

(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
2022年11月3日(03.11.2022)



(10) 国際公開番号

WO 2022/230425 A1

(51) 国際特許分類:
G10L 25/27 (2013.01) G10K 15/04 (2006.01)

(21) 国際出願番号: PCT/JP2022/012222

(22) 国際出願日: 2022年3月17日(17.03.2022)

(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 2021-074895 2021年4月27日(27.04.2021) JP

(71) 出願人:株式会社NTTドコモ(NTT DOCOMO, INC.) [JP/JP]; 〒1006150 東京都千代田区永田町二丁目1番1号 Tokyo (JP).

(72) 発明者:横山 啓太 (YOKOYAMA Keita); 〒1006150 東京都千代田区永田町二丁目1番1号 山王パークタワー 株式会社NTTドコモ知的財産部内 Tokyo (JP).

(74) 代理人:長谷川 芳樹, 外(HASEGAWA Yoshiki et al.); 〒1000005 東京都千代田区丸の内二

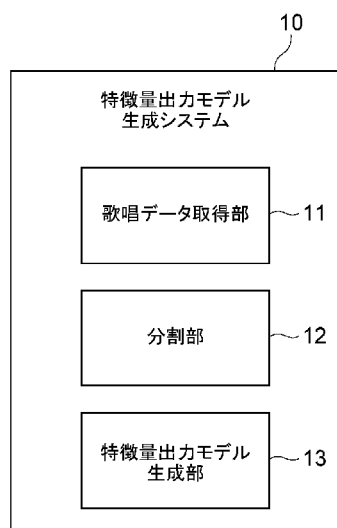
丁目1番1号丸の内 M Y P L A Z A (明治安田生命ビル) 9階 創英国際特許法律事務所 Tokyo (JP).

(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, IT, JM, JO, 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, 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, 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,

(54) Title: FEATURE AMOUNT OUTPUT MODEL GENERATION SYSTEM

(54) 発明の名称: 特徴量出力モデル生成システム



10 ... Feature amount output model generation system
11 ... Singing data acquisition unit
12 ... Division unit
13 ... Feature amount output model generation unit

(57) Abstract: This feature amount output model generation system generates a feature amount output model which suitably outputs a feature amount from singing data. A feature amount output model generation system 10 generates a feature amount output model that receives input of information based on singing data, which is time series voice data associated with singing of a song, and that outputs a feature amount of the singing data, said feature amount output model generation system comprising: a singing data acquisition unit 11 that acquires pieces of singing data for a respective plurality of songs; a division unit 12 that divides each piece of the singing data into a plurality of temporal sections; and a feature amount output model generation unit 13 that generates, by machine learning, the feature amount output model which receives input of information based on the singing data of a divided section and which outputs a feature amount of the divided section from the divided singing data, wherein the feature amount output model generation unit 13 performs machine learning by a reference based on the distance between feature amounts of the singing data associated with the same song and on the distance between feature amounts of the singing data associated with songs different from each other.

WO 2022/230425 A1

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：

一 国際調査報告（条約第21条(3)）

(57) 要約：歌唱データから適切に特徴量を出力する特徴量出力モデルを生成する。特徴量出力モデル生成システム10は、楽曲の歌唱に係る音声の時系列のデータである歌唱データに基づく情報を入力して、当該歌唱データの特徴量を出力する特徴量出力モデルを生成するシステムであって、複数の楽曲についての楽曲毎の歌唱データを取得する歌唱データ取得部11と、各歌唱データを、複数の時間的な区間に分割する分割部12と、分割された歌唱データから、分割された区間の歌唱データに基づく情報を入力して、当該区間の特徴量を出力する特徴量出力モデルを機械学習によって生成する特徴量出力モデル生成部13と、を備え、特徴量出力モデル生成部13は、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習を行う。

明 細 書

発明の名称：特徴量出力モデル生成システム

技術分野

[0001] 本発明は、楽曲の歌唱に係る音声の時系列のデータである歌唱データに基づき情報を入力して、当該歌唱データの特徴量を出力する特徴量出力モデルを生成する特徴量出力モデル生成システムに関する。

背景技術

[0002] 従来、カラオケにおけるユーザの歌唱履歴に基づいて、ユーザに楽曲をレコメンドすることが提案されている（例えば、特許文献1参照）。

先行技術文献

特許文献

[0003] 特許文献1：特開2012-78387号公報

発明の概要

発明が解決しようとする課題

[0004] カラオケで歌唱する際には楽曲のキーが重要になる。そのため、上記のような楽曲のレコメンドと同様に、ユーザが歌唱する際のキーをレコメンドすることが考えられる。キーをレコメンドする際に、ユーザが過去に歌唱した音声のデータである歌唱データを用いてレコメンドすることが考えられる。歌唱データを用いることでより適切なキーをレコメンドすることができる。また、楽曲をレコメンドする際にも、歌唱データを用いることでより適切にレコメンドすることができる。

[0005] レコメンドするキー又は楽曲の決定には、予め用意されたレコメンド用のモデル等を用いることが考えられる。適切なレコメンドを行うために、レコメンド用のモデル等への入力として、歌唱データそのものではなく、歌唱データの特徴量を用いることが考えられる。しかしながら、従来、レコメンドのために歌唱データから特徴量を生成する方法は提案されていなかった。そのため、従来、歌唱データを用いた適切なレコメンドを行うことができな

った。

[0006] 本発明の一実施形態は、上記に鑑みてなされたものであり、歌唱データから適切に特徴量を出力する特徴量出力モデルを生成することができる特徴量出力モデル生成システムを提供することを目的とする。

課題を解決するための手段

[0007] 上記の目的を達成するために、本発明の一実施形態に係る特徴量出力モデル生成システムは、楽曲の歌唱に係る音声の時系列のデータである歌唱データに基づく情報を入力して、当該歌唱データの特徴量を出力する特徴量出力モデルを生成する特徴量出力モデル生成システムであって、特徴量出力モデルの生成に用いる、複数の楽曲についての楽曲毎の歌唱データを取得する歌唱データ取得部と、歌唱データ取得部によって取得された各歌唱データを、複数の時間的な区間に分割する分割部と、分割部によって分割された歌唱データから、分割された区間の歌唱データに基づく情報を入力して、当該区間の歌唱データの特徴量を出力する特徴量出力モデルを機械学習によって生成する特徴量出力モデル生成部と、を備え、特徴量出力モデル生成部は、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習を行う。

[0008] 本発明の一実施形態に係る特徴量出力モデル生成システムでは、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習が行われて、特徴量出力モデルが生成される。このように生成される特徴量出力モデルは、楽曲同士の類似度が考慮されたレコメンド等に用いるのに適切な特徴量を出力することができる。即ち、本発明の一実施形態に係る特徴量出力モデル生成システムによれば、歌唱データから適切に特徴量を出力する特徴量出力モデルを生成することができる。

発明の効果

[0009] 本発明の一実施形態によって生成される特徴量出力モデルは、楽曲同士の類似度が考慮されたレコメンド等に用いるのに適切な特徴量を出力すること

ができる。即ち、本発明の一実施形態によれば、歌唱データから適切に特徴量を出力する特徴量出力モデルを生成することができる。

図面の簡単な説明

[0010] [図1]本発明の実施形態に係る特徴量出力モデル生成システムの機能構成を示す図である。

[図2]特徴量出力モデルの生成に用いられる歌唱データの例を示す図である。

[図3]機械学習による特徴量出力モデルの生成を説明するための図である。

[図4]特徴量出力モデルによって出力される特徴量の例を示すグラフである。

[図5]本発明の実施形態に係る特徴量出力モデル生成システムで実行される処理を示すフローチャートである。

[図6]本発明の実施形態に係る特徴量出力モデル生成システムのハードウェア構成を示す図である。

発明を実施するための形態

[0011] 以下、図面と共に本発明に係る特徴量出力モデル生成システムの実施形態について詳細に説明する。なお、図面の説明においては同一要素には同一符号を付し、重複する説明を省略する。

