

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2018-190371

(P2018-190371A)

(43) 公開日 平成30年11月29日(2018.11.29)

(51) Int.Cl.			F I			テーマコード (参考)	
G06T	7/00	(2017.01)	G06T	7/00	350C	5L096	
G06F	17/30	(2006.01)	G06F	17/30	320Z		
G06N	3/04	(2006.01)	G06F	17/30	220Z		
			G06N	3/04			

審査請求 未請求 請求項の数 7 O L (全 15 頁)

(21) 出願番号	特願2017-182215 (P2017-182215)	(71) 出願人	504133110 国立大学法人電気通信大学 東京都調布市調布ヶ丘一丁目5番地1
(22) 出願日	平成29年9月22日 (2017.9.22)	(74) 代理人	100107766 弁理士 伊東 忠重
(31) 優先権主張番号	特願2017-100317 (P2017-100317)	(74) 代理人	100070150 弁理士 伊東 忠彦
(32) 優先日	平成29年4月28日 (2017.4.28)	(72) 発明者	坂本 真樹 東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
(33) 優先権主張国	日本国 (JP)	(72) 発明者	川嶋 卓也 東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内

最終頁に続く

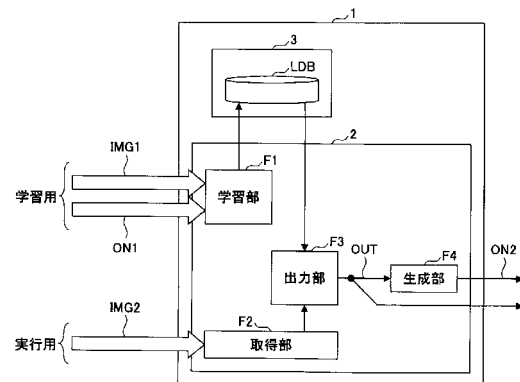
(54) 【発明の名称】 情報処理システム、情報処理方法及びプログラム

(57) 【要約】

【課題】 画像から、質感を精度良く表現する。

【解決手段】 情報処理システムが、第1画像に対する質感を音象徴語で被験者が回答した結果を入力して畳み込みニューラルネットワークによって学習を行う学習部と、第2画像を取得する取得部と、前記学習部による学習結果に基づいて、前記第2画像の質感を示す音韻の要素の出現確率を出力する出力部とを含む。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

第 1 画像に対する質感を音象徴語で被験者が回答した結果を入力して畳み込みニューラルネットワークによって学習を行う学習部と、

第 2 画像を取得する取得部と、

前記学習部による学習結果に基づいて、前記第 2 画像の質感を示す音韻の要素の出現確率を出力する出力部と

を含む情報処理システム。

【請求項 2】

前記音韻の要素を組み合わせて、音象徴語を生成する生成部を更に含む請求項 1 に記載の情報処理システム。

10

【請求項 3】

前記音韻の要素は、母音、子音、反復及び特殊音である請求項 1 又は 2 に記載の情報処理システム。

【請求項 4】

前記第 1 画像は、FMD 画像及び前記 FMD 画像の一部を切り出した部分画像を含む請求項 1 乃至 3 のいずれか 1 項に記載の情報処理システム。

【請求項 5】

前記学習部は、深層畳み込みニューラルネットワークによって学習を行う請求項 1 乃至 4 のいずれか 1 項に記載の情報処理システム。

20

【請求項 6】

情報処理システムが行う情報処理方法であって、

情報処理システムが、第 1 画像に対する質感を音象徴語で被験者が回答した結果を入力して畳み込みニューラルネットワークによって学習を行う学習手順と、

情報処理システムが、第 2 画像を取得する取得手順と、

情報処理システムが、前記学習手順による学習結果に基づいて、前記第 2 画像の質感を示す音韻の要素の出現確率を出力する出力手順と

を含む情報処理方法。

【請求項 7】

コンピュータに情報処理方法を実行させるためのプログラムであって、

30

コンピュータが、第 1 画像に対する質感を音象徴語で被験者が回答した結果を入力して畳み込みニューラルネットワークによって学習を行う学習手順と、

コンピュータが、第 2 画像を取得する取得手順と、

コンピュータが、前記学習手順による学習結果に基づいて、前記第 2 画像の質感を示す音韻の要素の出現確率を出力する出力手順と

を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理システム、情報処理方法及びプログラムに関する。

40

【背景技術】

【0002】

従来、音象徴語と、物理特徴値とを対応付けることで、入力される音象徴語を摩擦係数等の物理特徴値にしたり、入力される物理特徴値を最適な音象徴語にしたり、入力される印象値を物理特徴値にしたり、又は、入力される物理特徴値を物理特徴値にしたりする方法が知られている（例えば、特許文献 1 等）。

