラベル無し特徴情報ｚは、動作特徴情報のうち、ラベル有り特徴情報ｙ以外の特徴を示す。本実施形態では、ラベル無し特徴情報ｚを動作スタイル情報とも称する

[0122] ＜５．２．ニューラルネットの構成＞

本実施形態に係るニューラルネットの構成を、図１６～図１８を参照して説明する。

[0123] 図１６は、本実施形態に係る識別ネット１１の構成の一例を示す図である。図１６に示すように、識別器１２は、センサデータから動作種別情報を識別する。

[0124] 図１７は、本実施形態に係る推論ネット１３の構成の一例を示す図である。図１７に示すように、推論器１４は、動作種別情報及びセンサデータから動作スタイル情報を推論する。

[0125] 図１８は、本実施形態に係る生成ネット１５の構成の一例を示す図である。図１８に示すように、生成器１６は、動作種別情報及び動作スタイル情報からセンサデータを生成する。

[0126] ＜５．３．学習＞

システム１（例えば、学習部１４５）は、ニューラルネット（即ち、推論器、識別器、及び生成器の各々）のパラメータを学習する。例えば、システム１は、所定の目的関数を最適化するように学習する。この学習に関する処理の流れの一例を、図１９を参照して説明する。

[0127] 図１９は、本実施形態に係る処理装置１００において実行される学習処理の流れの一例を示すフローチャートである。図１９に示すように、まず、取

得部１３１は、動物体をセンシングしたセンサデータＸ＝｛ｘ$_n$｝、及び当該動物体の動作種別情報Ｙ＝｛ｙ$_n$｝の組み合わせから成る学習データセットＤ＝｛Ｘ，Ｙ｝を取得する（ステップＳ４０２）。次いで、学習部１３５は、次式に従い、目的関数Ｌを最小化するようニューラルネットのパラメータを学習する（ステップＳ４０４）。

[0128] [数8]

$$w,u,v = argminL(X,Y;h_w,f_u,g_v) \quad \cdots（８）$$

[0129] ここで、ｗは識別器１２のパラメータであり、ｕは推論器１４のパラメータであり、ｖは生成器１６のパラメータである。また、目的関数Ｌは、上記数式（３）に示した形式であってもよい。そして、記憶部１２０は、学習結果を示す情報を記憶する（ステップＳ４０６）。

[0130] また、システム１は、動作特徴情報を学習してもよい。この場合の学習とは、上記パラメータ学習後の識別器１２又は推論器１４により、センサデータに対応する動作種別情報又は動作スタイル情報を識別又は推論することを指す。システム１は、学習した動作特徴情報を識別情報と対応付けて記憶してもよい。これにより、後述する識別情報の選択による動作スタイル情報の編集が可能となる。

[0131] ＜５．４．センサデータの生成＞

（１）生成

システム１（例えば、演算部１３３）は、生成器１６を用いて、動作種別情報及び動作スタイル情報からセンサデータ（より正確には、疑似的なセンサデータ）を生成する。ユーザは、所望する動作種別情報及びスタイル情報をシステム１に入力することで、所望するセンサデータを生成させることが可能となる。

[0132] （２）動作特徴情報の編集

ユーザは、動作特徴情報を編集可能である。例えば、ユーザは、端末装置２００を介して動作特徴情報を編集する。動作特徴情報の編集は、上述した

音声特徴情報の編集と同様にして行われ得る。一例として、図２０を参照してＧＵＩに関して説明し、他については説明を省略する。

[0133] 図２０は、本実施形態に係るユーザインタフェースの一例を示す図である。図２０に示す動作特徴情報編集画面５０は、例えば端末装置２００により表示され、ユーザによる動作特徴情報の編集操作を受け付ける。図２０に示すように、動作特徴情報編集画面５０は、動作スタイル情報である多次元の特徴量z_0～z_{49}の調節部５１、及び動作種別情報である多次元の特徴量の調節部５２を含む。動作種別情報である多次元の特徴量には、それぞれ「歩く」「走る」及び「立つ」といったラベルが対応付けられる。ラベルは、「歩く」「走る」及び「立つ」が「人の動作」にカテゴリ分けされてもよい。ユーザは、各々の特徴量の値を示すつまみ５３を上下させることで、各々の特徴量を調節可能である。

[0134] ＜５．５．制御信号の生成＞

（１）概要

動物体は、アクチュエータを有するロボット又は車等の装置であってもよい。この場合、システム１は、動物体に含まれるアクチュエータを動作させるための制御信号を、時系列データｘとして生成してもよい。

[0135] 例えば、システム１（例えば、演算部１３３）は、センサデータに対応する制御信号を生成する。制御信号と当該制御信号に基づいて動作するアクチュエータを有する装置をセンシングしたセンサデータとの対応関係が既知であれば、システム１は、当該装置に所望の動作を行わせるための制御信号を生成することが可能となる。このような制御信号を生成するための生成ネット１５の構成の一例を、図２１を参照して説明する。

[0136] 図２１は、本実施形態に係る生成ネット１５の構成の一例を示す図である。図２１に示すように、本実施形態に係る生成ネット１５において、生成器１６に制御信号生成器６２が連結される。制御信号生成器６２は、センサデータｘから対応する制御信号ｓを生成する機能を有する。よって、生成ネット１５は、動作種別情報ｙ及び動作スタイル情報が示す特徴を有する動作を

アクチュエータを有する装置に行わせるための、制御信号ｓを生成することが可能である。

[0137] （２）学習

制御信号生成器６２は、ニューラルネット、より具体的にはフィードフォワードネットワークにより構成されていてもよい。制御信号生成器６２のパラメータの学習処理の流れの一例を、図２２を参照して説明する。

[0138] 図２２は、本実施形態に係る処理装置１００において実行される学習処理の流れの一例を示すフローチャートである。図２２に示すように、まず、取得部１３１は、アクチュエータに入力した制御信号Ｓ＝｛s_n｝、及び当該アクチュエータを有する装置をセンシングしたセンサデータＸ＝｛x_n｝の組み合わせから成る学習データセットＤ＝｛Ｘ，Ｓ｝を取得する（ステップＳ５０２）。次いで、学習部１３５は、次式に従い、目的関数Ｌを最小化するようニューラルネットのパラメータを学習する（ステップＳ５０４）。

[0139] [数9]

$$a = argminL(X, S; h_a) \quad \cdots (9)$$

[0140] ここで、ａは制御信号生成器６２のパラメータである。また、目的関数Ｌは、上記数式（３）に示した形式であってもよい。そして、記憶部１２０は、学習結果を示す情報を記憶する（ステップＳ５０６）。

[0141] なお、制御信号生成器６２と他のニューラルネット（例えば、識別器１２、推論器１４及び／又は生成器１６）とは、別箇に学習が行われてもよいし、まとめて学習が行われてもよい。

[0142] （３）生成

システム１（例えば、演算部１３３）は、生成器１６及び制御信号生成器６２を用いて、動作種別情報及び動作スタイル情報から制御信号を生成する。ユーザは、所望する動作種別情報及びスタイル情報をシステム１に入力することで、アクチュエータを有する装置に所望する動作を行わせるための制御信号を生成させることが可能となる。

[0143] システム１は、図２１に示した生成ネット１５を用いて制御信号を生成してもよいし、生成ネット１５に他の構成要素を組み合わせて制御信号を生成してもよい。その一例を、図２３～図２５を参照して説明する。

[0144] 図２３は、本実施形態に係る制御信号生成処理の一例を説明するための図である。図２３に示すように、システム１は、図２１を参照して説明した生成ネット１５に、図１７を参照して説明した推論ネット１３を組み合わせて制御信号を生成してもよい。図２３に示した構成によれば、推論器１４は、センシング対象の動物体の動作スタイル情報を推論し、生成器１６に入力する。そして、生成器１６はセンサデータを生成し、制御信号生成器６２は制御信号を生成する。センシング対象の動物体は、アクチュエータを有する装置でない、例えば人であってもよい。この場合、システム１は、人をセンシングしたセンサデータに基づいて、例えば人型ロボットに人と同様の動作を行わせることが可能となる。

[0145] 図２４は、本実施形態に係る制御信号生成処理の一例を説明するための図である。図２４に示すように、システム１は、図２１を参照して説明した生成ネット１５に、動作スタイル情報選択器６４を組み合わせて制御信号を生成してもよい。動作スタイル情報選択器６４は、例えばユーザにより選択された識別情報に対応する動作スタイル情報を出力する機能を有する。識別情報と動作スタイル情報との対応付けは、上述した学習により行われる。

[0146] 図２５は、本実施形態に係る制御信号生成処理の一例を説明するための図である。図２５に示すように、システム１は、図２１を参照して説明した生成ネット１５に、図１６を参照して説明した識別ネット１１及び図１７を参照して説明した推論ネット１３を組み合わせて制御信号を生成してもよい。ここで注目すべきは、識別器１２から出力された動作種別情報ｙが推論器１４に入力される点と、それに伴い入力がセンサデータｘのみである点である。これにより、ユーザの操作負荷が軽減される。

[0147] ＜＜６．補足＞＞

上記説明したように、本開示の一実施形態では、ラベル有り特徴情報ｙ及

びラベル無し特徴情報ｚから時系列データｘが生成される。時系列データｘ、ラベル有り特徴情報ｙ、及びラベル無し特徴情報ｚは、上述した例に限定されず、多様に考えられる。

[0148] 例えば、ラベル有り特徴情報ｙが楽譜であり、ラベル無し特徴情報ｚが演奏特徴及び歌唱特徴であり、時系列データｘが歌唱データであってもよい。この場合、システム１は、あるアーティストの歌詞を含んだ楽譜と歌唱データとの組み合わせを複数含むデータセットにより演奏特徴及び歌唱特徴を学習し、当該アーティストがあたかも歌唱したかのような、異なるアーティストの楽譜に対応する歌唱データを生成することが可能となる。

[0149] 例えば、ラベル有り特徴情報ｙが小説のプロットであり、ラベル無し特徴情報ｚが作家特徴（例えば、作風）であり、時系列データｘが小説であってもよい。この場合、システム１は、ある作家のプロットと小説との組み合わせを複数含むデータセットにより作家特徴を学習し、当該作家があたかも執筆したかのような、新しく入力されたプロットに対応する小説を生成することが可能となる。

[0150] 例えば、ラベル有り特徴情報ｙは第１の言語のテキスト情報であり、ラベル無し特徴情報ｚは言語特徴であり、時系列データｘは第２のテキスト情報であってもよい。この場合、システム１は、日本語のテキストと英語のテキストを含む対訳コーパス（即ち、データセット）を用いて言語特徴を学習し、新しく入力された日本語のテキストを英語のテキストに変換して出力することが可能となる。

[0151] 例えば、ラベル有り特徴情報ｙは絵のプロットであり、ラベル無し特徴情報ｚは画家特徴であり、時系列データｘは絵の描き順であってもよい。この場合、システム１は、「たくさんのヒマワリが青空の下にある」といった絵のプロットと、それをもとに描かれた絵の描き順との組み合わせを複数含むデータセットにより画家特徴を学習し、新しく入力された絵のプロットに対する絵の描き順を生成することが可能となる。

[0152] ＜＜７．ハードウェア構成例＞＞

最後に、図２６を参照して、本実施形態に係る情報処理装置のハードウェア構成について説明する。図２６は、本実施形態に係る情報処理装置のハードウェア構成の一例を示すブロック図である。なお、図２６に示す情報処理装置９００は、例えば、図２及び図３にそれぞれ示した処理装置１００又は端末装置２００を実現し得る。本実施形態に係る処理装置１００又は端末装置２００による情報処理は、ソフトウェアと、以下に説明するハードウェアとの協働により実現される。

[0153] 図２６に示すように、情報処理装置９００は、ＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）９０１、ＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）９０２、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）９０３及びホストバス９０４ａを備える。また、情報処理装置９００は、ブリッジ９０４、外部バス９０４ｂ、インタフェース９０５、入力装置９０６、出力装置９０７、ストレージ装置９０８、ドライブ９０９、接続ポート９１１及び通信装置９１３を備える。情報処理装置９００は、ＣＰＵ９０１に代えて、又はこれとともに、ＤＳＰ若しくはＡＳＩＣ等の処理回路を有してもよい。

[0154] ＣＰＵ９０１は、演算処理装置および制御装置として機能し、各種プログラムに従って情報処理装置９００内の動作全般を制御する。また、ＣＰＵ９０１は、マイクロプロセッサであってもよい。ＲＯＭ９０２は、ＣＰＵ９０１が使用するプログラムや演算パラメータ等を記憶する。ＲＡＭ９０３は、ＣＰＵ９０１の実行において使用するプログラムや、その実行において適宜変化するパラメータ等を一時記憶する。ＣＰＵ９０１は、例えば、図２に示す制御部１３０、又は図３に示す制御部２５０を形成し得る。

[0155] ＣＰＵ９０１、ＲＯＭ９０２及びＲＡＭ９０３は、ＣＰＵバスなどを含むホストバス９０４ａにより相互に接続されている。ホストバス９０４ａは、ブリッジ９０４を介して、ＰＣＩ（Ｐｅｒｉｐｈｅｒａｌ Ｃｏｍｐｏｎｅｎｔ Ｉｎｔｅｒｃｏｎｎｅｃｔ／Ｉｎｔｅｒｆａｃｅ）バスなどの外部バス９０４ｂに接続されている。なお、必ずしもホストバス９０４ａ、ブリッ