[0012] 図1に本実施形態に係る特徴量出力モデル生成システム10の機能構成を示す。特徴量出力モデル生成システム10は、特徴量出力モデルを生成するシステム（装置）である。特徴量出力モデル（特徴量生成モデル）は、楽曲の歌唱に係る音声の時系列のデータである歌唱データに基づく情報を入力して、当該歌唱データの特徴量を生成して出力するものである。歌唱データは、例えば、カラオケにおいてユーザが楽曲を歌唱した際に録音されたユーザの音声に基づくデータである。歌唱データが、具体的にどのようなデータであるかは後述する。なお、歌唱データは、楽曲の歌唱に係る音声の時系列のデータであれば、必ずしもユーザの歌唱を前提とするものでなくてもよい。例えば、歌唱データは、カラオケの採点システムにおいて満点を取れる歌唱を想定した音声に基づくデータ（手本データ）であってもよい。

[0013] 特徴量出力モデルによって出力される歌唱データの特徴量は、例えば、カ

ラオケにおいてユーザが楽曲を歌唱する際のキー（キー変更）又は楽曲のレコメンド（推薦）に用いられる。出力される特徴量が、具体的にどのように用いられるかは後述する。特徴量出力モデル生成システム10は、機械学習によって特徴量出力モデルを生成（推論）する。即ち、特徴量出力モデルは、機械学習による学習済モデル（機械学習モデル）である。

[0014] 特徴量出力モデル生成システム10は、例えば、PC（パーソナルコンピュータ）又はサーバ装置等のコンピュータによって実現される。また、特徴量出力モデル生成システム10は、複数のコンピュータ、即ち、コンピュータシステムによって実現されてもよい。

[0015] 引き続き、本実施形態に係る特徴量出力モデル生成システム10の機能を説明する。図1に示すように特徴量出力モデル生成システム10は、歌唱データ取得部11と、分割部12と、特徴量出力モデル生成部13とを備えて構成される。

[0016] 歌唱データ取得部11は、特徴量出力モデルの生成に用いる、複数の楽曲についての楽曲毎の歌唱データを取得する機能部である。歌唱データ取得部11は、時系列のピッチの長さを示すデータを含む歌唱データを取得してもよい。

[0017] 歌唱データは、例えば、図2に示すような、音声の同一のピッチの長さの継続時間を示す情報（楽曲ノート情報）である。また、歌唱データは、楽曲単位の情報であり、例えば、楽曲のIDに対応付けられる等でどの楽曲に係るデータであるかを識別できるようになっている。継続時間は、例えば、楽曲が再生される場合の開始時刻からの経過時間で示される。図2に示す“pitch”の欄の値は、ピッチを示す情報である。具体的には、“pitch”の欄の値は、ノートナンバー（MIDIナンバー、MIDIキー）である。例えば、“pitch”の欄の62との値は、国際式キーのD4に対応する。図2に示す“time_from”及び“time_to”の欄の情報は、それぞれ対応するピッチでの歌唱（発声）が開始されたタイミング及び終了したタイミングを示している。“time_from”及び“time

e__t o”の単位は秒である。図2に示す左側の数値は、情報の通し番号である。

[0018] 例えば、図2の0番の行のデータは、ノートナンバーが62のピッチが、楽曲の開始時間を基準（0秒）として8.005秒の時点から8.337秒の時点まで、0.332秒継続したことを示している。

[0019] なお、前奏があること等により、ユーザが歌唱を開始するのは、通常、楽曲の開始時刻からある程度時間が経ってからである。そのため、最初の“time__from”は0秒ではない。図2に示す歌唱データでは、最初の“time__from”は8.005秒である。

[0020] 図2に示す歌唱データは、例えば、従来のカラオケシステム等で歌唱された音声（カラオケの楽曲ファイル）の解析による得ることができる。歌唱データ取得部11は、例えば、カラオケシステムにおいて、ユーザが歌唱した際に録音された音声から得られた歌唱データを当該カラオケシステムから取得する。あるいは、歌唱データ取得部11は、ユーザの歌唱に係る音声の生データを取得して、従来の方法によって歌唱データを生成して取得してもよい。また、歌唱データ取得部11は、上記以外の方法で歌唱データを取得してもよい。なお、歌唱データ取得部11によって取得される歌唱データは、実際のユーザの歌唱に基づくものである必要はなく、実際の歌唱を想定した図2に示す形式のデータであればよい。また、歌唱データは、必ずしも図2に示すものである必要はなく、楽曲の歌唱に係る音声の時系列のデータであればよい。

[0021] 歌唱データ取得部11は、互いに異なる楽曲について、複数の歌唱データを取得する。なお、歌唱データ取得部11によって取得される歌唱データに係る歌唱したユーザは、任意のユーザでよく、また、複数のユーザであってもよい。また、当該ユーザは、特徴量出力モデルによって特徴量が出力される歌唱データに係る歌唱したユーザ（即ち、例えば、レコメンドの対象となるユーザ）と同一のユーザであっても、別のユーザであってもよい。歌唱データ取得部11は、後述する機械学習を行うのに十分な数の歌唱データを取

得する。歌唱データ取得部 11 は、取得した歌唱データを分割部 12 に出力する。

[0022] 分割部 12 は、歌唱データ取得部 11 によって取得された各歌唱データを、複数の時間的な区間に分割する機能部である。後述するように機械学習は、時間的な区間に分割された歌唱データが用いられて行われる。分割部 12 は、以下のように分割を行う。

[0023] 分割部 12 は、歌唱データ取得部 11 から歌唱データを入力する。分割部 12 は、予め記憶した分割ルールに基づいて、入力した歌唱データ毎に複数の区間に分割する。分割部 12 は、各歌唱データを一定数の区間に分割する。分割部 12 は、歌唱データの最初の “time_from” から最後の “time_to” までの時間帯、即ち、歌唱データに係る時間帯を上記の一定数になるように等分して、等分されたタイミング（時刻）を区切りに用いるタイミングとする。区切りに用いるタイミングが、連続する同一のピッチの時間帯（即ち、1つの “pitch” の値に対応付いた “time_from” から “time_to” までの時間帯）に含まれる場合、当該時間帯の先端及び末尾のうち、当該タイミングに近い方を区間の区切りにする。このように区間に区切ることで、連続する同一のピッチは、複数の区間に分割されることがない。例えば、図 2 に示す 0 番～4 番のデータが、1つの区間（最初の区間）のデータとされる。

[0024] なお、分割する区間の数は、一定数ではなく、楽曲毎に予め設定された数としてもよい。また、区切りに用いるタイミングは、上記のように等分されたものではなく、楽曲毎に予め設定されたタイミングであってもよい。分割部 12 は、分割対象となった歌唱データ、及び分割した区間を示す情報を特徴量出力モデル生成部 13 に出力する。

[0025] 特徴量出力モデル生成部 13 は、分割部 12 によって分割された歌唱データから特徴量出力モデルを機械学習によって生成する機能部である。特徴量出力モデルは、分割された区間の歌唱データに基づく情報を入力して、当該区間の歌唱データの特徴量を出力するモデルである。特徴量出力モデル生成

部13は、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習を行う。

[0026] 特徴量出力モデル生成部13は、同一の楽曲に係る歌唱データの特徴量間の距離が、互いに異なる楽曲に係る歌唱データの特徴量間の距離よりも短くなるように機械学習を行ってもよい。特徴量出力モデル生成部13は、生成途中の特徴量出力モデルによって出力される特徴量間の距離に基づいて、機械学習に用いる歌唱データの区間を決定してもよい。特徴量出力モデル生成部13は、分割部12によって分割された歌唱データに含まれるピッチの長さを示すデータを、連続する同一のピッチ毎にピッチの長さに応じた文字列である単語に変換して、変換した単語に基づく情報を入力する特徴量出力モデルを生成してもよい。

[0027] 特徴量出力モデルによって出力される特徴量は、予め設定された次元数のベクトル（N次元ベクトル）である。即ち、特徴量出力モデルは、Embeddingを行うモデルである。後述するように区間の歌唱データは単語に変換されて、変換された単語から特徴量が生成される。特徴量出力モデルは、例えば、ニューラルネットワークを含んで構成される。より具体的には、特徴量出力モデルは、時系列データに適したLSTM（Long Short Term Memory）である。但し、特徴量出力モデルは、機械学習によって生成され、区間の歌唱データを入力して、当該区間の歌唱データの特徴量を出力するモデルであれば、上記以外のものであってもよい。

[0028] キー又は楽曲のレコメンドを行う際に、楽曲同士の類似度（類似性）はユーザの楽曲の好み又は歌唱しやすさの非常に重要な特徴となり得る。ここで楽曲同士の類似度は、楽曲同士がどの程度同じリズム、旋律及び音階のパターンを持っているか、音の高さが近いかというものである。上記の通り、特徴量出力モデルは、歌唱データに基づく情報を入力するものである。また、特徴量出力モデルは、楽曲に係る歌唱データの特徴量間の距離を考慮して生成される。このようにして生成される特徴量出力モデルによって出力され