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】国際公開第 2017 043424 号

50

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、従来の技術は、画像を入力し、入力される画像から、質感を精度良く表現できない場合が多い。

【0005】

本発明の1つの側面は、このような問題に鑑みてなされたものであり、画像から、質感を精度良く表現することを目的とする。

【課題を解決するための手段】

【0006】

上記の課題を解決するため、情報処理システムは、
第1画像に対する質感を音象徴語で被験者が回答した結果を入力して畳み込みニューラルネットワークによって学習を行う学習部と、
第2画像を取得する取得部と、
前記学習部による学習結果に基づいて、前記第2画像の質感を示す音韻の要素の出現確率を出力する出力部と
を含む。

10

【発明の効果】

【0007】

画像から、質感を精度良く表現できる。

20

【図面の簡単な説明】

【0008】

【図1】情報処理システムの全体構成例を示す図である。

【図2】情報処理装置のハードウェア構成例を示すブロック図である。

【図3】学習処理例を示すフローチャートである。

【図4】学習処理の処理結果例を示す図である。

【図5】音韻の要素を説明する図である。

【図6】実行処理例を示すフローチャートである。

【図7】実行処理の出力結果例を示す図である。

【図8】音象徴語の生成例を示す図である。

30

【図9】音韻の要素の組み合わせ例を示す図である。

【図10】精度評価に用いた数式を示す図である。

【発明を実施するための形態】

【0009】

以下、本発明の好適な実施形態の具体例を示して説明する。

【0010】

<全体構成例>

図1は、情報処理システムの全体構成例を示す概要図である。図示するように、情報処理システム1は、1以上の情報処理装置を有する。具体的には、情報処理システム1は、情報処理装置の例であるサーバ2と、情報処理装置の例であるPC(Personal Computer)3とを有する。

40

【0011】

図示するように、サーバ2と、PC3とは、ネットワーク等によって接続され、相互にデータを送受信することができる。

【0012】

なお、サーバ2と、PC3とは、一体となる構成、すなわち、情報処理システム1は、1つの情報処理装置で構成されてもよい。一方で、情報処理システム1は、3以上の情報処理装置を有してもよい。以下、図示する全体構成を例に説明する。

【0013】

<機能構成例>

50

情報処理システム 1 は、画像が示す被写体の質感を音韻の要素に基づいて表現する。

【 0 0 1 4 】

質感は、人が物体を見て知覚又は推定する物体の特徴である。具体的には、質感は、物体の物性、材質、状態、美感又は組み合わせ等である。物性は、例えば、物体の光沢感又は透明感等である。また、材質は、例えば、物体が陶器又は金属等であるか否か等である。状態は、例えば、物体が乾燥又は凍結しているか否か等である。なお、質感は、物体の物性、材質、状態又は美感等に限られず、他の性質であってもよい。ゆえに、情報処理システム 1 は、画像が示す被写体の「柔らかさ」等の質感を音韻の要素に基づいて表現する。

【 0 0 1 5 】

また、情報処理システム 1 は、実行処理の準備となる学習用の機能構成と、学習後、実行処理を実行するための実行用の機能構成とを含む機能構成である。

【 0 0 1 6 】

以下、学習用に用いられる画像を「第 1 画像 I M G 1」という。一方で、実行用、すなわち、質感を情報処理システム 1 に表現させる対象となる被写体が写った画像を「第 2 画像 I M G 2」という。第 1 画像 I M G 1 及び第 2 画像 I M G 2 の詳細は、後述する。