ジ９０４および外部バス９０４ｂを分離構成する必要はなく、１つのバスにこれらの機能を実装してもよい。

[0156] 入力装置９０６は、例えば、マウス、キーボード、タッチパネル、ボタン、マイクロフォン、スイッチ及びレバー等、ユーザによって情報が入力される装置によって実現される。また、入力装置９０６は、例えば、赤外線やその他の電波を利用したリモートコントロール装置であってもよいし、情報処理装置９００の操作に対応した携帯電話やＰＤＡ等の外部接続機器であってもよい。さらに、入力装置９０６は、例えば、上記の入力手段を用いてユーザにより入力された情報に基づいて入力信号を生成し、ＣＰＵ９０１に出力する入力制御回路などを含んでいてもよい。情報処理装置９００のユーザは、この入力装置９０６を操作することにより、情報処理装置９００に対して各種のデータを入力したり処理動作を指示したりすることができる。

[0157] 他にも、入力装置９０６は、ユーザに関する情報を検知する装置により形成され得る。例えば、入力装置９０６は、画像センサ（例えば、カメラ）、深度センサ（例えば、ステレオカメラ）、加速度センサ、ジャイロセンサ、地磁気センサ、光センサ、音センサ、測距センサ、力センサ等の各種のセンサを含み得る。また、入力装置９０６は、情報処理装置９００の姿勢、移動速度等、情報処理装置９００自身の状態に関する情報や、情報処理装置９００の周辺の明るさや騒音等、情報処理装置９００の周辺環境に関する情報を取得してもよい。また、入力装置９０６は、ＧＮＳＳ（Global Navigation Satellite System）衛星からのＧＮＳＳ信号（例えば、ＧＰＳ（Global Positioning System）衛星からのＧＰＳ信号）を受信して装置の緯度、経度及び高度を含む位置情報を測定するＧＮＳＳモジュールを含んでもよい。また、位置情報に関しては、入力装置９０６は、Ｗｉ－Ｆｉ（登録商標）、携帯電話・ＰＨＳ・スマートフォン等との送受信、または近距離通信等により位置を検知するものであってもよい。入力装置９０６は、例えば、図３に示す入力部２１０を形成し得る。

[0158] 出力装置９０７は、取得した情報をユーザに対して視覚的又は聴覚的に通

知することが可能な装置で形成される。このような装置として、ＣＲＴディスプレイ装置、液晶ディスプレイ装置、プラズマディスプレイ装置、ＥＬディスプレイ装置、レーザープロジェクタ、ＬＥＤプロジェクタ及びランプ等の表示装置や、スピーカ及びヘッドホン等の音声出力装置や、プリンタ装置等がある。出力装置９０７は、例えば、情報処理装置９００が行った各種処理により得られた結果を出力する。具体的には、表示装置は、情報処理装置９００が行った各種処理により得られた結果を、テキスト、イメージ、表、グラフ等、様々な形式で視覚的に表示する。他方、音声出力装置は、再生された音声データや音響データ等からなるオーディオ信号をアナログ信号に変換して聴覚的に出力する。上記表示装置又は上記音声出力装置は、例えば、図３に示す出力部２２０を形成し得る。

[0159] ストレージ装置９０８は、情報処理装置９００の記憶部の一例として形成されたデータ格納用の装置である。ストレージ装置９０８は、例えば、ＨＤＤ等の磁気記憶部デバイス、半導体記憶デバイス、光記憶デバイス又は光磁気記憶デバイス等により実現される。ストレージ装置９０８は、記憶媒体、記憶媒体にデータを記録する記録装置、記憶媒体からデータを読み出す読出し装置および記憶媒体に記録されたデータを削除する削除装置などを含んでもよい。このストレージ装置９０８は、ＣＰＵ９０１が実行するプログラムや各種データ及び外部から取得した各種のデータ等を格納する。ストレージ装置９０８は、例えば、図２に示す記憶部１２０又は図３に示す記憶部２４０を形成し得る。

[0160] ドライブ９０９は、記憶媒体用リーダライタであり、情報処理装置９００に内蔵、あるいは外付けされる。ドライブ９０９は、装着されている磁気ディスク、光ディスク、光磁気ディスク、または半導体メモリ等のリムーバブル記憶媒体に記録されている情報を読み出して、ＲＡＭ９０３に出力する。また、ドライブ９０９は、リムーバブル記憶媒体に情報を書き込むこともできる。

[0161] 接続ポート９１１は、外部機器と接続されるインタフェースであって、例

えばＵＳＢ（Ｕｎｉｖｅｒｓａｌ　Ｓｅｒｉａｌ　Ｂｕｓ）などによりデータ伝送可能な外部機器との接続口である。

[0162] 通信装置９１３は、例えば、ネットワーク９２０に接続するための通信デバイス等で形成された通信インタフェースである。通信装置９１３は、例えば、有線若しくは無線ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、ＬＴＥ（Ｌｏｎｇ　Ｔｅｒｍ　Ｅｖｏｌｕｔｉｏｎ）、Ｂｌｕｅｔｏｏｔｈ（登録商標）又はＷＵＳＢ（Ｗｉｒｅｌｅｓｓ　ＵＳＢ）用の通信カード等である。また、通信装置９１３は、光通信用のルータ、ＡＤＳＬ（Ａｓｙｍｍｅｔｒｉｃ　Ｄｉｇｉｔａｌ　Ｓｕｂｓｃｒｉｂｅｒ　Ｌｉｎｅ）用のルータ又は各種通信用のモデム等であってもよい。この通信装置９１３は、例えば、インターネットや他の通信機器との間で、例えばＴＣＰ／ＩＰ等の所定のプロトコルに則して信号等を送受信することができる。通信装置９１３は、例えば、図２に示す通信部１１０又は図３に示す通信部２３０を形成し得る。

[0163] なお、ネットワーク９２０は、ネットワーク９２０に接続されている装置から送信される情報の有線、または無線の伝送路である。例えば、ネットワーク９２０は、インターネット、電話回線網、衛星通信網などの公衆回線網や、Ｅｔｈｅｒｎｅｔ（登録商標）を含む各種のＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、ＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）などを含んでもよい。また、ネットワーク９２０は、ＩＰ－ＶＰＮ（Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ－Ｖｉｒｔｕａｌ　Ｐｒｉｖａｔｅ　Ｎｅｔｗｏｒｋ）などの専用回線網を含んでもよい。

[0164] 以上、本実施形態に係る情報処理装置９００の機能を実現可能なハードウェア構成の一例を示した。上記の各構成要素は、汎用的な部材を用いて実現されていてもよいし、各構成要素の機能に特化したハードウェアにより実現されていてもよい。従って、本実施形態を実施する時々の技術レベルに応じて、適宜、利用するハードウェア構成を変更することが可能である。

[0165] なお、上述のような本実施形態に係る情報処理装置９００の各機能を実現

するためのコンピュータプログラムを作製し、ＰＣ等に実装することが可能である。また、このようなコンピュータプログラムが格納された、コンピュータで読み取り可能な記録媒体も提供することができる。記録媒体は、例えば、磁気ディスク、光ディスク、光磁気ディスク、フラッシュメモリ等である。また、上記のコンピュータプログラムは、記録媒体を用いずに、例えばネットワークを介して配信されてもよい。

[0166] ＜＜８．まとめ＞＞

以上、図１～図２６を参照して、本開示の一実施形態について詳細に説明した。上記説明したように、本実施形態に係るシステム１は、意味付けされた第１の特徴情報（即ち、ラベル有り特徴情報）及び意味付けされない第２の特徴情報（即ち、ラベル無し特徴情報）を取得して、ラベル有り特徴情報及びラベル無し特徴情報が示す特徴を有する時系列データを生成する。これにより、データの特徴付けを適切に行うことが可能となる。より具体的には、ユーザは、所望する特徴に合わせてラベル有り特徴情報及びラベル無し特徴情報を選択したり編集したりすることで、所望する特徴を有する時系列データを、容易に生成させることが可能である。

[0167] 以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

[0168] 例えば、上記実施形態では、処理装置１００と端末装置２００とが別箇の装置であるものとして説明したが、本技術はかかる例に限定されない。例えば、処理装置１００と端末装置２００とはひとつの装置として実現されてもよい。

[0169] また、本明細書においてフローチャート及びシーケンス図を用いて説明した処理は、必ずしも図示された順序で実行されなくてもよい。いくつかの処

理ステップは、並列的に実行されてもよい。また、追加的な処理ステップが採用されてもよく、一部の処理ステップが省略されてもよい。

[0170] また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

[0171] なお、以下のような構成も本開示の技術的範囲に属する。

（１）

意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得する取得部と、

前記取得部により取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを生成する生成部と、

を備える情報処理装置。

（２）

前記第１の特徴情報は、前記時系列データのコンテキストを示す情報を含む、前記（１）に記載の情報処理装置。

（３）

前記第１の特徴情報は、前記時系列データの枠組みを示す情報を含む、前記（２）に記載の情報処理装置。

（４）

前記枠組みを示す情報は、テキスト情報であり、

前記時系列データは、前記テキスト情報を読み上げた音声データである、前記（３）に記載の情報処理装置。

（５）

前記コンテキストを示す情報は、話者自身に関する情報を含む、前記（４）に記載の情報処理装置。

（６）

前記コンテキストを示す情報は、話者が話す環境に関する情報を含む、前

記（４）又は（５）に記載の情報処理装置。

（７）

前記コンテキストを示す情報は、話者が話す話題に関する情報を含む、前記（４）～（６）のいずれか一項に記載の情報処理装置。

（８）

前記枠組みを示す情報は、楽譜情報であり、

前記時系列データは、前記楽譜情報に従って演奏された音楽データである、前記（３）に記載の情報処理装置。

（９）

前記コンテキストを示す情報は、動作種別を示す情報であり、

前記時系列データは、動物体をセンシングしたセンサデータである、前記（２）又は（３）に記載の情報処理装置。

（１０）

前記コンテキストを示す情報は、動作種別を示す情報であり、

前記時系列データは、動物体に含まれるアクチュエータを動作させるための制御信号である、前記（２）又は（３）に記載の情報処理装置。

（１１）

前記第１の特徴情報は、ユーザ音声による指示に応じて編集される、前記（１）～（１０）のいずれか一項に記載の情報処理装置。

（１２）

前記第１の特徴情報及び前記第２の特徴情報は、ユーザ音声が有する特徴に対応する特徴を示す、前記（１）～（１１）のいずれか一項に記載の情報処理装置。

（１３）

前記取得部は、前記第１の特徴情報及び前記第２の特徴情報に対応する識別情報を取得し、

前記生成部は、前記識別情報に対応する特徴を有する時系列データを生成する、前記（１）～（１２）のいずれか一項に記載の情報処理装置。

（１４）

前記生成部は、ニューラルネットにより前記時系列データを生成する、前記（１）～（１３）のいずれか一項に記載の情報処理装置。

（１５）

前記情報処理装置は、前記ニューラルネットのパラメータを学習する学習部をさらに備える、前記（１４）に記載の情報処理装置。

（１６）

前記情報処理装置は、前記時系列データから前記第１の特徴情報を識別する識別部をさらに備える、前記（１）～（１５）のいずれか一項に記載の情報処理装置。

（１７）

前記情報処理装置は、前記時系列データ及び前記第１の特徴情報から前記第２の特徴情報を推論する推論部をさらに備える、前記（１）～（１６）のいずれか一項に記載の情報処理装置。

（１８）

意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を他の装置に通知する通知部と、

前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを前記他の装置から取得する取得部と、

を備える情報処理装置。

（１９）

意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得することと、

取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データをプロセッサにより生成することと、

を含む情報処理方法。

（２０）

意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を他

の装置に通知することと、

前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを前記他の装置からプロセッサにより取得することと、

を含む情報処理方法。

符号の説明

[0172]

１	システム
１１	識別ネット
１２	識別器
１３	推論ネット
１４	推論器
１５	生成ネット
１６	生成器
４２	翻訳器
６２	制御信号生成器
６４	動作スタイル情報選択器
１００	処理装置
１１０	通信部
１２０	記憶部
１３０	制御部
１３１	取得部
１３３	演算部
１３５	学習部
１３７	通知部
２００	端末装置
２１０	入力部
２２０	出力部
２３０	通信部
２４０	記憶部