る特徴量は、上記の楽曲同士の類似度が考慮されたものとなる。楽曲は、リズム、旋律及び音階のパターン等によって表現されるが、従来、リズム、旋律及び音階のパターンといった個々の特徴に対して楽曲同士の類似度を算出することは困難であった。

[0029] 特徴量出力モデル生成部13は、以下のように特徴量出力モデルを生成する。特徴量出力モデル生成部13は、分割部12から、歌唱データ及び分割した区間を示す情報を入力する。特徴量出力モデル生成部13は、歌唱データの連続する同一のピッチ毎（図2の行毎）にピッチの長さに応じた文字列である単語に変換する。まず、特徴量出力モデル生成部13は、連続する同一のピッチの時間的な長さを算出する。秒単位の当該長さは、“time__to”の値から“time__from”の値を引くことで算出できる。続いて、特徴量出力モデル生成部13は、算出した長さを予め設定された単位時間で割る。予め設定された単位時間は、例えば、0.1秒である。特徴量出力モデル生成部13は、算出した値を端数処理して整数にする。端数処理は、予め設定された方法、例えば、四捨五入、切り捨て又は切り上げによって行われる。特徴量出力モデル生成部13は、ピッチを示す値を、算出した整数だけ連続して並べた文字列を、連続する同一のピッチの単語とする。

[0030] 例えば、図2の0番のピッチのデータについて、まず、時間的な長さが、 $8.337 - 8.005 = 0.332$ 秒と算出される。続いて、算出された値が予め設定された単位時間で割られた値 $0.332 / 0.1 = 3.32$ が算出される。その値が、四捨五入されて整数である3とされる。その数分、ピッチを示す値62が連続して並べられた“626262”が、当該ピッチのデータについての単語（ノートEmbedding）とされる。上記の単語は、ピッチがどの程度の長さ継続したかを示している。上記の例では、ピッチが0.何秒継続したかに基づいて単語化が行われる。単語は、音の高低と長さとを表現している。特徴量出力モデル生成部13は、全てのピッチのデータについて単語に変換する。

[0031] 特徴量出力モデル生成部13は、このように同一のピッチ（同一音）の経

過時間に基づく歌唱データの符号化を行う。これによって不規則に連続する歌唱データ（音情報）を特徴量出力モデルの生成に用いる情報として取り扱うことができる。上記の符号化によって、同一のピッチで長さが若干異なるもの、及び近いピッチで同じ長さのもの等を類似する情報として取り扱えることが期待される。

[0032] 特徴量出力モデル生成部13は、特徴量出力モデルの生成の前処理として、変換した各単語を特徴量出力モデルに入力する入力用特徴量に変換する。なお、以降の説明では、単に特徴量といった場合、特徴量出力モデルの出力である特徴量を指し、単語から変換された特徴量は入力用特徴量と呼ぶ。入力用特徴量は、予め設定された次元数のベクトルである。単語から入力用特徴量の変換は、例えば、従来の自然言語処理の手法によって行われる。具体的には、当該変換は、fastTextによって生成されるモデルによって行われる。機械学習であるfastTextによるモデルの生成に、上記の単語が用いられてもよい。

[0033] fastTextを用いることで、与えられた単語を構成する文字のまとまりを考慮して、字面の近い単語同士を意味のあるまとまりにすることができる。そのため、“7171”と“717171”とは別の単語になるが、その構成要素である“71”に共通性があるため、意味的な距離では近いものとして取り扱われる。

[0034] 特徴量出力モデル生成部13は、区間に含まれる単語から変換された入力用特徴量を、単語が歌唱データに登場する順番にしたものを、当該区間の特徴量出力モデルへの入力とする。例えば、特徴量出力モデルには、入力層に入力用特徴量であるベクトルの次元数のニューロンが設けられる。また、出力層に特徴量であるベクトルの次元数（上述したN）のニューロンが設けられる。特徴量出力モデルは、区間に含まれる入力用特徴量を入力用特徴量毎に順番に入力する。特徴量出力モデルは、区間に含まれる全ての入力用特徴量を入力すると特徴量を出力する。

[0035] 特徴量出力モデル生成部13は、3つの区間の情報を1つのセットとして

、特徴量出力モデル生成のための機械学習を行う。3つの区間は、所定の区間（アンカー）、所定の区間と同一の楽曲の別の区間（ポジティブ）、及び所定の区間と互いに異なる楽曲の別の区間（ネガティブ）である。特徴量出力モデル生成部13は、上記の3つの区間を選択（抽出）する。

[0036] 特徴量出力モデル生成部13は、図3に示す上記のように選択した3つの区間の情報（入力用特徴量の集合）をそれぞれ特徴量出力モデルの入力として機械学習を行う。特徴量出力モデルからは、3つの区間それぞれについて特徴量が出力として得られる。具体的には、図3に示すように入力となるアンカーの情報であるAnchor Seqから、特徴量出力モデルであるEmbedding Netによってアンカーの特徴量であるAnchor Embeddingが得られる。ポジティブの情報であるPositive Seqから、ポジティブの特徴量であるPositive Embeddingが得られる。ネガティブの情報であるNegative Seqから、ネガティブの特徴量であるNegative Embeddingが得られる。なお、図3では、単語が特徴量出力モデルに入力されるように示されているが、実際には特徴量出力モデルには入力用特徴量が入力される。

[0037] 特徴量出力モデル生成部13は、出力される特徴量、即ち、上記のAnchor Embedding、Positive Embedding及びNegative Embeddingに基づく機械学習を行う。特徴量出力モデル生成部13は、図3の右側に示すようにAnchor EmbeddingとPositive Embeddingとの距離（アンカー—ポジティブ間の距離）D1が、Anchor EmbeddingとNegative Embeddingとの距離（アンカー—ネガティブ間の距離）D2よりも短くなるように機械学習を行う。上記の距離は、特徴量のベクトル空間（特徴空間）であるN次元空間におけるユークリッド距離でもよいし、それ以外の距離でもよい。

[0038] 具体的には、特徴量出力モデル生成部13は、損失関数Loss () を以下として機械学習を行う。

$$\text{Loss}(A, P, N) = \text{Max}(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0)$$

上記の式において、Aは、Anchor Seqである。Pは、Positive Seqである。Nは、Negative Seqである。f(X)は、XをEmbedding Netに入力した際の出力として得られるEmbeddingのベクトルである。 $\|f(A) - f(P)\|$ は、アンカー-ポジティブ間である。 $\|f(A) - f(N)\|$ は、アンカー-ネガティブ間の距離である。 α は、予め設定されるハイパーパラメータであり、アンカー-ポジティブ間の距離とアンカー-ネガティブ間の距離との差分をどの程度とするかを示す値である。Max(X, Y)は、X, Yのうち、大きい方の値を関数値とする関数である。損失関数に基づく機械学習自体は、従来と同様に行われればよい。

[0039] 特徴量出力モデル生成部13は、分割部12から入力された情報によって示される区間から、アンカー、ポジティブ及びネガティブの3つの区間を選択して、選択された区間の情報を用いて機械学習を行う。特徴量出力モデル生成部13は、3つの区間の選択と機械学習とを繰り返し行って特徴量出力モデルを生成する。例えば、特徴量出力モデル生成部13は、従来と同様に特徴量出力モデルの生成が予め設定した条件に基づいて収束するまで、あるいは予め設定された規定回数、上記の繰り返しを行って特徴量出力モデルを生成する。

[0040] アンカー、ポジティブ及びネガティブの3つの区間の選択は、以下のように行われる。例えば、特徴量出力モデル生成部13は、ランダムサンプリングで3つの区間を選択する。この場合、特徴量出力モデル生成部13は、上記の繰り返し毎に、即ち、学習のエポック毎にランダムに楽曲を選択して、選択した楽曲に含まれる2つの区間をランダムに選択してアンカー及びポジティブとする。更に、特徴量出力モデル生成部13は、選択した楽曲とは別の楽曲をランダムに選択して、選択した楽曲に含まれる1つの区間をランダムに選択してネガティブとする。