【 0 0 1 7 】

また、以下の説明では、音象徴語の例としてオノマトペを例に説明するが、音象徴語は、オノマトペに限られず、他の種類であってもよい。

【 0 0 1 8 】

例えば、情報処理システム 1 は、図示するように、学習部 F 1 と、取得部 F 2 と、出力部 F 3 とを含む機能構成である。なお、情報処理システム 1 は、図示するように、生成部 F 4 を更に含む機能構成であるのが望ましいが、生成部 F 4 は、必須構成ではない。以下、図示する機能構成を例に説明する。

【 0 0 1 9 】

学習部 F 1 は、まず、学習処理において、第 1 画像 I M G 1 を入力する。そして、学習部 F 1 は、第 1 画像 I M G 1 に対する質感を音象徴語で被験者が回答した結果を入力して畳み込みニューラルネットワークによって学習を行う学習手順を行う。例えば、学習部 F 1 は、後述する C P U 2 H 0 1 等によって実現される。

【 0 0 2 0 】

取得部 F 2 は、実行処理において、第 2 画像 I M G 2 を取得する取得手順を行う。例えば、取得部 F 2 は、後述する入力装置 2 H 0 3 等によって実現される。

【 0 0 2 1 】

出力部 F 3 は、学習部 F 1 による学習処理によって得られる学習結果に基づいて、第 2 画像 I M G 2 の質感を示す音韻の要素の出現確率 O U T を出力する出力手順を行う。例えば、出力部 F 3 は、後述する出力装置 2 H 0 4 等によって実現される。

【 0 0 2 2 】

生成部 F 4 は、出力部 F 3 が出力する音韻の要素の出現確率 O U T に基づいて、音韻の要素を組み合わせて、オノマトペ O N 2 等の音象徴語を生成する生成手順を行う。例えば、生成部 F 4 は、後述する C P U 2 H 0 1 等によって実現される。

【 0 0 2 3 】

図示する機能構成のうち、学習部 F 1 が、学習処理、いわゆる学習フェーズを行うための機能構成である。一方で、図示する機能構成のうち、取得部 F 2 と、出力部 F 3 と、生成部 F 4 とが、実行処理、いわゆる実行フェーズを行うための機能構成である。

【 0 0 2 4 】

< ハードウェア構成例 >

図 2 は、情報処理装置のハードウェア構成例を示すブロック図である。例えば、サーバ 2 及び P C 3 は、同一のハードウェア構成である。以下、サーバ 2 を例に説明し、P C 3 の説明を省略する。

【 0 0 2 5 】

10

20

30

40

50

サーバ2は、CPU (Central Processing Unit) 2H01と、記憶装置2H02と、入力装置2H03と、出力装置2H04と、インタフェース (interface) 2H05とを有する。各ハードウェア資源は、バス (bus) で接続され、相互にデータを送受信する。

【0026】

CPU 2H01は、演算装置及び制御装置の例である。記憶装置2H02は、主記憶装置等である。また、記憶装置2H02は、SSD (Solid State Drive) 又はハードディスク等の補助記憶装置があってもよい。入力装置2H03は、ユーザから操作を入力する装置である。例えば、入力装置2H03は、キーボード、マウス又はこれらの組み合わせ等である。出力装置2H04は、ユーザに処理結果等を表示する装置である。例えば、出力装置2H04は、ディスプレイ等である。インタフェース2H05は、外部装置とデータを入出力する装置である。例えば、インタフェース2H05は、コネクタ又は通信装置等である。

10

【0027】

なお、ハードウェア構成は、図示する構成に限られない。例えば、ハードウェア構成は、演算装置又は制御装置等が内部又は外部に更にある構成等でもよい。

【0028】

< 学習処理例 >

まず、情報処理システム1は、学習処理によって学習データを生成し、データベースLD Bを構築する。そして、データベースLD Bが構築された後、情報処理システム1は、データベースLD Bを用いて実行処理を行う。例えば、以下のような学習処理が情報処理システム1によって行われる。