２５０　制御部

２５１　通知部

２５３　取得部

請求の範囲

[請求項1] 意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得する取得部と、

前記取得部により取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを生成する生成部と、

を備える情報処理装置。

[請求項2] 前記第１の特徴情報は、前記時系列データのコンテキストを示す情報を含む、請求項１に記載の情報処理装置。

[請求項3] 前記第１の特徴情報は、前記時系列データの枠組みを示す情報を含む、請求項２に記載の情報処理装置。

[請求項4] 前記枠組みを示す情報は、テキスト情報であり、

前記時系列データは、前記テキスト情報を読み上げた音声データである、請求項３に記載の情報処理装置。

[請求項5] 前記コンテキストを示す情報は、話者自身に関する情報を含む、請求項４に記載の情報処理装置。

[請求項6] 前記コンテキストを示す情報は、話者が話す環境に関する情報を含む、請求項４に記載の情報処理装置。

[請求項7] 前記コンテキストを示す情報は、話者が話す話題に関する情報を含む、請求項４に記載の情報処理装置。

[請求項8] 前記枠組みを示す情報は、楽譜情報であり、

前記時系列データは、前記楽譜情報に従って演奏された音楽データである、請求項３に記載の情報処理装置。

[請求項9] 前記コンテキストを示す情報は、動作種別を示す情報であり、

前記時系列データは、動物体をセンシングしたセンサデータである、請求項２に記載の情報処理装置。

[請求項10] 前記コンテキストを示す情報は、動作種別を示す情報であり、

前記時系列データは、動物体に含まれるアクチュエータを動作させるための制御信号である、請求項２に記載の情報処理装置。

[請求項11] 前記第１の特徴情報は、ユーザ音声による指示に応じて編集される、請求項１に記載の情報処理装置。

[請求項12] 前記第１の特徴情報及び前記第２の特徴情報は、ユーザ音声が有する特徴に対応する特徴を示す、請求項１に記載の情報処理装置。

[請求項13] 前記取得部は、前記第１の特徴情報及び前記第２の特徴情報に対応する識別情報を取得し、

前記生成部は、前記識別情報に対応する特徴を有する時系列データを生成する、請求項１に記載の情報処理装置。

[請求項14] 前記生成部は、ニューラルネットにより前記時系列データを生成する、請求項１に記載の情報処理装置。

[請求項15] 前記情報処理装置は、前記ニューラルネットのパラメータを学習する学習部をさらに備える、請求項１４に記載の情報処理装置。

[請求項16] 前記情報処理装置は、前記時系列データから前記第１の特徴情報を識別する識別部をさらに備える、請求項１に記載の情報処理装置。

[請求項17] 前記情報処理装置は、前記時系列データ及び前記第１の特徴情報から前記第２の特徴情報を推論する推論部をさらに備える、請求項１に記載の情報処理装置。

[請求項18] 意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を他の装置に通知する通知部と、

前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを前記他の装置から取得する取得部と、

を備える情報処理装置。

[請求項19] 意味付けされた第１の特徴情報及び意味付けされない第２の特徴情報を取得することと、

取得された前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データをプロセッサにより生成することと、

を含む情報処理方法。

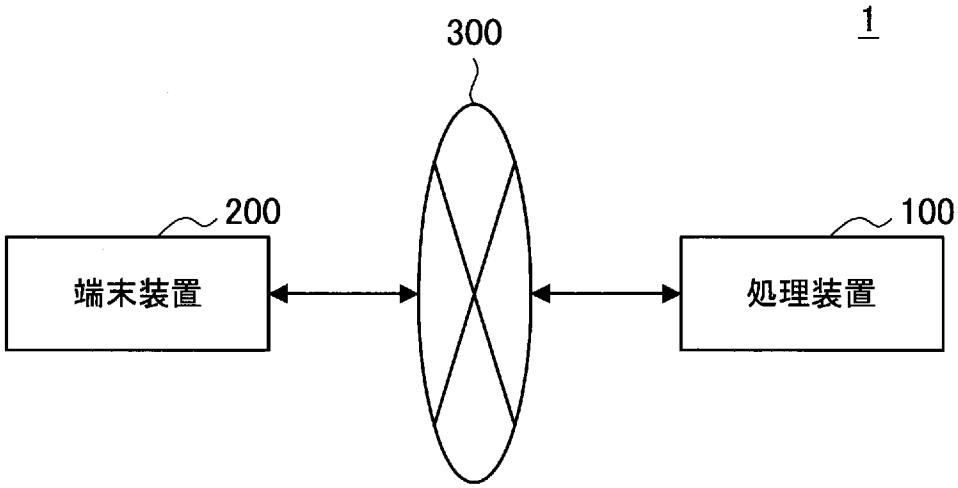
[請求項20] 意味付けされた第１の特徴情報及び意味付けされない第２の特徴情

報を他の装置に通知することと、

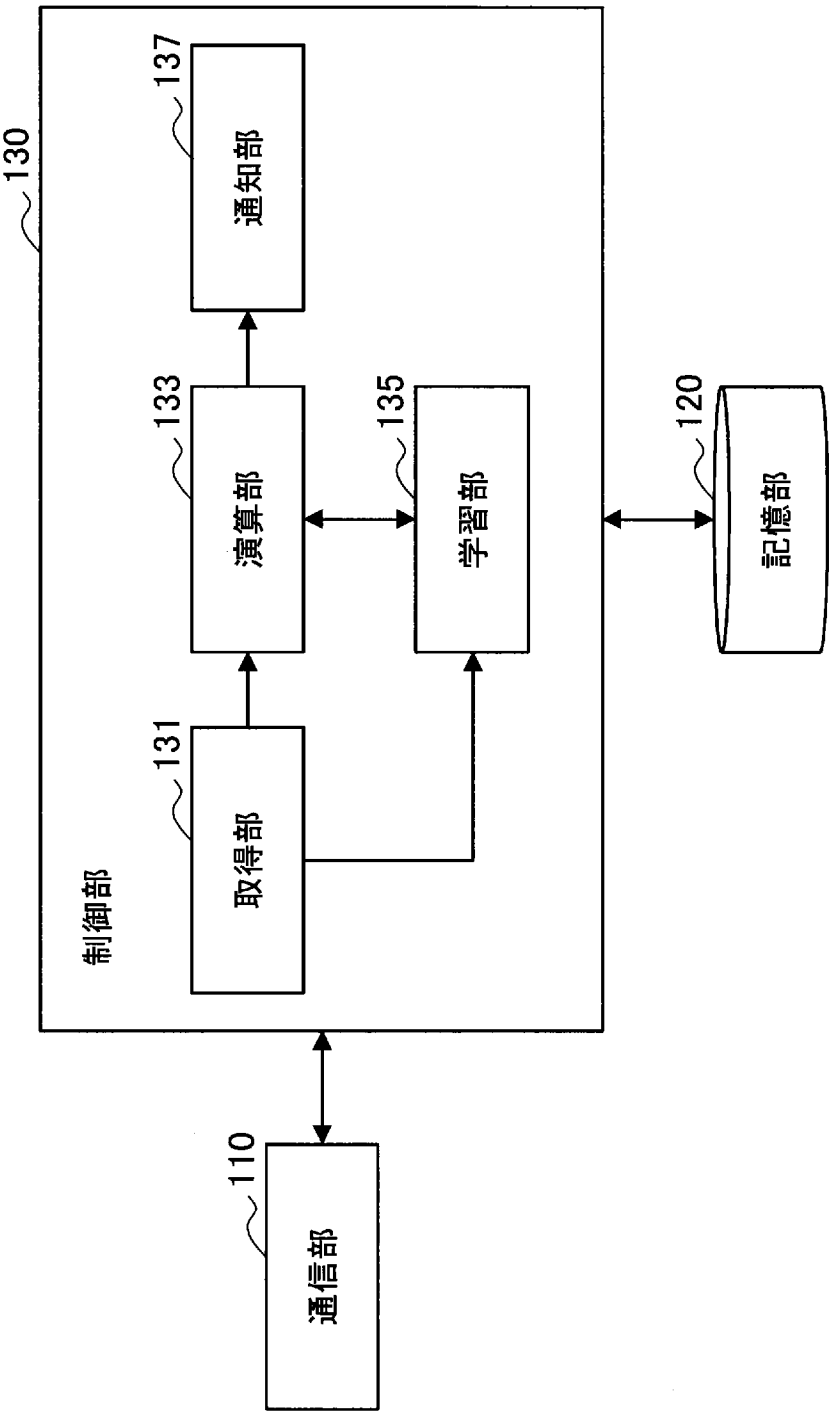
前記第１の特徴情報及び前記第２の特徴情報が示す特徴を有する時系列データを前記他の装置からプロセッサにより取得することと、

を含む情報処理方法。

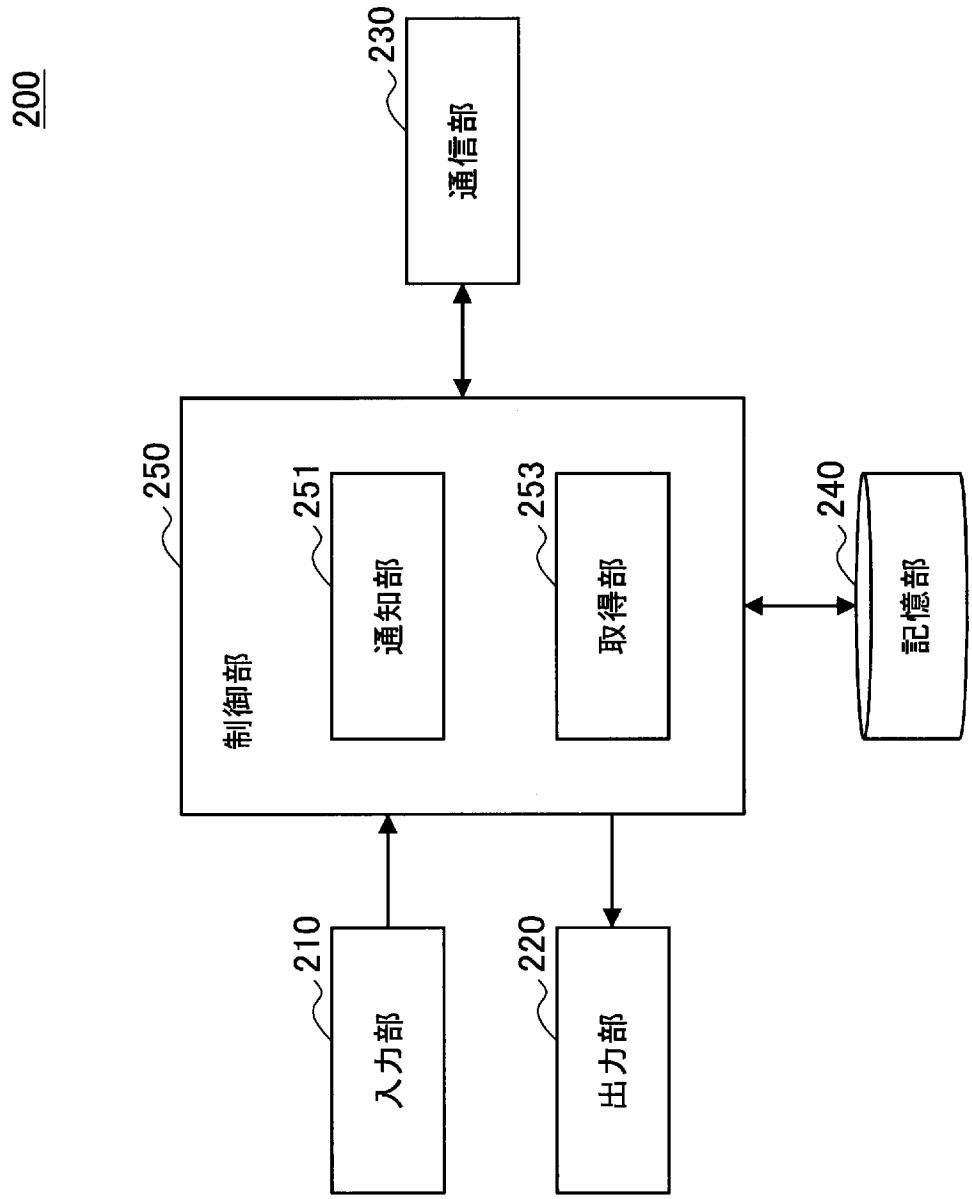
[図1]



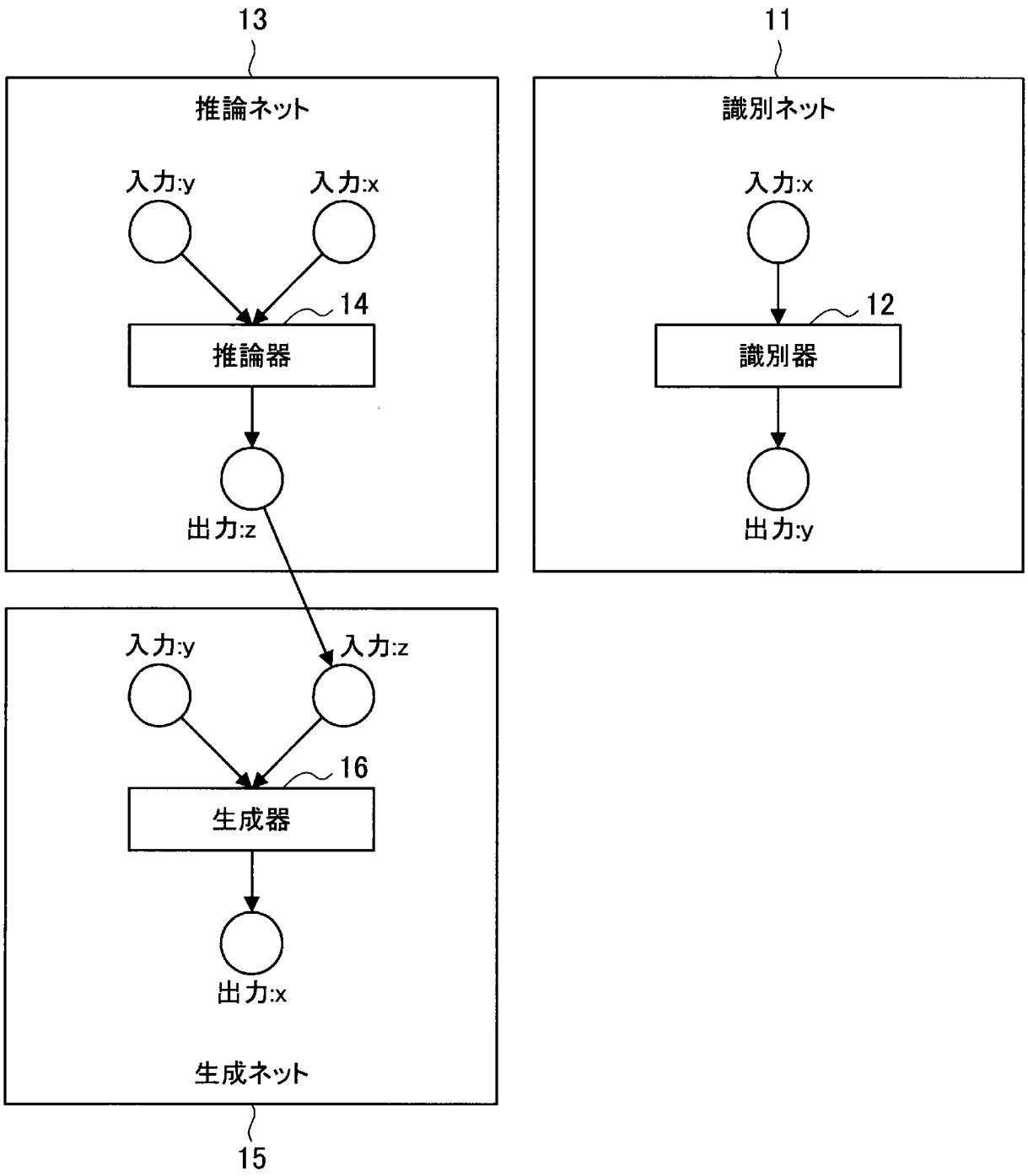
[図2]