- [0041] あるいは、特徴量出力モデル生成部 13 は、生成途中の特徴量出力モデルを用いてネガティブを選択してもよい。この場合、アンカー及びポジティブを上記と同様に選択する。特徴量出力モデル生成部 13 は、アンカー及びポジティブを含む楽曲とは別の楽曲をランダムに選択する。この際、別の楽曲は複数であってもよい。特徴量出力モデル生成部 13 は、選択した楽曲に含まれる区間をランダムに選択してネガティブの候補とする。この際、ネガティブの候補は複数、例えば、予め設定した数（N 個）である。
- [0042] 特徴量出力モデル生成部 13 は、生成途中の特徴量出力モデルを用いてアンカーの特徴量及びネガティブの候補それぞれの特徴量を算出する。続いて、特徴量出力モデル生成部 13 は、算出した特徴量から、ネガティブの候補それぞれについて、アンカーーネガティブの候補間の距離を算出する。特徴量出力モデル生成部 13 は、算出した距離が予め設定した閾値以内のネガティブの候補を機械学習に用いるネガティブとして決定する。このようにネガティブを採用することで、サンプリング毎の距離計算が必要になる反面、本来、アンカーとの距離が遠くなるべきネガティブに対して集中的に学習が進む。
- [0043] なお、ネガティブの候補は、アンカーの楽曲以外全ての楽曲の全ての区間としてもよいが、距離計算を考慮して全てを候補にするのではなく、上記のように一定数である N 個のネガティブの候補を用いてネガティブを決定してもよい。これによって、機械学習の処理速度を速くすることができる。なお、機械学習に用いる区間の決定の処理は、ミニバッチ処理として行われてもよい。
- [0044] 特徴量出力モデル生成部 13 は、生成した特徴量出力モデルを出力する。例えば、特徴量出力モデルを用いる他の装置又はモジュールに特徴量出力モデルを送信又は出力する。あるいは、特徴量出力モデル生成部 13 は、特徴量出力モデル生成システム 10 に生成した特徴量出力モデルを記憶させて、特徴量出力モデルを用いる他の装置又はモジュールに利用できるようにしてもよい。

[0045] 図4に、特徴量出力モデル生成システム10によって生成された特徴量出力モデルによって出力される特徴量の例を示す。図4において、1つの点が1つの区間の特徴量に対応する。また、同一の色（同一の濃さ）の点が、同一の楽曲の点に対応する。図4においては、高次元のベクトルである特徴量を3次元のベクトルに変換（次元圧縮）している。同一の楽曲に含まれる区間の特徴量は、互いに近接した位置の特徴量となる。図4において、矩形の領域A1, A2に含まれる点は、それぞれの領域A1, A2毎に同一の楽曲に含まれる区間の特徴量に対応している。なお、楽曲が別であっても、曲調及びジャンル等が類似している楽曲間の区間の特徴量は、それらが異なる楽曲間の区間の特徴量と比べて互いに近接している。例えば、互いに異なる楽曲であっても、ポップスの楽曲同士の間の特徴量間の距離は、ポップスと演歌との区間の特徴量間の距離よりも近接している。

[0046] 特徴量出力モデル生成システム10によって生成される学習済モデルである特徴量出力モデルは、人工知能ソフトウェアの一部であるプログラムモジュールとしての利用が想定される。特徴量出力モデルは、例えば、CPU（Central Processing Unit）及びメモリを備えるコンピュータにて用いられ、コンピュータのCPUが、メモリに記憶された特徴量出力モデルからの指令に従って動作する。例えば、コンピュータのCPUが、当該指令に従って、特徴量出力モデルに対して情報を入力して、特徴量出力モデルに応じた演算を行って、特徴量出力モデルから結果を出力するように動作する。具体的には、コンピュータのCPUが、当該指令に従って、ニューラルネットワークの入力層に情報を入力して、ニューラルネットワークにおける学習済の重み付け係数等に基づく演算を行って、ニューラルネットワークの出力層から結果を出力するように動作する。

[0047] 上記のように生成された特徴量出力モデルは、以下のように用いられる。例えば、特徴量出力モデルは、カラオケにおいて楽曲が歌唱される際のキーのレコメンド又は楽曲のレコメンドに用いられる。具体的には、レコメンドの対象のユーザの過去の歌唱実績のデータから上記のレコメンドを行う際に

用いられる。歌唱実績のデータには、上記の歌唱データ取得部11によって取得される歌唱データと同様の歌唱データが含まれる。

[0048] 当該歌唱データから、特徴量出力モデルが用いられて特徴量 (Embedding) が生成される。なお、この際、歌唱データから、特徴量出力モデル生成システム10における特徴量出力モデルの生成の際と同様に特徴量出力モデルへの入力とされる情報 (例えば、上述した入力用特徴量) が生成される。なお、特徴量出力モデルへの入力とされる情報は、歌唱データの区間毎の情報である。区間への分割も上記と同様に行われる。

[0049] 区間毎の特徴量は、楽曲毎に区間の順番に並べられて連結されて楽曲毎の特徴量とされてもよい。即ち、特徴量の concatenate が行われてもよい。単純な連結の場合、楽曲当たりの特徴量であるベクトルの次元は、特徴量出力モデルにより出力される特徴量であるベクトルの次元数×楽曲の区間の数となる。また、連結の際にベクトルを集約処理 (平均又は加法) することでより低次元の楽曲毎の特徴量としてもよい。

[0050] レコメンドの対象のユーザの過去の歌唱実績のデータを上記のように歌唱に係る楽曲毎の特徴量とすることで、レコメンドに用いることができる。キーをレコメンドする場合には、ユーザが歌唱しようとする楽曲のキー毎の手本データである歌唱データを、上記と同様に特徴量出力モデルを用いて特徴量にしたものを用いてもよい。なお、手本データである歌唱データのキーを変更する場合は、図2に示すピッチの値を変えればよい。例えば、キーを1つ上げる場合、ピッチの値が62であったら63に変更する。これにより、ピッチに対応する変更前の単語が“626262”であったら、変更後の単語は“636363”となる。

[0051] レコメンドは、例えば、上記の歌唱実績に係る特徴量、及び手本データに係る特徴量を入力として、ユーザの歌唱が手本データに係るキーにどの程度合致しているかを示す値を出力するレコメンド用モデル (Dense) を用いることにより行われる。複数の互いに異なるキーの手本データを用いて上記の値を算出して、最も高い値のキーをレコメンドすればよい。レコメンド

用モデルは、従来の機械学習の方法等を用いて作成すればよい。

- [0052] 上記のようにある楽曲のキーの高低を変化させることにより、例えば、女性に対するレコメンド対象の楽曲が男性歌手の楽曲であっても、キーを変更することで女性が過去に歌った女性歌手の楽曲に近いという類似度を求めることが可能になる。
- [0053] 楽曲をレコメンドする場合も、ユーザに歌唱をレコメンドする候補となる楽曲の手本データを用いればよい。その場合、レコメンドは、例えば、上記の歌唱実績に係る特徴量、及び手本データに係る特徴量を入力として、手本データに係る楽曲をどの程度レコメンドすべきかを示す値を出力するレコメンド用モデルを用いることにより行われる。複数の互いに異なる楽曲の手本データを用いて上記の値を算出して、最も高い値の楽曲をレコメンドすればよい。
- [0054] 上記は、特徴量出力モデルによって出力される特徴量を用いたレコメンドの例であり、上記以外のレコメンドに用いられてもよい。また、特徴量出力モデルは、レコメンド以外の用途に用いられてもよい。以上が、本実施形態に係る特徴量出力モデル生成システム10の機能である。
- [0055] 引き続き、図5のフローチャートを用いて、本実施形態に係る特徴量出力モデル生成システム10で実行される処理（特徴量出力モデル生成システム10が行う動作方法）を説明する。
- [0056] 本処理では、歌唱データ取得部11によって、複数の楽曲についての楽曲毎の歌唱データが取得される（S01）。続いて、分割部12によって、各歌唱データが、複数の時間的な区間に分割される（S02）。続いて、特徴量出力モデル生成部13によって、各区間の歌唱データが、特徴量出力モデルの入力の形式に変換される（S03）。具体的には、歌唱データは、ピッチの長さに応じた単語に変換される。また、当該単語は入力用特徴量に変換される。
- [0057] 続いて、特徴量出力モデル生成部13によって、機械学習に用いられる区間が決定される（S04）。具体的には、上述したアンカー、ポジティブ及