20

【0029】

図3は、学習処理例を示すフローチャートである。

【0030】

ステップS01では、情報処理システム1は、第1画像IMG1を表示する。例えば、第1画像IMG1は、FMD (Flickr Material Database) 画像等である。FMD画像は、幅広い素材の画像が集まっており、様々な照明条件、色、テクスチャ、表面形状及び材料のサブタイプ等の特徴とする画像である。そのため、FMD画像が学習処理に用いられると、情報処理システム1は、質感を精度良く学習できる。

30

【0031】

なお、第1画像IMG1は、FMD画像と、FMD画像の一部を切り出した画像 (以下「部分画像」という。) とを含むのが望ましい。具体的には、FMD画像を1000枚とし、さらに、部分画像を1946枚とする。このように、部分画像を用いると、学習処理用の画像となる第1画像IMG1の枚数を増やすことができる。そのため、情報処理システム1は、学習データを高精度化することができる。

【0032】

さらに、部分画像は、FMD画像の一部であるため、被写体の一部しか写っていない場合がある。このような場合では、学習処理において、回答する被験者は、被写体の種類等が何であるか分かりにくい場合が多い。そのため、被写体の種類等による被験者の先入観を少なくできる。ゆえに、部分画像を用いると、情報処理システム1は、先入観による被写体全体の種類が回答結果に与える影響を少なくできる。

40

【0033】

ステップS02では、情報処理システム1は、第1画像IMG1の質感を示す音象徴語の回答結果を入力し、畳み込みニューラルネットワーク (CNN、Convolutional Neural Network) によって学習を行う。

【0034】

なお、学習処理に用いられる畳み込みニューラルネットワークは、深層畳み込みニューラルネットワーク (DCNN、Deep Convolutional Neural Network) であるのが望ましい。具体的には、以下に説明する例では、深層畳み込

50

みニューラルネットワークは、VGG (Visual Geometry Group) 16モデルにおける「f8」の出力層を音象徴語の出力形式とする例である。より具体的には、深層畳み込みニューラルネットワークは、2モーラ (mora) に該当する73次元を出力層とする設定である。

【0035】

以下、「モーラ」は、日本語リズムにおける拍数を示す単位とする。例えば、「ズキッ」という音象徴語は、「ズ」の部分が1モーラ目となり、「キッ」の部分が2モーラ目となる2モーラのおノマトペである。

【0036】

また、この例では、深層畳み込みニューラルネットワークは、ファインチューニングを行い、ISLVR (ImageNet Large Scale Visual Recognition Challenge) 2015のVGG16のデータを適用した設定とする。さらに、この例では、ドロップアウト (dropout) 等を設定し、過学習を抑制する設定とする。

10

【0037】

なお、設定は、上記の設定に限られない。

【0038】

畳み込みニューラルネットワークを用いると、SIFT (Scale Invariant Feature Transform) 等を用いる場合と比較して、情報処理システム1は、学習データを高精度化することができる。

20

【0039】

具体的には、学習は、表示される第1画像IMG1を被験者が見て、被験者が質感をおノマトペON1で回答する形式で行われる。したがって、学習処理が行われると、第1画像IMG1をいわゆるトレーニング画像とした被験者のおノマトペON1による回答結果がいわゆる正解データとなり、学習結果が蓄積される。

【0040】

具体的には、学習処理によって、学習済みモデルには、各層に重みが行列の形式で保存される。例えば、以下のような(1)式の関係があるとする。

$$y = a \cdot x + b \quad (1)$$

30

上記(1)式では、「x」が第1画像IMG1となる。そして、上記(1)式では、「a」が重みのパラメータ、すなわち、行列となる。さらに、上記(1)式では、「b」がバイアスのパラメータであり、「y」が出力値となる。上記(1)式が用いられる場合は、例えば、「a」と、「b」とが、学習結果として、各層に保存される。

【0041】

なお、図示する処理は、第1画像IMG1の枚数分繰り返し行われる。

【0042】

<学習処理の処理結果例>

図4は、学習処理の処理結果例を示す図である。例えば、学習処理が行われると、図示するような学習データが得られる。

40

【0043】

図示するように、学習処理で回答されたおノマトペON1は、音韻に分解されて保存される。この例は、第1画像IMG1が1946枚であり、「さーさー」及び「ふわふわ」といった1モーラ又は2モーラの29443語のおノマトペが回答の選択肢となる例である。以下、おノマトペのモーラ数を1モーラ又は2モーラとする例で説明するが、出力層の出力次元を増やすことで3モーラ以上のおノマトペを用いてもよい。