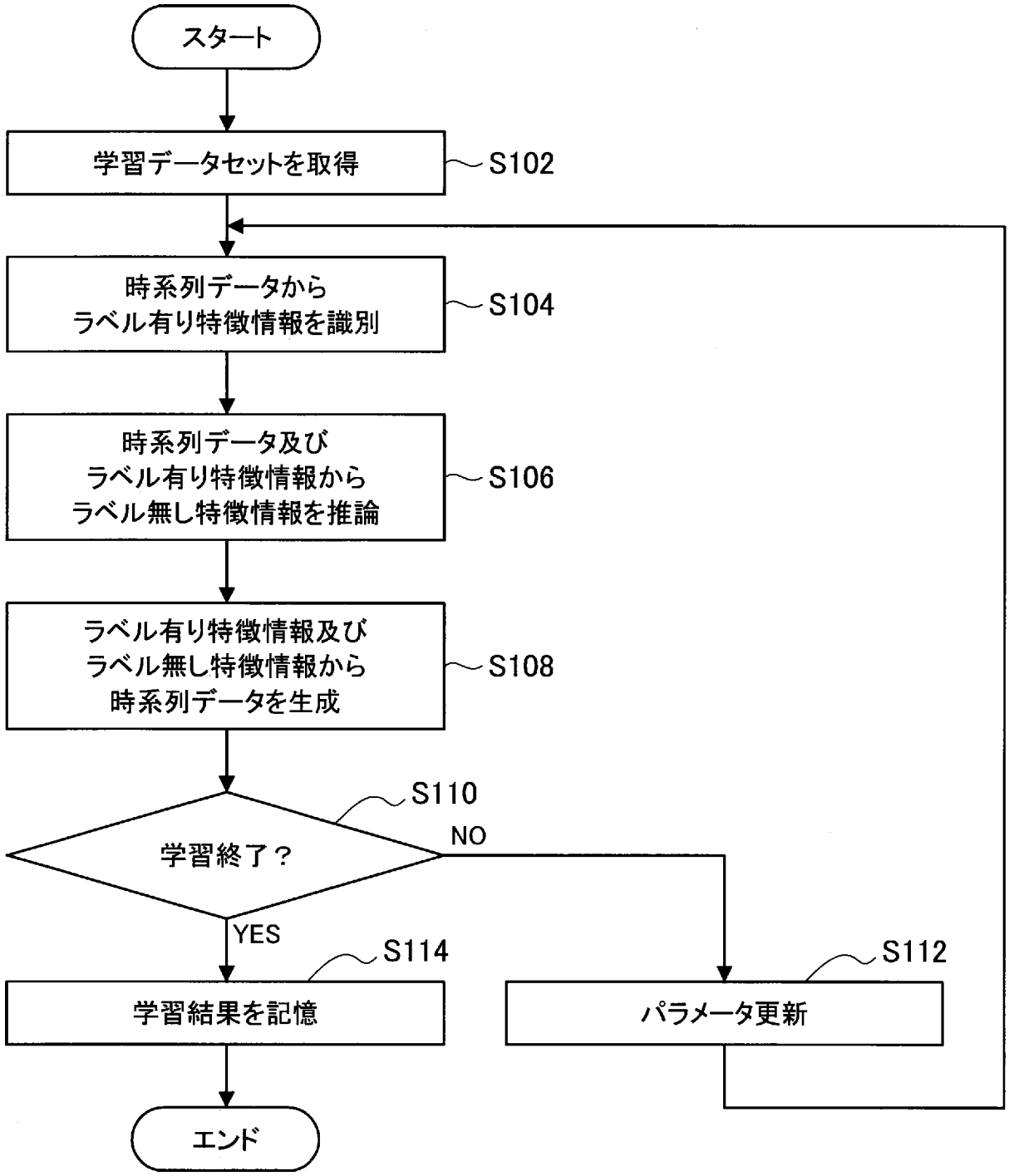
[図3]



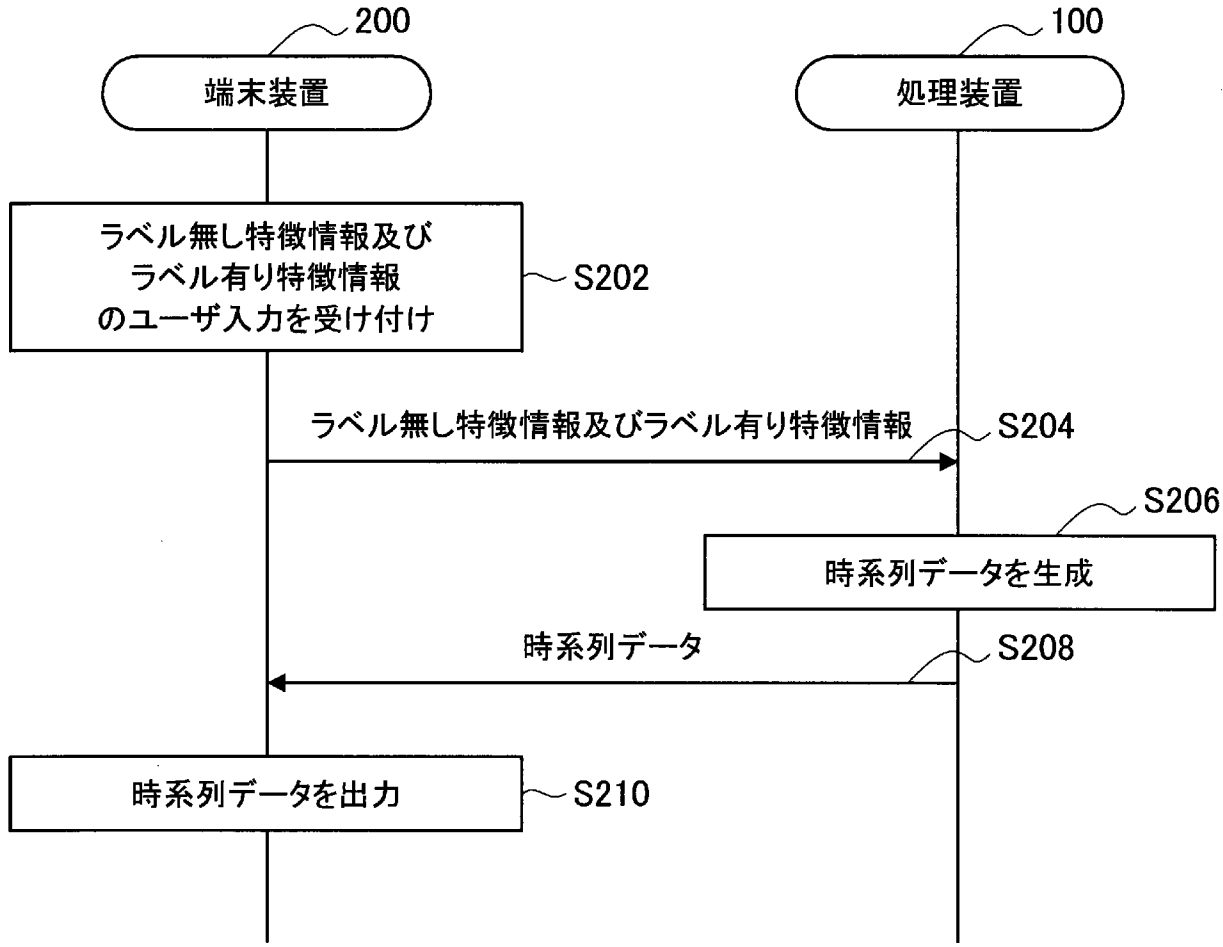
[図4]


10
13
推論ネット
入力:y
入力:x
14
推論器
出力:z
11
識別ネット
入力:x
12
識別器
出力:y
入力:y
入力:z
16
生成器
出力:x
生成ネット
15

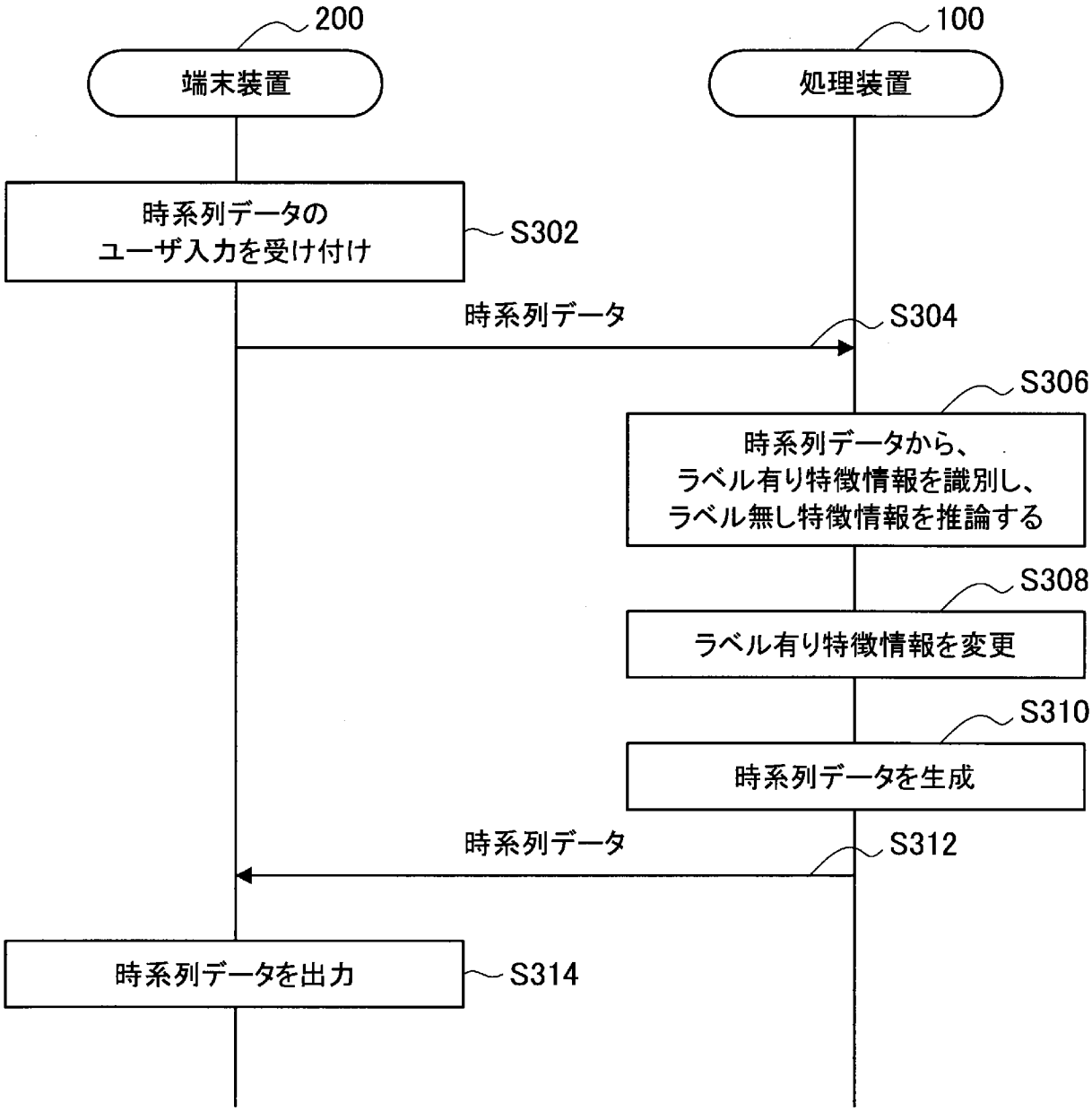
[図5]


スタート
学習データセットを取得
S102
時系列データから
ラベル有り特徴情報を識別
S104
時系列データ及び
ラベル有り特徴情報から
ラベル無し特徴情報を推論
S106
ラベル有り特徴情報及び
ラベル無し特徴情報から
時系列データを生成
S108
S110
学習終了？
NO
YES
S114
学習結果を記憶
S112
パラメータ更新
エンド