びネガティブの3つの区間が決定される。続いて、特徴量出力モデル生成部13によって、決定された区間の情報が用いられて特徴量出力モデルを生成するための機械学習が行われる(S05)。具体的には、上述したように、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習が行われる。続いて、特徴量出力モデル生成部13によって、機械学習を終了するか否かが判断される(S06)。

[0058] 機械学習を終了しないと判断された場合、上記のS04～S06の処理が再度行われる。機械学習を終了すると判断された場合、特徴量出力モデル生成部13から、生成された特徴量出力モデルが出力される(S07)。本実施形態に係る特徴量出力モデル生成システム10で実行される処理である。

[0059] 上述したように本実施形態では、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習が行われて、特徴量出力モデルが生成される。このように生成される特徴量出力モデルは、例えば、図4を用いて説明したように楽曲同士の類似度が考慮されたレコメンド等に用いるのに適切な特徴量を出力することができる。即ち、本実施形態によれば、歌唱データから適切に特徴量を出力する特徴量出力モデルを生成することができる。この結果、類似楽曲を抽出したり、キー又は楽曲のレコメンドをより精度よく行ったりすることができる。

[0060] また、上述した実施形態のように、同一の楽曲に係る歌唱データの特徴量間の距離が、互いに異なる楽曲に係る歌唱データの特徴量間の距離よりも短くなるように機械学習を行ってもよい。この構成によれば、確実に歌唱データから適切に特徴量を出力する特徴量出力モデルを生成することができる。但し、機械学習は、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって行われるものであれば、必ずしも上記のように行われる必要はない。

[0061] また、上述した実施形態のように、生成途中の特徴量出力モデルによって

出力される特徴量間の距離に基づいて、機械学習に用いる歌唱データの区間（例えば、上述したアンカー、ポジティブ及びネガティブの3つの区間）を決定してもよい。この構成によれば、上述したように機械学習を効率的に行うことができ、その結果、同一の学習処理に対してより学習が進んだ特徴量出力モデルを生成することができる。あるいは、特徴量出力モデルの生成における学習処理を軽減することができる。但し、機械学習に用いる区間の決定は上記のように行われる必要はない。

[0062] また、上述した実施形態のように、歌唱データは、時系列のピッチの長さを示すデータを含むものとしてもよい。この構成によれば、確実かつ適切に特徴量出力モデルを生成することができる。また、分割された歌唱データに含まれるピッチの長さを示すデータを、連続する同一のピッチ毎にピッチの長さに応じた文字列である単語に変換して、変換した単語に基づく情報を入力する特徴量出力モデルを生成してもよい。この構成によれば、特徴量出力モデルを生成する上で、上述した従来の自然言語処理の手法を用いて適切かつ容易に歌唱データを扱うことができる。その結果、特徴量出力モデルを適切かつ容易に生成することができる。但し、歌唱データは、上記のものである必要はなく、楽曲の歌唱に係る音声の時系列のデータであればよい。また、歌唱データは、上記のように単語として扱われる必要はなく、特徴量出力モデルへの入力とできる形式であればどのように扱われてもよい。

[0063] なお、上記実施形態の説明に用いたブロック図は、機能単位のブロックを示している。これらの機能ブロック（構成部）は、ハードウェア及びソフトウェアの少なくとも一方の任意の組み合わせによって実現される。また、各機能ブロックの実現方法は特に限定されない。すなわち、各機能ブロックは、物理的又は論理的に結合した1つの装置を用いて実現されてもよいし、物理的又は論理的に分離した2つ以上の装置を直接的又は間接的に（例えば、有線、無線などを用いて）接続し、これら複数の装置を用いて実現されてもよい。機能ブロックは、上記1つの装置又は上記複数の装置にソフトウェアを組み合わせて実現されてもよい。

[0064] 機能には、判断、決定、判定、計算、算出、処理、導出、調査、探索、確認、受信、送信、出力、アクセス、解決、選択、選定、確立、比較、想定、期待、見做し、報知 (broadcasting)、通知 (notifying)、通信 (communicating)、転送 (forwarding)、構成 (configuring)、再構成 (reconfiguring)、割り当て (allocating、mapping)、割り振り (assigning) などがあるが、これらに限られない。たとえば、送信を機能させる機能ブロック (構成部) は、送信部 (transmitting unit) 又は送信機 (transmitter) と呼称される。いずれも、上述したとおり、実現方法は特に限定されない。

[0065] 例えば、本開示の一実施の形態における特徴量出力モデル生成システム 10 は、本開示の情報処理を行うコンピュータとして機能してもよい。図 6 は、本開示の一実施の形態に係る特徴量出力モデル生成システム 10 のハードウェア構成の一例を示す図である。上述の特徴量出力モデル生成システム 10 は、物理的には、プロセッサ 1001、メモリ 1002、ストレージ 1003、通信装置 1004、入力装置 1005、出力装置 1006、バス 1007 などを含むコンピュータ装置として構成されてもよい。

[0066] なお、以下の説明では、「装置」という文言は、回路、デバイス、ユニットなどに読み替えることができる。特徴量出力モデル生成システム 10 のハードウェア構成は、図に示した各装置を 1 つ又は複数含むように構成されてもよいし、一部の装置を含まずに構成されてもよい。

[0067] 特徴量出力モデル生成システム 10 における各機能は、プロセッサ 1001、メモリ 1002 などのハードウェア上に所定のソフトウェア (プログラム) を読み込ませることによって、プロセッサ 1001 が演算を行い、通信装置 1004 による通信を制御したり、メモリ 1002 及びストレージ 1003 におけるデータの読み出し及び書き込みの少なくとも一方を制御したりすることによって実現される。

[0068] プロセッサ 1001 は、例えば、オペレーティングシステムを動作させてコンピュータ全体を制御する。プロセッサ 1001 は、周辺装置とのインターフェース、制御装置、演算装置、レジスタなどを含む中央処理装置 (CP

U : Central Processing Unit) によって構成されてもよい。例えば、上述の特微量出力モデル生成システム 10 における各機能は、プロセッサ 1001 によって実現されてもよい。

[0069] また、プロセッサ 1001 は、プログラム（プログラムコード）、ソフトウェアモジュール、データなどを、ストレージ 1003 及び通信装置 1004 の少なくとも一方からメモリ 1002 に読み出し、これらに従って各種の処理を実行する。プログラムとしては、上述の実施の形態において説明した動作の少なくとも一部をコンピュータに実行させるプログラムが用いられる。例えば、特微量出力モデル生成システム 10 における各機能は、メモリ 1002 に格納され、プロセッサ 1001 において動作する制御プログラムによって実現されてもよい。上述の各種処理は、1つのプロセッサ 1001 によって実行される旨を説明してきたが、2以上のプロセッサ 1001 により同時又は逐次に実行されてもよい。プロセッサ 1001 は、1以上のチップによって実装されてもよい。なお、プログラムは、電気通信回線を介してネットワークから送信されても良い。

[0070] メモリ 1002 は、コンピュータ読み取り可能な記録媒体であり、例えば、ROM (Read Only Memory)、EPROM (Erasable Programmable ROM)、EEPROM (Electrically Erasable Programmable ROM)、RAM (Random Access Memory) などの少なくとも1つによって構成されてもよい。メモリ 1002 は、レジスタ、キャッシュ、メインメモリ（主記憶装置）などと呼ばれてもよい。メモリ 1002 は、本開示の一実施の形態に係る情報処理を実施するために実行可能なプログラム（プログラムコード）、ソフトウェアモジュールなどを保存することができる。

[0071] ストレージ 1003 は、コンピュータ読み取り可能な記録媒体であり、例えば、CD-ROM (Compact Disc ROM) などの光ディスク、ハードディスクドライブ、フレキシブルディスク、光磁気ディスク（例えば、コンパクトディスク、デジタル多用途ディスク、Blu-ray（登録商標）ディスク）、スマートカード、フラッシュメモリ（例えば、カード、スティック、キ

ードライブ)、フロッピー(登録商標)ディスク、磁気ストリップなどの少なくとも1つによって構成されてもよい。ストレージ1003は、補助記憶装置と呼ばれてもよい。特徴量出力モデル生成システム10が備える記憶媒体は、例えば、メモリ1002及びストレージ1003の少なくとも一方を含むデータベース、サーバその他の適切な媒体であってもよい。

[0072] 通信装置1004は、有線ネットワーク及び無線ネットワークの少なくとも一方を介してコンピュータ間の通信を行うためのハードウェア(送受信デバイス)であり、例えばネットワークデバイス、ネットワークコントローラ、ネットワークカード、通信モジュールなどともいう。