【0044】

学習処理で回答されたおノマトペON1は、図示する例では、「0」乃至「72」の73次元に分解される。図では、縦軸は、学習処理で回答されたおノマトペON1とし、横

50

軸は、オノマトペON1が音韻の要素を持つか否か解析した結果である。なお、図示する例では、解析結果は、所定の音韻の要素をオノマトペON1が有する場合を「1」とし、所定の音韻の要素をオノマトペON1が有しない場合を「0」とする。

【0045】

図示する例では、73次元の音韻の要素は、以下のような要素である。

【0046】

図5は、音韻の要素を説明する図である。図における「番号」が図4に示す横軸、すなわち、73次元に対応する。そして、「要素」が音韻の要素の内容を示し、「()」内の数値が「1モーラ目」か「2モーラ目」かを示す。

【0047】

まず、「0」及び「1」の番号は、反復の有無を示す。すなわち、「ざらざら」のように、1モーラ目と、2モーラ目とが同じ音となり、反復がある場合には、「0」番目の「Re」の音韻の要素があると判断される。一方で、反復がない場合には、「1」番目の「noRe」の音韻の要素がないと判断される。

【0048】

また、「2」乃至「36」の要素が1モーラ目であり、「37」乃至「72」の要素が2モーラ目である。まず、これらの要素のうち、「a」、「i」、「u」、「e」及び「o」は、オノマトペON1が母音の要素を持つか否かを解析した結果である。すなわち、1モーラ目において、「a」、「i」、「u」、「e」又は「o」の母音がある場合には、「2」乃至「6」のいずれかの音韻の要素があると判断され、2モーラ目において、「a」、「i」、「u」、「e」又は「o」の母音がある場合には、「37」乃至「41」のいずれかの音韻の要素があると判断される。一方で、1モーラ目において、母音がない場合には、「7」の音韻の要素があると判断され、2モーラ目において、母音がない場合には、「42」の音韻の要素があると判断される。

【0049】

さらに、「k」等は、オノマトペON1が子音の要素を持つか否かを解析した結果である。すなわち、1モーラ目において、子音がある場合には、「8」乃至「33」のいずれかの音韻の要素があると判断される。また、2モーラ目において、子音がある場合には、「43」乃至「68」のいずれかの音韻の要素があると判断される。

【0050】

さらにまた、「N」、「Q」、「R」及び「L」は、特殊音の要素を持つか否かを解析した結果である。

【0051】

特殊音は、詰まり音又は伸ばし音等のように、各モーラに含まれる要素となる音韻である。具体的には、「N」は、「ん」の音を示す。そして、「Q」は、「っ」で記載される音、すなわち、促音を示す。「R」は、「ー」で記載される伸ばし音、すなわち、長音を示す。「L」は、「り」の音を示す。なお、これらの特殊音の要素までが1モーラと判断される。したがって、例えば、「が」、「がん」又は「がっ」といった音は、1モーラと判断される。

【0052】

以上のように、情報処理システム1は、学習処理で被験者が回答するオノマトペON1をモーラに分解して、反復、母音、子音及び特殊音の音韻の要素を持つか否かを判断して記憶する。

【0053】

< 実行処理例 >

上記のような学習処理によって、データベースLDBが構築された後、すなわち、いわゆる学習フェーズが完了した後、以下のような実行処理が行われる。

【0054】

図6は、実行処理例を示すフローチャートである。

【0055】

10

20

30

40

50

ステップ S 1 1 では、情報処理システム 1 は、第 2 画像 I M G 2 を取得する。つまり、ステップ S 1 1 で取得される第 2 画像 I M G 2 が写す被写体の質感が、実行処理によって評価され、出力される。

【 0 0 5 6 】

ステップ S 1 2 では、情報処理システム 1 は、学習結果に基づいて、第 2 画像 I M G 2 の質感を音韻の要素の出現確率で表示する。ステップ S 1 2 の詳細は、後述する。