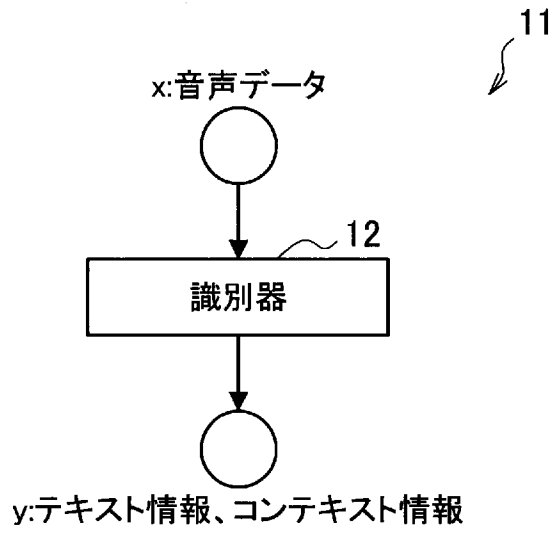
[図6]


200
端末装置
100
処理装置
ラベル無し特徴情報及び
ラベル有り特徴情報
のユーザ入力を受け付け
S202
ラベル無し特徴情報及びラベル有り特徴情報
S204
S206
時系列データを生成
時系列データ
S208
時系列データを出力
S210

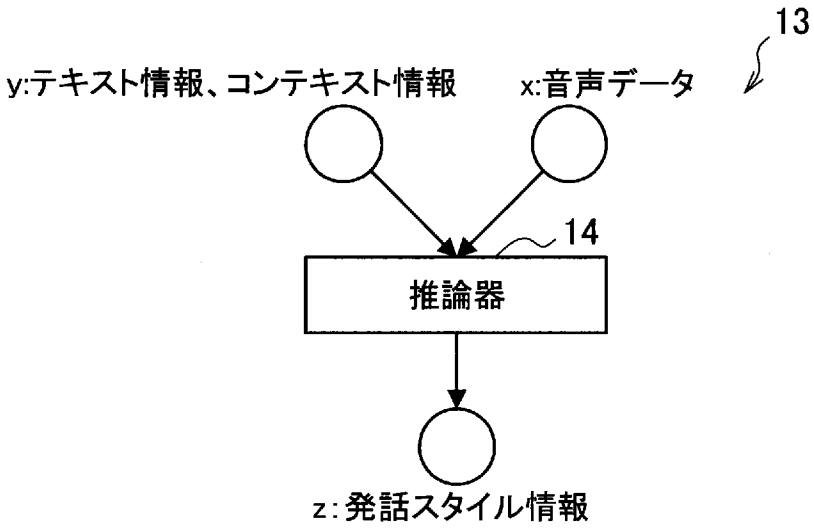
[図7]


200
端末装置
100
処理装置
時系列データの
ユーザ入力を受け付け
S302
時系列データ
S304
S306
時系列データから、
ラベル有り特徴情報を識別し、
ラベル無し特徴情報を推論する
S308
ラベル有り特徴情報を変更
S310
時系列データを生成
時系列データ
S312
時系列データを出力
S314

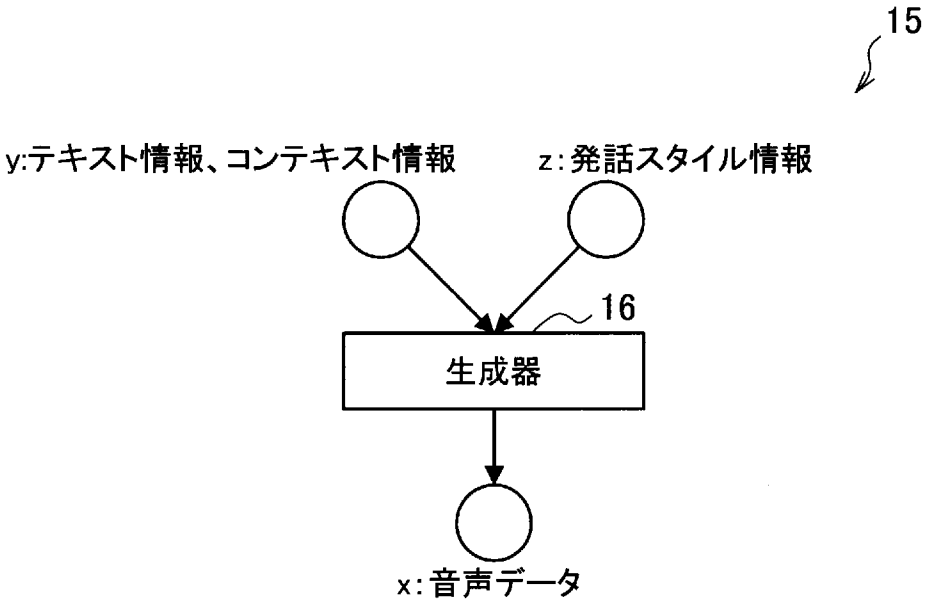
[図8]


11
x:音声データ
12
識別器
y:テキスト情報、コンテキスト情報

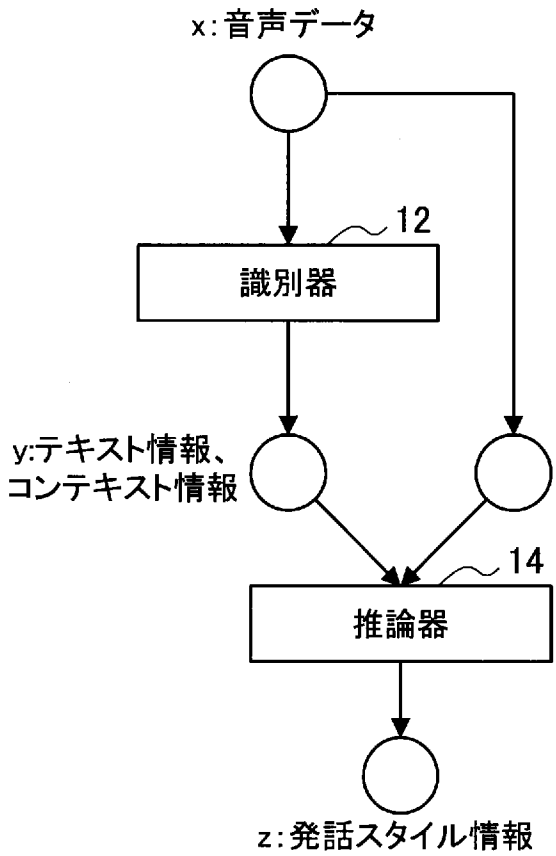
[図9]


13
y:テキスト情報、コンテキスト情報
x:音声データ
14
推論器
z:発話スタイル情報

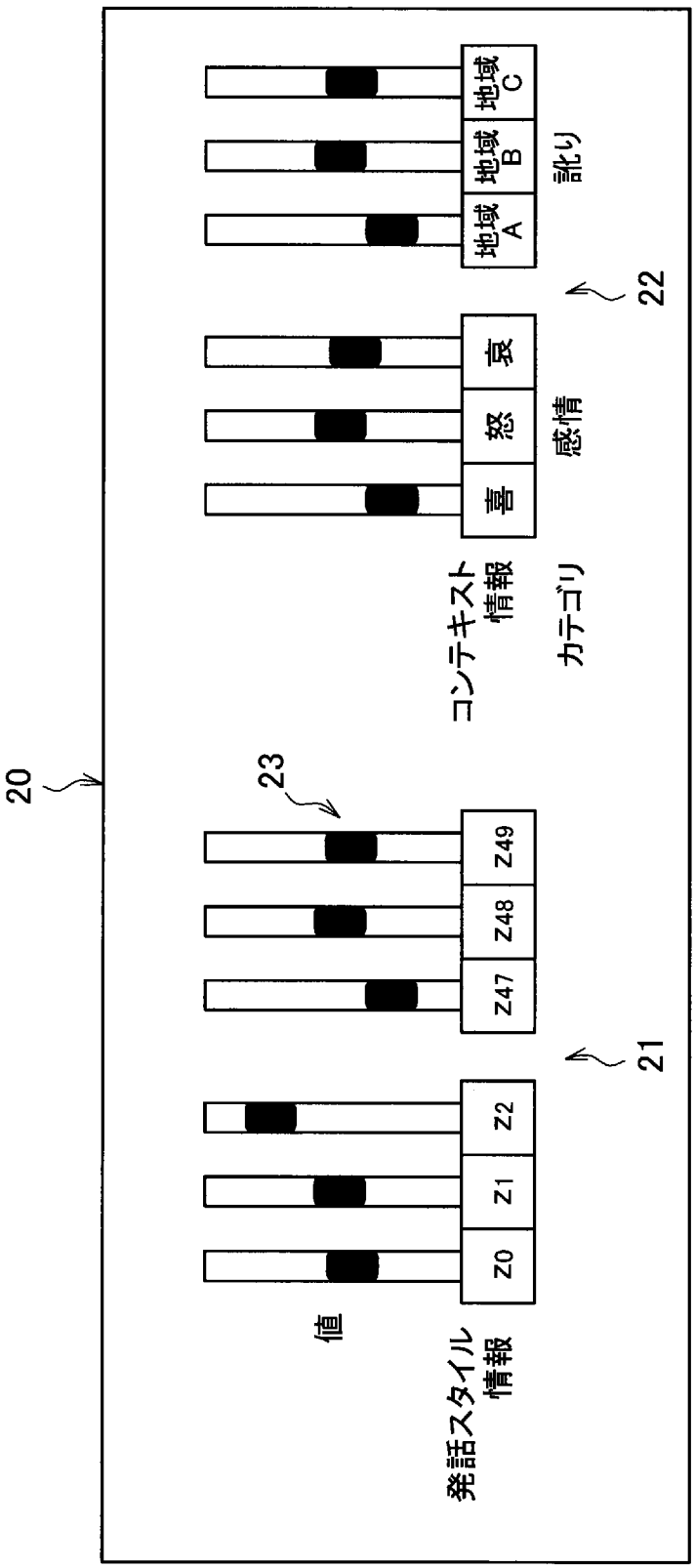
[図10]


15
y:テキスト情報、コンテキスト情報
z:発話スタイル情報
16
生成器
x:音声データ

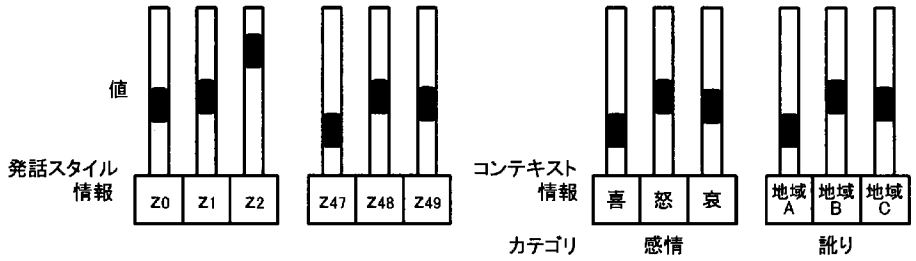
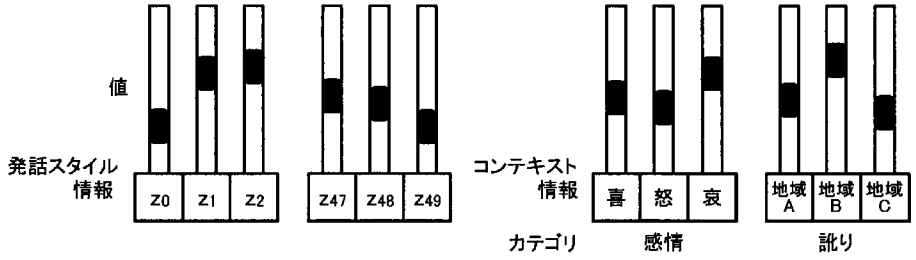
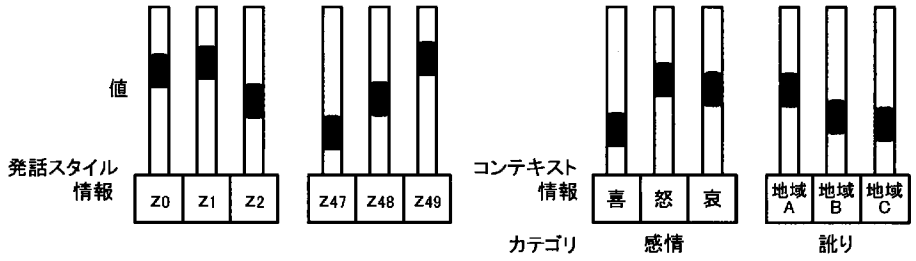
[図11]



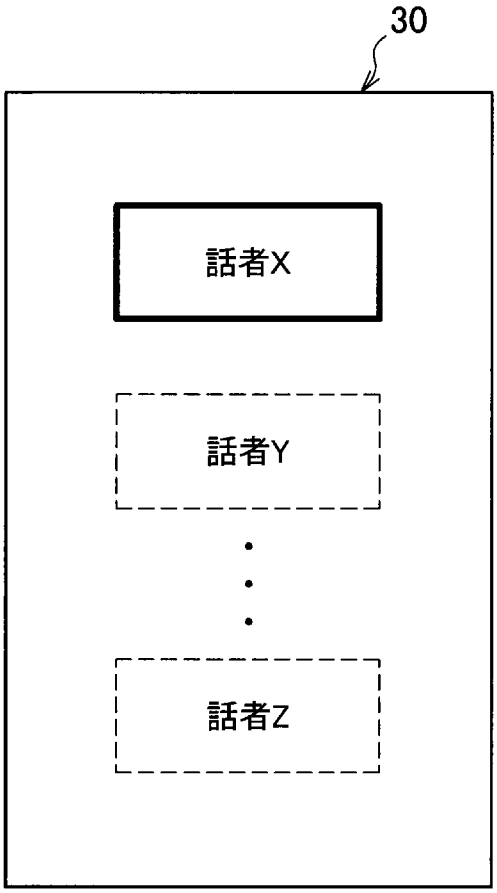
[図12]