[0073] 入力装置1005は、外部からの入力を受け付ける入力デバイス(例えば、キーボード、マウス、マイクロフォン、スイッチ、ボタン、センサなど)である。出力装置1006は、外部への出力を実施する出力デバイス(例えば、ディスプレイ、スピーカー、LEDランプなど)である。なお、入力装置1005及び出力装置1006は、一体となった構成(例えば、タッチパネル)であってもよい。

[0074] また、プロセッサ1001、メモリ1002などの各装置は、情報を通信するためのバス1007によって接続される。バス1007は、単一のバスを用いて構成されてもよいし、装置間ごとに異なるバスを用いて構成されてもよい。

[0075] また、特徴量出力モデル生成システム10は、マイクロプロセッサ、デジタル信号プロセッサ(DSP: Digital Signal Processor)、ASIC(Application Specific Integrated Circuit)、PLD(Programmable Logic Device)、FPGA(Field Programmable Gate Array)などのハードウェアを含んで構成されてもよく、当該ハードウェアにより、各機能ブロックの一部又は全てが実現されてもよい。例えば、プロセッサ1001は、これらのハードウェアの少なくとも1つを用いて実装されてもよい。

[0076] 本開示において説明した各態様/実施形態の処理手順、シーケンス、フローチャートなどは、矛盾の無い限り、順序を入れ替えてもよい。例えば、本

開示において説明した方法については、例示的な順序を用いて様々なステップの要素を提示しており、提示した特定の順序に限定されない。

[0077] 入出力された情報等は特定の場所（例えば、メモリ）に保存されてもよいし、管理テーブルを用いて管理してもよい。入出力される情報等は、上書き、更新、又は追記され得る。出力された情報等は削除されてもよい。入力された情報等は他の装置へ送信されてもよい。

[0078] 判定は、1ビットで表される値（0か1か）によって行われてもよいし、真偽値（Boolean：true又はfalse）によって行われてもよいし、数値の比較（例えば、所定の値との比較）によって行われてもよい。

[0079] 本開示において説明した各態様／実施形態は単独で用いてもよいし、組み合わせで用いてもよいし、実行に伴って切り替えて用いてもよい。また、所定の情報の通知（例えば、「Xであること」の通知）は、明示的に行うものに限られず、暗黙的（例えば、当該所定の情報の通知を行わない）ことによって行われてもよい。

[0080] 以上、本開示について詳細に説明したが、当業者にとっては、本開示が本開示中に説明した実施形態に限定されるものではないということは明らかである。本開示は、請求の範囲の記載により定まる本開示の趣旨及び範囲を逸脱することなく修正及び変更態様として実施することができる。したがって、本開示の記載は、例示説明を目的とするものであり、本開示に対して何ら制限的な意味を有するものではない。

[0081] ソフトウェアは、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語と呼ばれるか、他の名称で呼ばれるかを問わず、命令、命令セット、コード、コードセグメント、プログラムコード、プログラム、サブプログラム、ソフトウェアモジュール、アプリケーション、ソフトウェアアプリケーション、ソフトウェアパッケージ、ルーチン、サブルーチン、オブジェクト、実行可能ファイル、実行スレッド、手順、機能などを意味するよう広く解釈されるべきである。

[0082] また、ソフトウェア、命令、情報などは、伝送媒体を介して送受信されて

もよい。例えば、ソフトウェアが、有線技術（同軸ケーブル、光ファイバケーブル、ツイストペア、デジタル加入者回線（DSL: Digital Subscriber Line）など）及び無線技術（赤外線、マイクロ波など）の少なくとも一方を使用してウェブサイト、サーバ、又は他のリモートソースから送信される場合、これらの有線技術及び無線技術の少なくとも一方は、伝送媒体の定義内に含まれる。

- [0083] 本開示において使用する「システム」及び「ネットワーク」という用語は、互換的に使用される。
- [0084] また、本開示において説明した情報、パラメータなどは、絶対値を用いて表されてもよいし、所定の値からの相対値を用いて表されてもよいし、対応する別の情報を用いて表されてもよい。
- [0085] 本開示で使用する「判断(determining)」、「決定(determining)」という用語は、多種多様な動作を包含する場合がある。「判断」、「決定」は、例えば、判定(judging)、計算(calculating)、算出(computing)、処理(processing)、導出(deriving)、調査(investigating)、探索(looking up、search、inquiry)（例えば、テーブル、データベース又は別のデータ構造での探索）、確認(ascertaining)した事を「判断」「決定」したとみなす事などを含み得る。また、「判断」、「決定」は、受信(receiving)（例えば、情報を受信すること）、送信(transmitting)（例えば、情報を送信すること）、入力(input)、出力(output)、アクセス(accessing)（例えば、メモリ中のデータにアクセスすること）した事を「判断」「決定」したとみなす事などを含み得る。また、「判断」、「決定」は、解決(resolving)、選択(selecting)、選定(choosing)、確立(establishing)、比較(comparing)などした事を「判断」「決定」したとみなす事を含み得る。つまり、「判断」「決定」は、何らかの動作を「判断」「決定」したとみなす事を含み得る。また、「判断（決定）」は、「想定する(assuming)」、「期待する(expecting)」、「みなす(considering)」などで読み替えられてもよい。
- [0086] 「接続された(connected)」、「結合された(coupled)」という用語、又は

これらのあらゆる変形は、2又はそれ以上の要素間の直接的又は間接的なあらゆる接続又は結合を意味し、互いに「接続」又は「結合」された2つの要素間に1又はそれ以上の中間要素が存在することを含むことができる。要素間の結合又は接続は、物理的なものであっても、論理的なものであっても、或いはこれらの組み合わせであってもよい。例えば、「接続」は「アクセス」で読み替えられてもよい。本開示で使用する場合、2つの要素は、1又はそれ以上の電線、ケーブル及びプリント電気接続の少なくとも一つを用いて、並びにいくつかの非限定的かつ非包括的な例として、無線周波数領域、マイクロ波領域及び光（可視及び不可視の両方）領域の波長を有する電磁エネルギーなどを用いて、互いに「接続」又は「結合」されることができると考えることができる。

[0087] 本開示において使用する「に基づいて」という記載は、別段に明記されていない限り、「のみに基づいて」を意味しない。言い換えれば、「に基づいて」という記載は、「のみに基づいて」と「に少なくとも基づいて」の両方を意味する。

[0088] 本開示において使用する「第1の」、「第2の」などの呼称を使用した要素へのいかなる参照も、それらの要素の量又は順序を全般的に限定しない。これらの呼称は、2つ以上の要素間を区別する便利な方法として本開示において使用され得る。したがって、第1及び第2の要素への参照は、2つの要素のみが採用され得ること、又は何らかの形で第1の要素が第2の要素に先行しなければならないことを意味しない。

[0089] 本開示において、「含む (include)」、「含んでいる (including)」及びそれらの変形が使用されている場合、これらの用語は、用語「備える (comprising)」と同様に、包括的であることが意図される。さらに、本開示において使用されている用語「又は (or)」は、排他的論理和ではないことが意図される。

[0090] 本開示において、例えば、英語でのa, an及びtheのように、翻訳により冠詞が追加された場合、本開示は、これらの冠詞の後に続く名詞が複数形であ

ることを含んでもよい。

[0091] 本開示において、「AとBが異なる」という用語は、「AとBが互いに異なる」ことを意味してもよい。なお、当該用語は、「AとBがそれぞれCと異なる」ことを意味してもよい。「離れる」、「結合される」などの用語も、「異なる」と同様に解釈されてもよい。