【 0 0 5 7 】

なお、ステップ S 1 2 によって出力される音韻の要素の出現確率に基づいて、以下のような処理が更に行われるのが望ましい。

【 0 0 5 8 】

ステップ S 1 3 では、情報処理システム 1 は、音韻の要素の出現確率に基づいて、音象徴語を生成し、出力する。ステップ S 1 3 の詳細は、後述する。

【 0 0 5 9 】

< 音韻の要素の出現確率の表示例 >

上記 (1) 式のように学習が行われる場合には、学習済みモデルでは、出力関数をシグモイド (s i g m o i d) 関数とする。そのため、この例では、第 2 画像 I M G 2 における音韻の要素が、図 4 に示す「出現確率」のように、確率で出力される。

【 0 0 6 0 】

そして、実行処理において、テスト画像となる第 2 画像 I M G 2 が取得されると、情報処理システム 1 は、あらかじめ第 1 画像 I M G 1 によって学習された学習済みモデルにおける各層に保存される重みと、バイアスとに基づいて第 2 画像 I M G 2 を処理する。このような処理によって、例えば、以下のような音韻の要素の出現確率 O U T が表示できる。

【 0 0 6 1 】

図 7 は、実行処理の出力結果例を示す図である。図は、音韻の要素の出現確率 O U T を「反復」、「1 モーラ目」及び「2 モーラ目」に分けて表示する例である。また、図は、音韻の要素を出現確率の高い順に並べて示す。したがって、図において、数値は、出現確率の順位を示し、「1」で示す要素が各構成において、最も出現確率が高い要素である。

【 0 0 6 2 】

具体的には、図示する例では、「反復」は、「R e」が「0.96」、すなわち、「96%」の出現確率である。なお、「反復」における「R e」は、図 5 における「1」番目の要素に相当する。

【 0 0 6 3 】

また、図示する例では、「1 モーラ目」における「母音」の要素では、「o」が「0.61」と最も高い出現確率である。なお、「1 モーラ目」における「o」は、図 5 における「6」番目の要素に相当する。

【 0 0 6 4 】

同様に、「1 モーラ目」における「子音」の要素では、「g」が「0.34」と最も高い出現確率である。なお、「1 モーラ目」における「g」は、図 5 における「10」番目の要素に相当する。

【 0 0 6 5 】

さらに、「特殊音」の要素では、「N」が「0.02」、「Q」が「0.01」、「R」が「0.01」の出現確率である。なお、「1 モーラ目」における「特殊音」は、図 5 における「34」乃至「36」番目の要素に相当する。

【 0 0 6 6 】

「1 モーラ目」と同様に、「2 モーラ目」の出現確率も表示される。

【 0 0 6 7 】

情報処理システム 1 は、以上のように各要素を出現確率で示して、第 2 画像 I M G 2 の質感を出力する。質感は、人によって異なる表現となり、表現に曖昧さがある場合が多い。例えば、同じ「柔らかい」という質感を表現する場合でも、形容詞を用いると、人によって、「とても柔らかい」と表現されたり、「少し柔らかい」と表現されたりするため、

10

20

30

40

50

同一の画像であっても、表現する人によって差が生じる場合がある。これに対して、上記のように、音韻の要素を出現確率で示すと、情報処理システム1は、人による曖昧さを表現することができる。

【0068】

<音象徴語の生成例>

以上のような音韻の要素の出現確率OUTがステップS12で出力されると、例えば、ステップS13では、情報処理システム1は、以下のように音象徴語を生成できる。

【0069】

図8は、音象徴語の生成例を示す図である。以下、図7に示す音韻の要素の出現確率OUTが出力される場合を例に説明する。また、この例は、音韻の要素の出現確率OUTにおける「反復」、「1モーラ目」及び「2モーラ目」の要素のうち、最も出現確率が高いそれぞれの要素を組み合わせ、オノマトペを生成する例である。

10

【0070】

まず、情報処理システム1によるオノマトペの「1モーラ目」（以下「第1モーラ成分MR1」という。）の生成例を説明する。図示するように、音韻の要素の出現確率OUTでは、「母音」は、「o」の要素が最も出現確率が高く、「子音」は、「g」の要素が最も出現確率が高い要素である。