20
23
Z49
Z48
Z47
Z2
Z1
Z0
値
発話スタイル情報
21
地域C
地域B
地域A
訛り
喜
怒
哀
感情
コンテキスト情報
カテゴリ
22

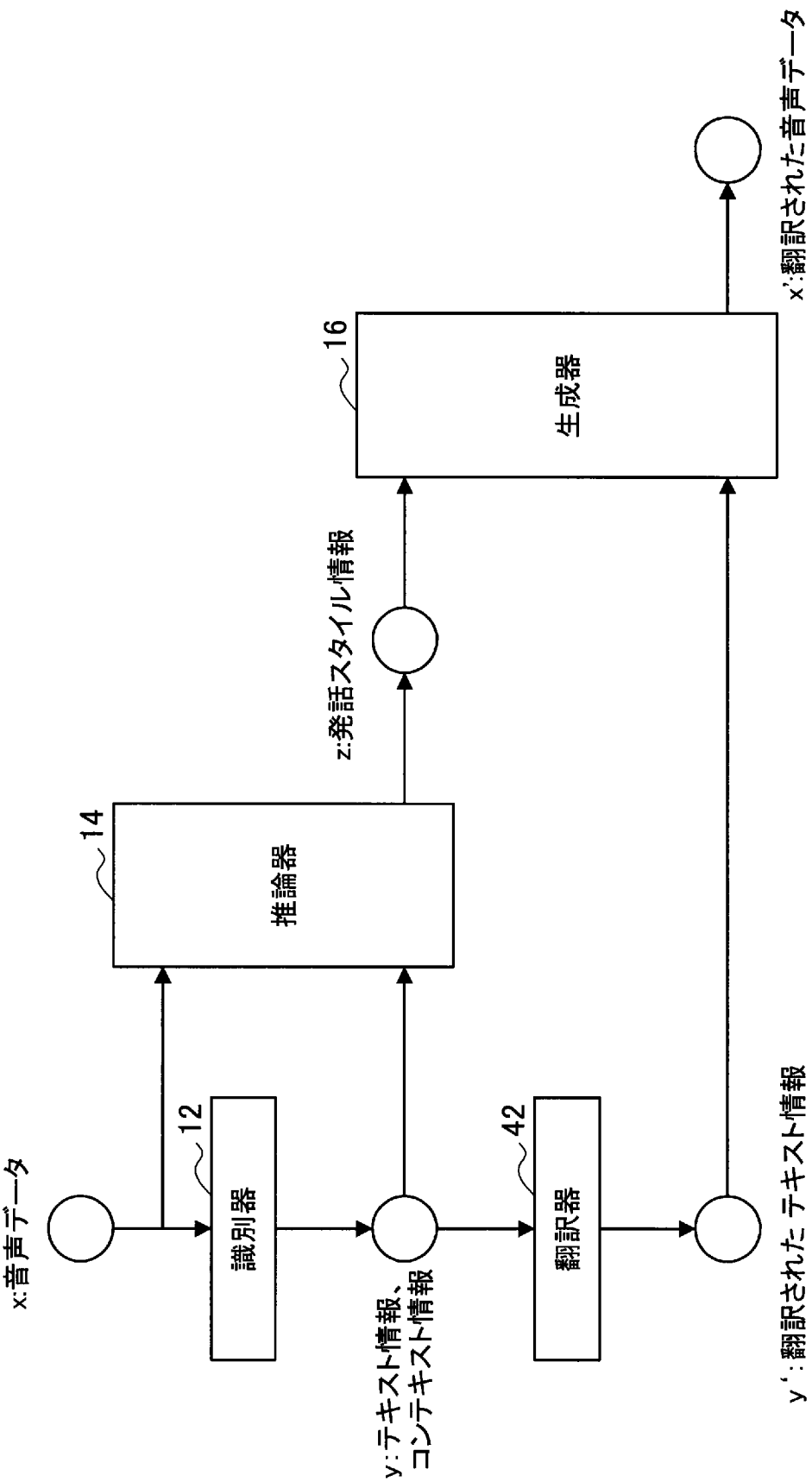
[図13]

識別情報	音声特徴情報
話者X	
話者Y	
⋮	⋮
話者Z	
⋮	⋮

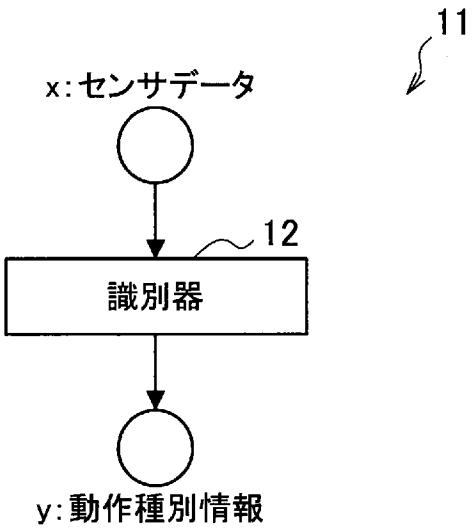
[図14]



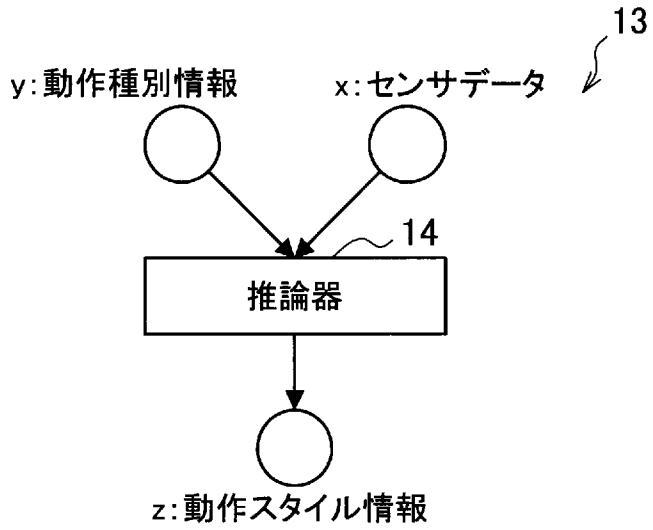
[図15]


x:音声データ
12
識別器
y:テキスト情報、
コンテキスト情報
14
推論器
z:発話スタイル情報
42
翻訳器
y':翻訳されたテキスト情報
16
生成器
x':翻訳された音声データ

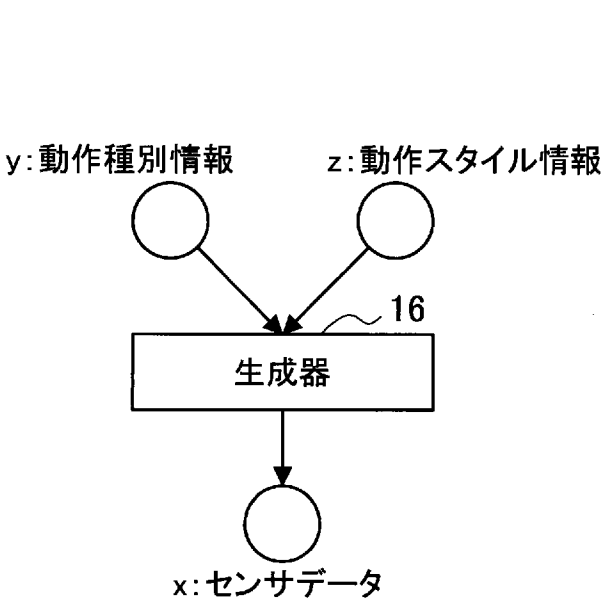
[図16]


11
x:センサデータ
12
識別器
y:動作種別情報

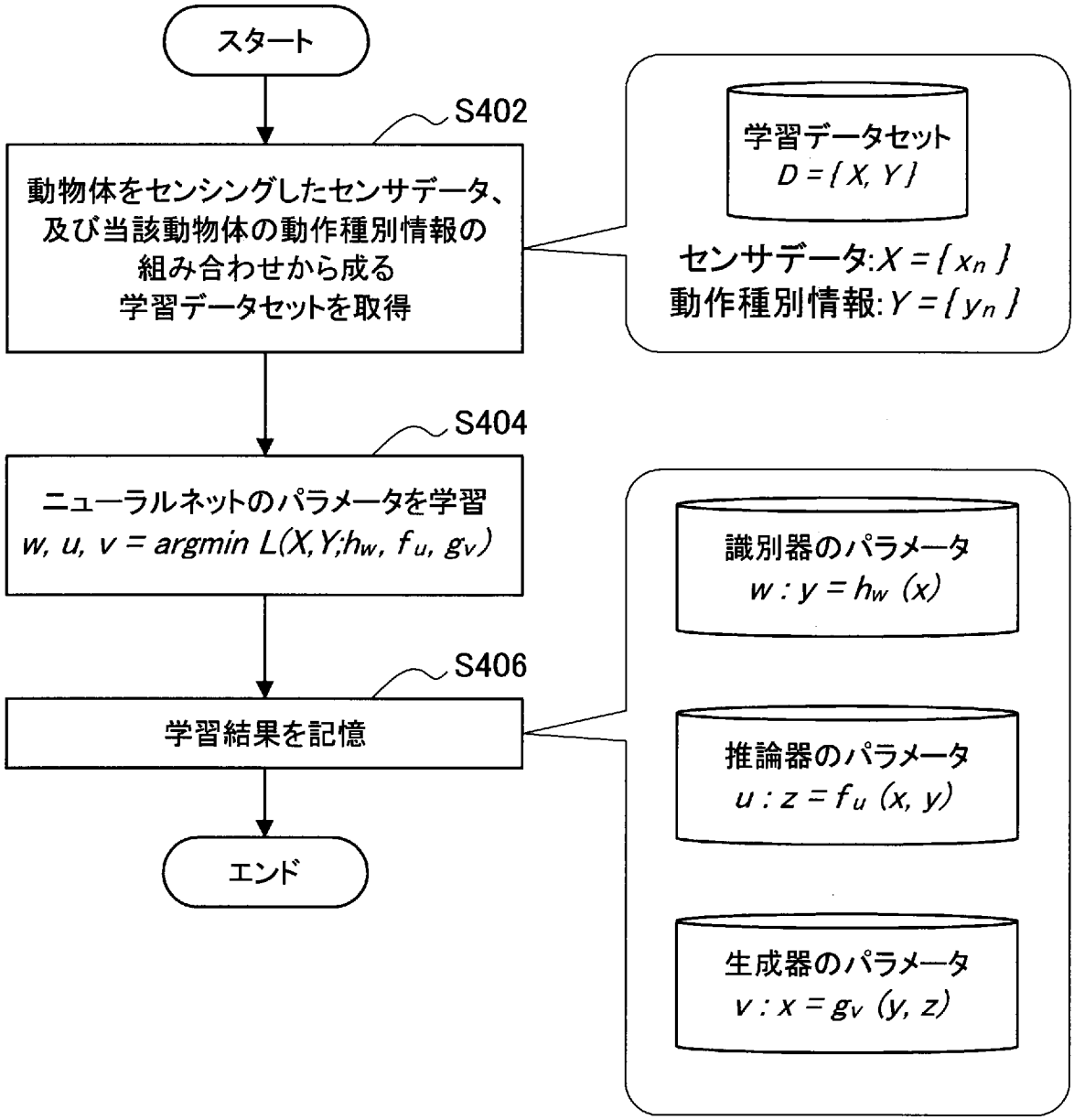
[図17]