符号の説明

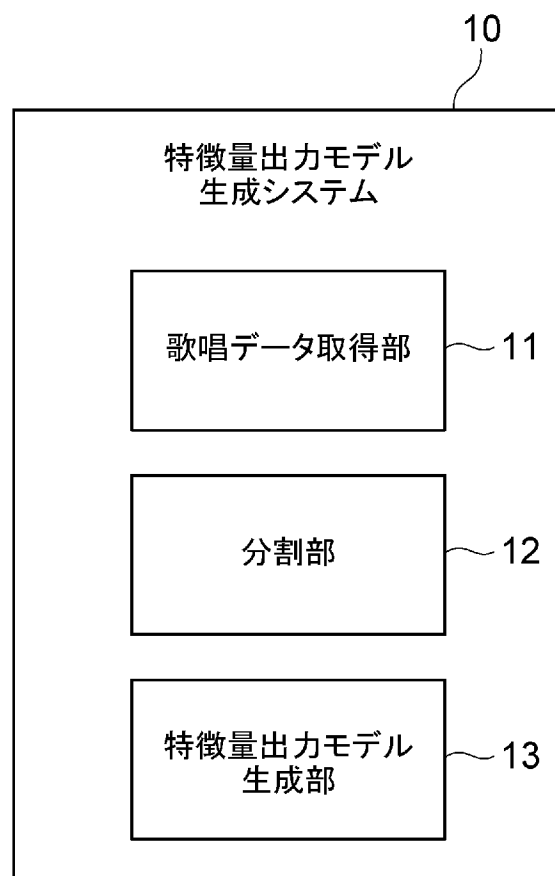
[0092] 10…特徴量出力モデル生成システム、11…歌唱データ取得部、12…分割部、13…特徴量出力モデル生成部、1001…プロセッサ、1002…メモリ、1003…ストレージ、1004…通信装置、1005…入力装置、1006…出力装置、1007…バス。

請求の範囲

- [請求項1] 楽曲の歌唱に係る音声の時系列のデータである歌唱データに基づく情報を入力して、当該歌唱データの特徴量を出力する特徴量出力モデルを生成する特徴量出力モデル生成システムであって、
- 特徴量出力モデルの生成に用いる、複数の楽曲についての楽曲毎の歌唱データを取得する歌唱データ取得部と、
- 前記歌唱データ取得部によって取得された各歌唱データを、複数の時間的な区間に分割する分割部と、
- 前記分割部によって分割された歌唱データから、分割された区間の歌唱データに基づく情報を入力して、当該区間の歌唱データの特徴量を出力する特徴量出力モデルを機械学習によって生成する特徴量出力モデル生成部と、を備え、
- 前記特徴量出力モデル生成部は、同一の楽曲に係る歌唱データの特徴量間の距離と、互いに異なる楽曲に係る歌唱データの特徴量間の距離とに基づく基準によって機械学習を行う特徴量出力モデル生成システム。
- [請求項2] 前記特徴量出力モデル生成部は、同一の楽曲に係る歌唱データの特徴量間の距離が、互いに異なる楽曲に係る歌唱データの特徴量間の距離よりも短くなるように機械学習を行う請求項1に記載の特徴量出力モデル生成システム。
- [請求項3] 前記特徴量出力モデル生成部は、生成途中の特徴量出力モデルによって出力される特徴量間の距離に基づいて、機械学習に用いる歌唱データの区間を決定する請求項1又は2に記載の特徴量出力モデル生成システム。
- [請求項4] 前記歌唱データ取得部は、時系列のピッチの長さを示すデータを含む歌唱データを取得する請求項1～3の何れか一項に記載の特徴量出力モデル生成システム。
- [請求項5] 前記特徴量出力モデル生成部は、前記分割部によって分割された歌

唱データに含まれるピッチの長さを示すデータを、連続する同一のピッチ毎にピッチの長さに応じた文字列である単語に変換して、変換した単語に基づく情報を入力する特徴量出力モデルを生成する請求項4に記載の特徴量出力モデル生成システム。

[図1]

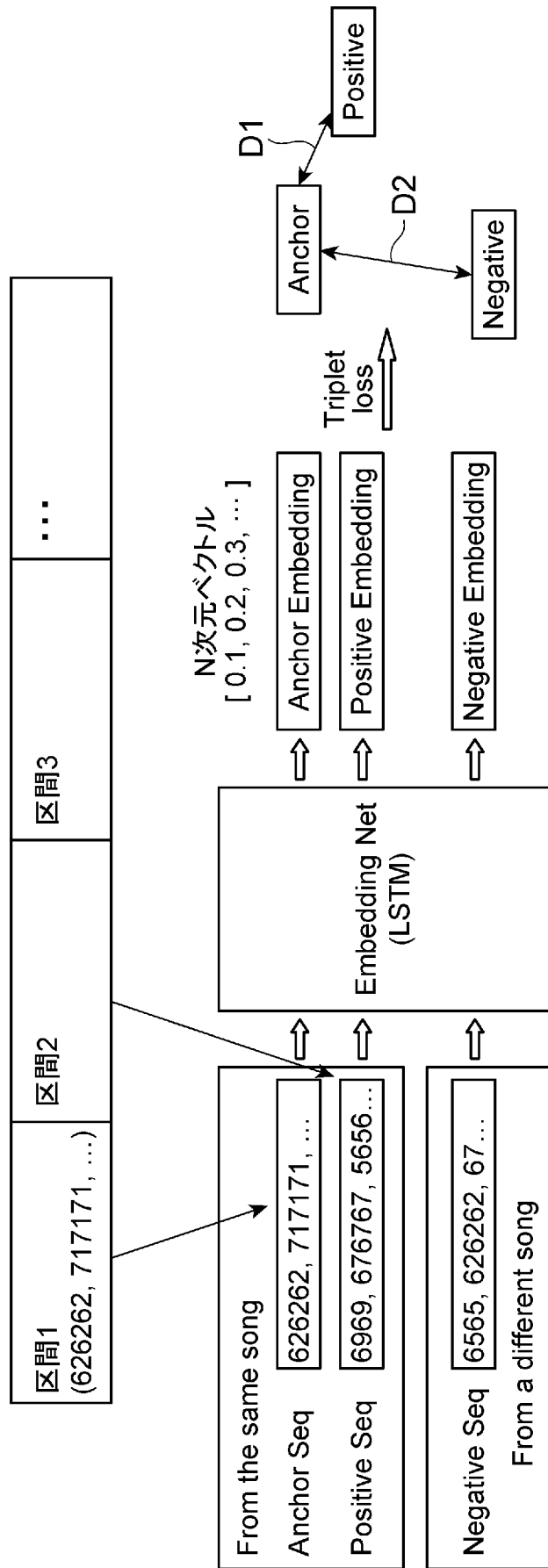


[図2]

	time_from	time_to	pitch
0	8.005	8.337	62
1	8.339	8.671	71
2	8.672	9.004	69
3	9.005	9.337	67
4	9.338	9.670	69
...

626262,
717171,
696969,
676767,
...

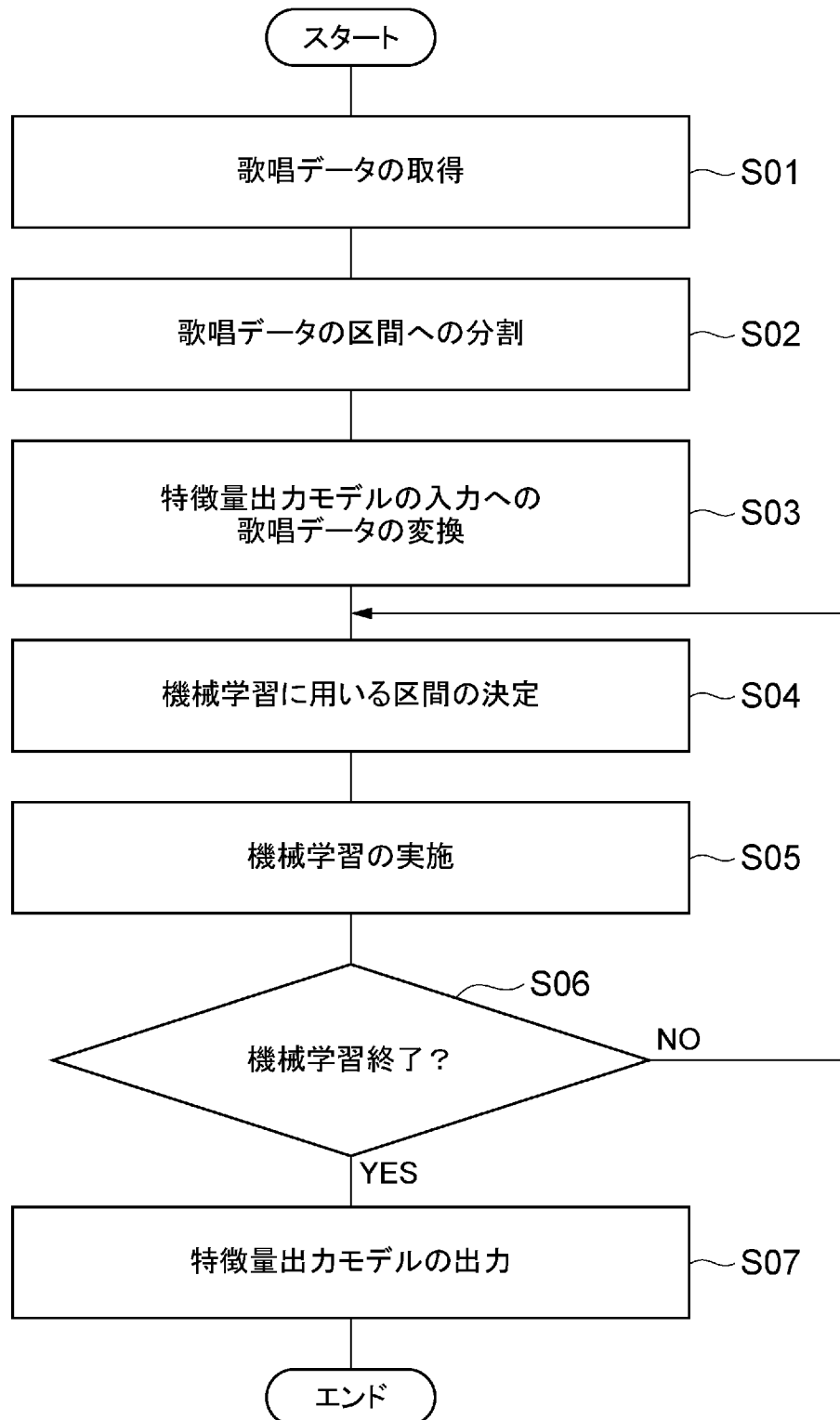
[図3]



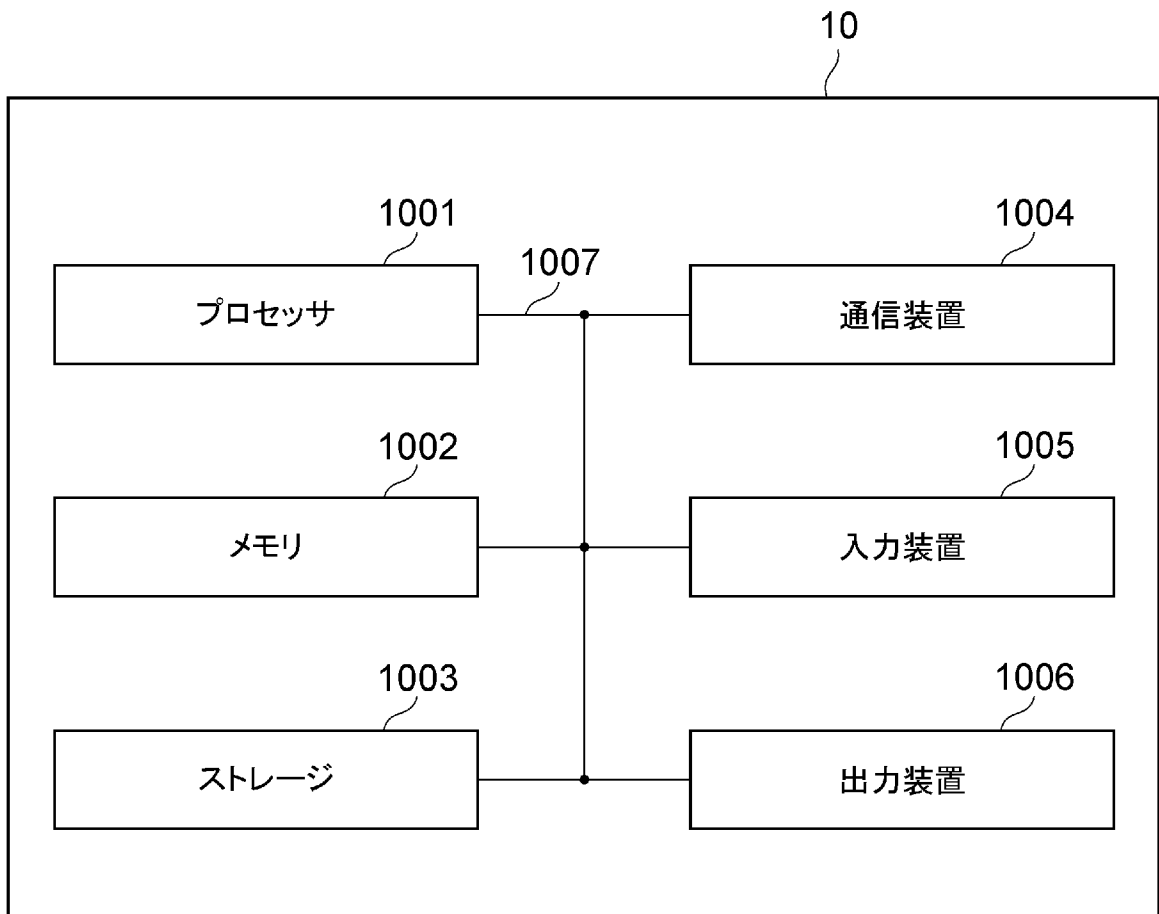
[図4]



[図5]



[図6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/012222

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G10L 25/27</i> (2013.01)i; <i>G10K 15/04</i> (2006.01)i FI: G10L25/27; G10K15/04 302D		
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) G10L25/00-25/93; G10K15/04		
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 1971-2022 Registered utility model specifications of Japan 1996-2022 Published registered utility model applications of Japan 1994-2022		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-510176 A (MUSICGENOME.COM INC.) 02 April 2004 (2004-04-02) paragraphs [0005]-[0026]	1-5
A	WO 2021/005984 A1 (NTT DOCOMO, INC.) 14 January 2021 (2021-01-14) entire text	1-5
A	WO 2021/005985 A1 (NTT DOCOMO, INC.) 14 January 2021 (2021-01-14) entire text	1-5
A	US 2015/0220633 A1 (APERTURE INVESTMENTS, LLC) 06 August 2015 (2015-08-06) entire text	1-5
P, A	CN 113297412 A (BEIJING DAJIA INTERNET INFORMATION TECHNOLOGY CO., LTD.) 24 August 2021 (2021-08-24) entire text	1-5
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 06 June 2022		Date of mailing of the international search report 14 June 2022
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/JP2022/012222

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2004-510176	A	02 April 2004	US 2003/0055516 A1 paragraphs [0013]-[0057] WO 2002/001439 A2 EP 1297471 A1	
WO	2021/005984	A1	14 January 2021	(Family: none)	
WO	2021/005985	A1	14 January 2021	(Family: none)	
US	2015/0220633	A1	06 August 2015	(Family: none)	
CN	113297412	A	24 August 2021	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G10L 25/27(2013.01)i; G10K 15/04(2006.01)i FI: G10L25/27; G10K15/04 302D		
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）） G10L25/00-25/93; G10K15/04 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 1971-2022年 日本国実用新案登録公報 1996-2022年 日本国登録実用新案公報 1994-2022年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2004-510176 A (ミュージックゲノム、コム インコーポレイテッド) 02.04.2004 (2004-04-02) 段落[0005]-[0026]	1-5
A	WO 2021/005984 A1 (株式会社NTTドコモ) 14.01.2021 (2021-01-14) 全文	1-5
A	WO 2021/005985 A1 (株式会社NTTドコモ) 14.01.2021 (2021-01-14) 全文	1-5
A	US 2015/0220633 A1 (APERTURE INVESTMENTS, LLC) 06.08.2015 (2015-08-06) 全文	1-5
P, A	CN 113297412 A (BEIJING DAJIA INTERNET INFORMATION TECHNOLOGY CO., LTD.) 24.08.2021 (2021-08-24) 全文	1-5
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> 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、一般的な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 06.06.2022	国際調査報告の発送日 14.06.2022	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 山下 剛史 5Z 8946 電話番号 03-3581-1101 内線 3591	

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2022/012222

引用文献			公表日	パテントファミリー文献		公表日
JP	2004-510176	A	02.04.2004	US	2003/0055516	A1
					段落[0013]-[0057]	
				WO	2002/001439	A2
				EP	1297471	A1
WO	2021/005984	A1	14.01.2021	(ファミリーなし)		
WO	2021/005985	A1	14.01.2021	(ファミリーなし)		
US	2015/0220633	A1	06.08.2015	(ファミリーなし)		
CN	113297412	A	24.08.2021	(ファミリーなし)		