13
y:動作種別情報
x:センサデータ
14
推論器
z:動作スタイル情報

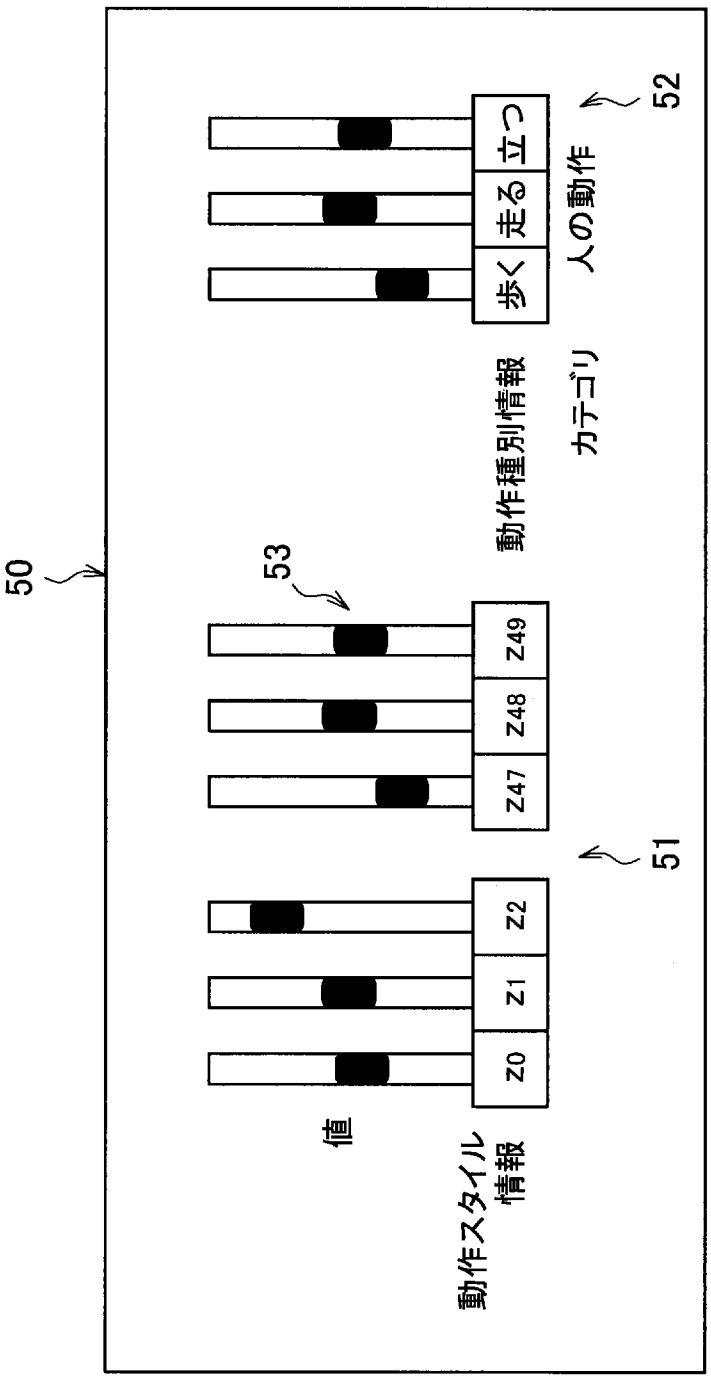
[図18]


15
y:動作種別情報
z:動作スタイル情報
16
生成器
x:センサデータ

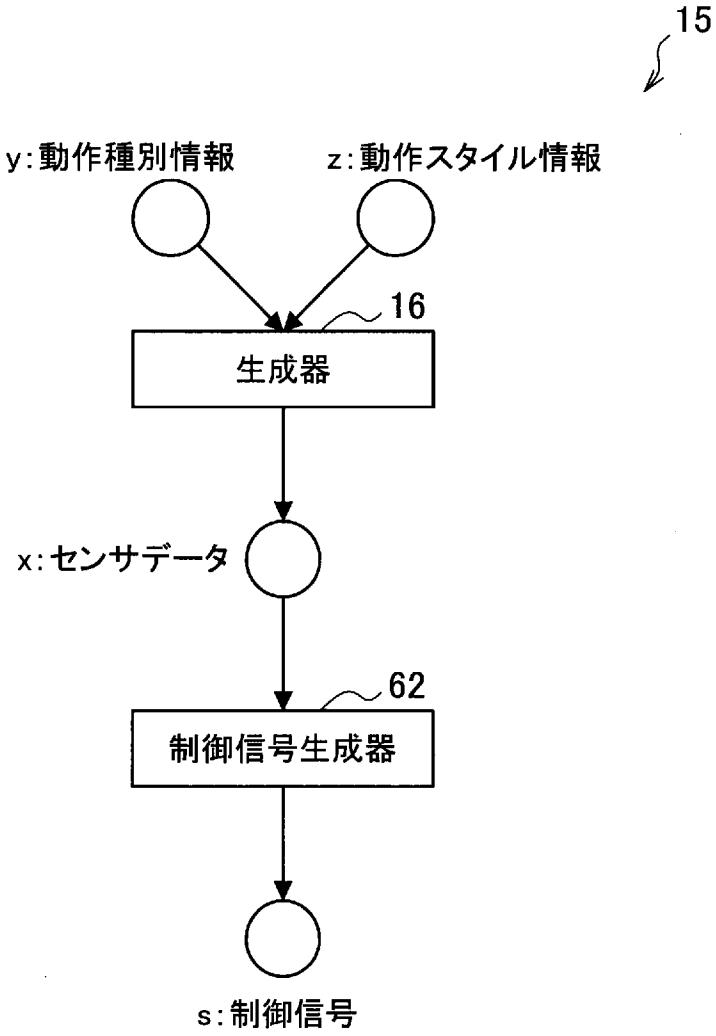
[図19]


スタート
S402
動物体をセンシングしたセンサデータ、
及び当該動物体の動作種別情報の
組み合わせから成る
学習データセットを取得
学習データセット
D = { X, Y }
センサデータ:X = { xn }
動作種別情報:Y = { yn }
S404
ニューラルネットのパラメータを学習
w, u, v = argmin L(X,Y;hw, fu, gv)
S406
学習結果を記憶
識別器のパラメータ
w : y = hw (x)
推論器のパラメータ
u : z = fu (x, y)
生成器のパラメータ
v : x = gv (y, z)
エンド

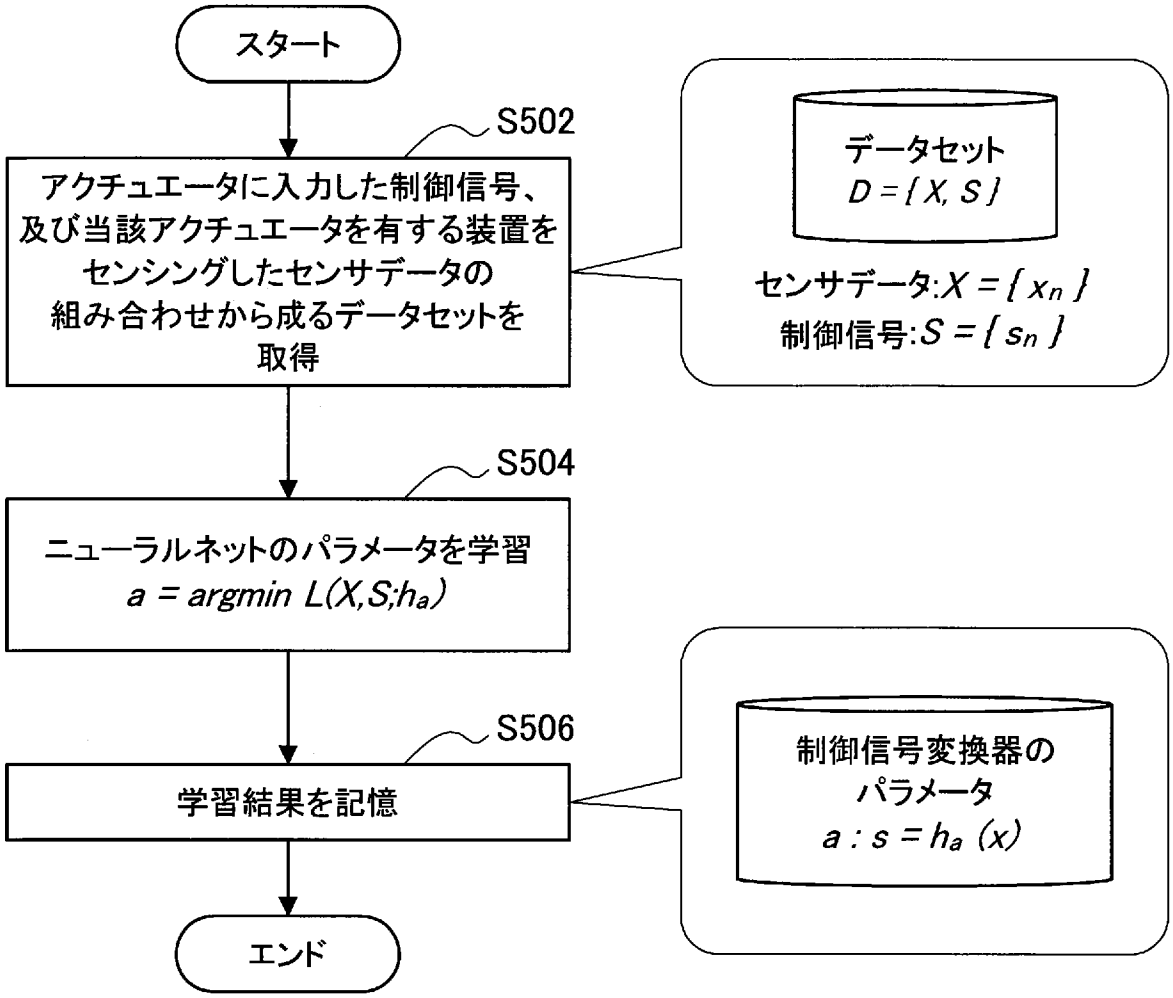
[図20]


50
53
52
51
動作種別情報
カテゴリ
人の動作
歩く
走る
立つ
Z49
Z48
Z47
Z2
Z1
Z0
値
動作スタイル情報

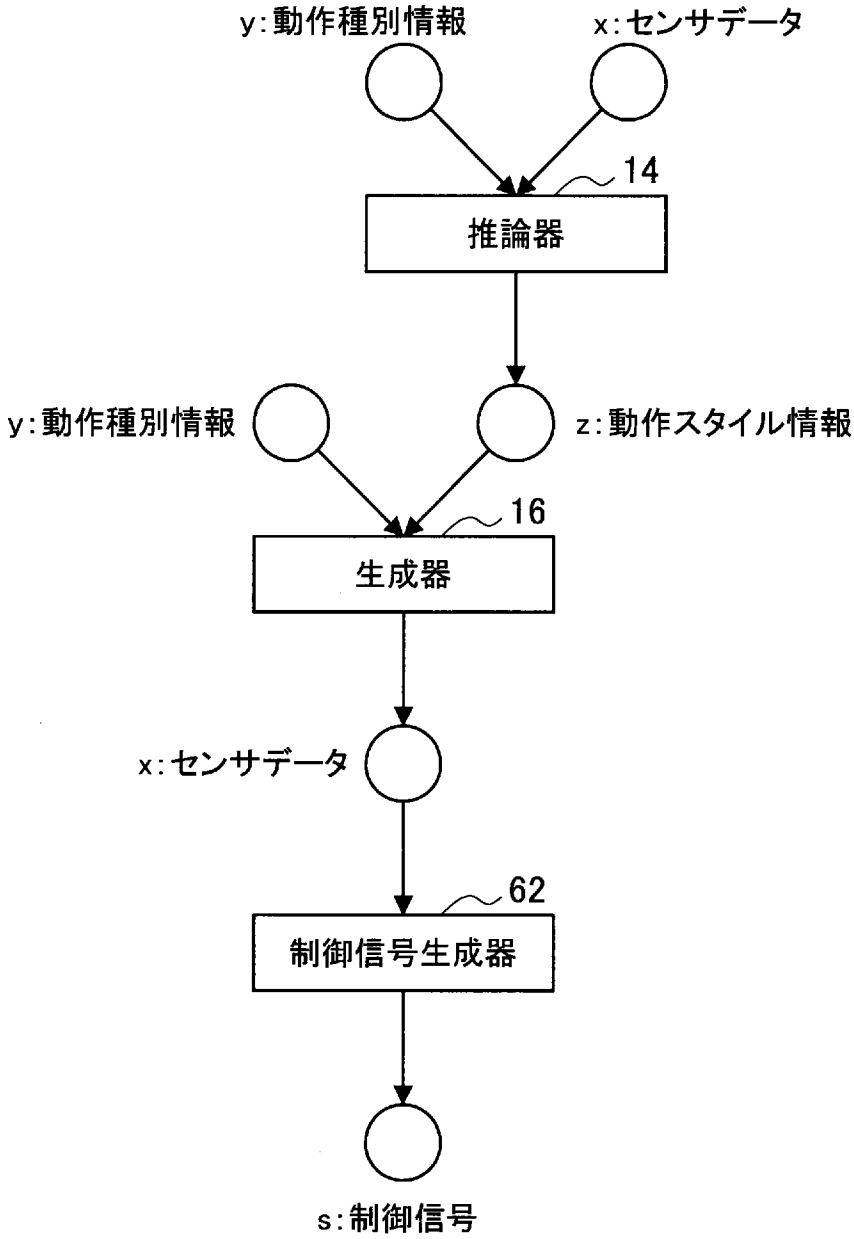
[図21]



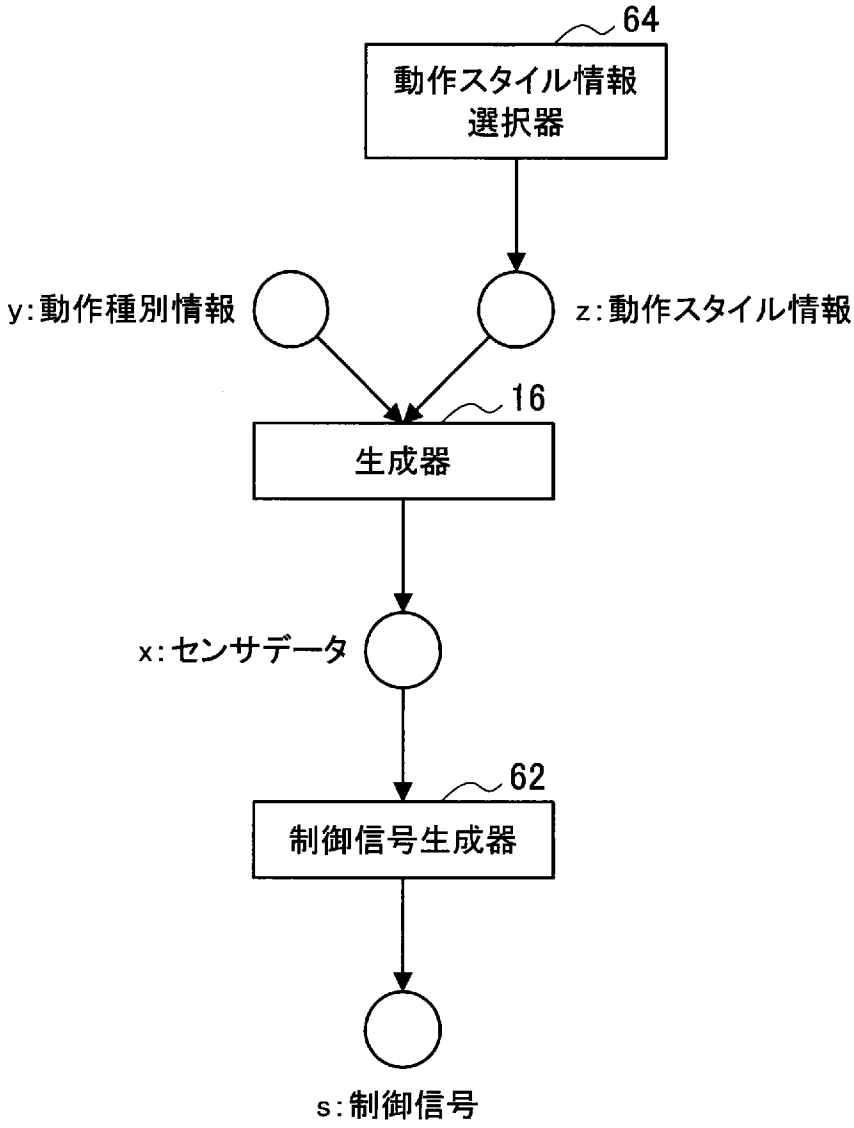
[図22]


スタート
S502
アクチュエータに入力した制御信号、及び当該アクチュエータを有する装置をセンシングしたセンサデータの組み合わせから成るデータセットを取得
データセット
D = { X, S }
センサデータ:X = { xn }
制御信号:S = { sn }
S504
ニューラルネットのパラメータを学習
a = argmin L(X,S;ha)
S506
学習結果を記憶
制御信号変換器のパラメータ
a : s = ha (x)
エンド

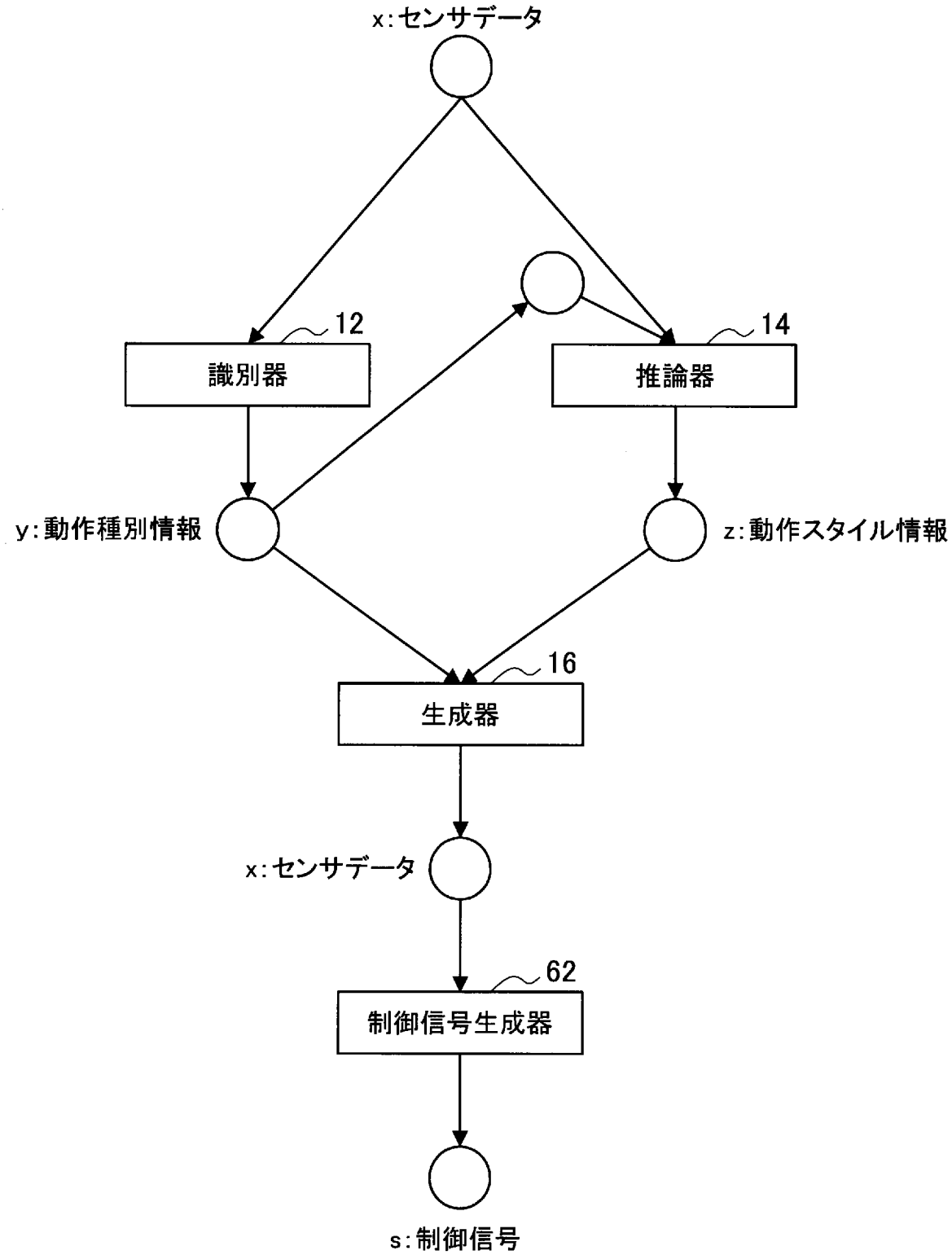
[図23]


y:動作種別情報
x:センサデータ
14
推論器
y:動作種別情報
z:動作スタイル情報
16
生成器
x:センサデータ
62
制御信号生成器
s:制御信号

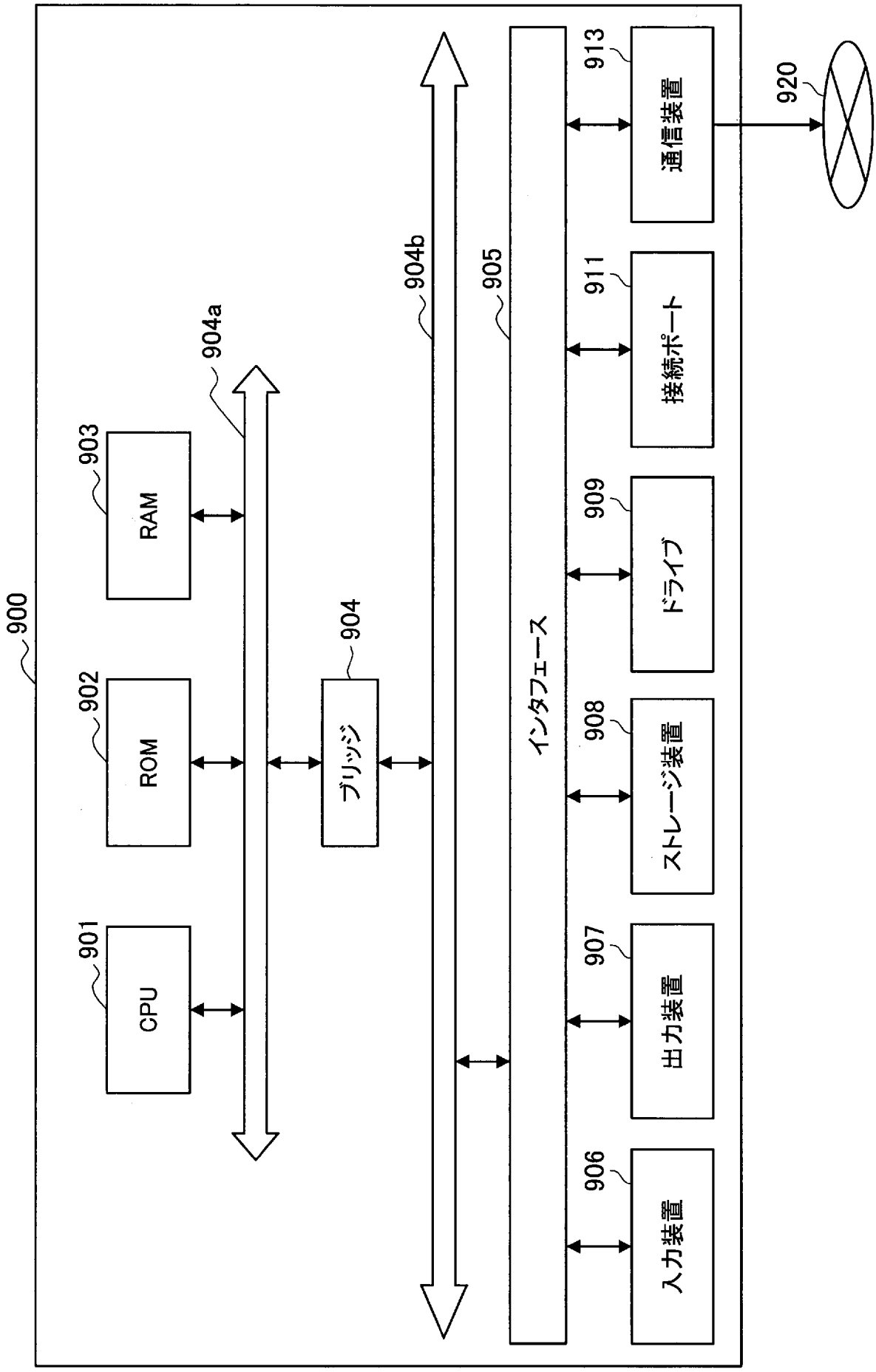
[図24]


64
動作スタイル情報
選択器
y:動作種別情報
z:動作スタイル情報
16
生成器
x:センサデータ
62
制御信号生成器
s:制御信号

[図25]


x:センサデータ
12
識別器
14
推論器
y:動作種別情報
z:動作スタイル情報
16
生成器
x:センサデータ
62
制御信号生成器
s:制御信号

[図26]


900
CPU
901
ROM
902
RAM
903
904a
ブリッジ
904
904b
インタフェース
905
入力装置
906
出力装置
907
ストレージ装置
908
ドライブ
909
接続ポート
911
通信装置
913
920

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2016/087316

A. CLASSIFICATION OF SUBJECT MATTER
G10L13/10(2013.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L13/10

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	Masamitsu MURASE, "Ningen Robot Kyocho no Tameno RNNPB ni yoru Giji Symbol no Kakutoku to sono Kaisosei no Kaiseki", Dai 24 Kai Annual Conference of the Robotics Society of Japan Yokoshu, The Robotics Society of Japan, 2006.09	1-3,9-16, 18-20 17 4-8
Y	Tetsuya OGATA, "Gesture and Sound Generation of Robot based on Cross-Modal Translation using Recurrent Neural Network with Parametric Bias", IEICE Technical Report, 2006.10, vol.106, no.298, pages 27 to 32	17
A	JP 2007-265345 A (Sony Corp.), 11 October 2007 (11.10.2007), paragraphs [0029] to [0037]; fig. 1, 2 & US 2007/0288407 A1 paragraphs [0033] to [0042]; fig. 1, 2	1-20

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

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

Date of the actual completion of the international search	Date of mailing of the international search report
21 February 2017 (21.02.17)	28 February 2017 (28.02.17)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3,Kasumigaseki,Chiyoda-ku, Tokyo 100-8915,Japan	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2016/087316

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2-304493 A (Hitachi, Ltd.), 18 December 1990 (18.12.1990), entire text; all drawings (Family: none)	1-20

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１６／０８７３１６

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G10L13/10(2013.01)i

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　G10L13/10

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１７年
日本国実用新案登録公報	１９９６－２０１７年
日本国登録実用新案公報	１９９４－２０１７年

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

Ｃ．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	村瀬　昌満，人間ロボット協調のためのＲＮＮＰＢによる疑似シンボルの獲得とその階層性の解析，第２４回日本ロボット学会学術講演会予稿集，社団法人日本ロボット学会，2006.09	1-3，9-16，18-20
Y		17
A		4-8
Y	尾形　哲也，ＲＮＮＰＢによる視聴覚情報変換を利用したロボットの身体・音声表現，電子情報通信学会技術研究報告，2006.10，Vol. 106，No. 298，PP. 27-32	17

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

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
２１．０２．２０１７	２８．０２．２０１７

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
東京都千代田区霞が関三丁目４番３号

特許庁審査官（権限のある職員）　５Ｚ　４０６６
安田　勇太
電話番号　０３－３５８１－１１０１　内線　３５９１

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	JP 2007-265345 A（ソニー株式会社）2007.10.11，段落［００２９］－［００３７］，図１，２ &US 2007/0288407 A1, [0033]-[0042], Fig. 1, 2	1-20
A	JP 2-304493 A（株式会社日立製作所）1990.12.18，全文・全図（ファミリーなし）	1-20