【0071】

なお、この例は、あらかじめ設定する設定値以下の出現確率である場合には、「特殊音」を使用しないとする例である。つまり、「特殊音」の要素がいずれも、低い出現確率である場合には、「特殊音」の要素を含まないオノマトペを生成する設定であるとする。

20

【0072】

したがって、第1モーラ成分MR1は、「g」の「子音」と、「o」の「母音」と、「特殊音」を「なし」とする構成で生成され、「go」（「ご」という音である。）となる。なお、「母音」と、「子音」との組み合わせによって生成される音の詳細は、後述する。

【0073】

次に、情報処理システム1によるオノマトペの「2モーラ目」（以下「第2モーラ成分MR2」という。）の生成例を説明する。図示するように、音韻の要素の出現確率OUTでは、「母音」は、「a」の要素が最も出現確率が高く、「子音」は、「w」の要素が最も出現確率が高い要素である。また、「特殊音」の要素の出現確率は、いずれも設定値以下の低い出現確率であるとする。

30

【0074】

したがって、第2モーラ成分MR2は、「w」の「子音」と、「a」の「母音」と、「特殊音」を「なし」とする構成で生成され、「wa」（「わ」という音である。）となる。

【0075】

以上のように生成される「go」の第1モーラ成分MR1と、「wa」の第2モーラ成分MR2とを組み合わせると、図示するように、「gow a」という音が生成できる。

【0076】

続いて、「反復」の要素を反映させる例を説明する。図示する例では、「反復」において、「Re」が「noRe」より高い出現確率である。したがって、この例では、「反復」を「あり」とする例である。具体的には、情報処理システム1は、第1モーラ成分MR1及び第2モーラ成分MR2の組み合わせによって生成される「gow a」の音を繰り返すようにして「gow a g o w a」とする。

40

【0077】

以上のようにすると、情報処理システム1は、「ごわごわ」という第2画像IMG2の質感を表現するオノマトペON2を生成することができる。音象徴語では、音韻の要素が示すそれぞれの1音1音に印象が結びつくため、共通した印象、いわゆる「音象徴性」が強く現れる傾向がある。具体的には、「h」の音韻の要素は、「柔らかさ」の質感に結び

50

つきやすく、「s」の音韻の要素は、「滑らかさ」の質感に結びつきやすい。ほかにも、「g」の音韻の要素は、「硬い表面との接触」の質感に結びつきやすい。

【0078】

したがって、上記のように生成されるオノマトペON2で質感を表現すると、例えば、形容詞で質感を示す場合と比較して、多様な表現ができる場合が多い。形容詞で表現する場合には、表現に用いる形容詞をあらかじめ用意する必要がある。そのため、表現は、用意された形容詞を用いた範囲内に限定される。すなわち、あらかじめ用意した形容詞の種類が少ないと、表現できない質感が多くなる。もし、形容詞で多様な質感を表現する場合には、多数の種類形容詞をあらかじめ用意する必要がある。

【0079】

一方で、本発明の一実施形態のように、オノマトペON2で表現する場合には、音韻の要素を組み合わせてオノマトペON2を生成するため、多様な種類のオノマトペON2を生成することが可能である。ゆえに、情報処理システム1は、オノマトペON2によって、多様な表現で複雑な質感を統合的に表現できる。

【0080】

また、音象徴語による表現は、日本語以外の言語であっても通用する可能性が高い。すなわち、日本語を母国語としない人、いわゆる外国人であっても、質感を音象徴語で表現しようとする、同じような音象徴語を用いようとする傾向がある。特に、「1モーラ目」の「子音」は、共通する可能性が高い。ゆえに、音象徴語による表現であると、外国人にも通用し、質感を精度良く表現できる場合が多い。

【0081】

<音韻の要素の組み合わせ例>

図9は、音韻の要素の組み合わせ例を示す図である。図示する例では、縦軸と、横軸の一部とが「子音」の要素であり、横軸が「母音」の要素である。

【0082】

例えば、「子音」が「k」であって、「母音」が「a」であると、「母音」と、「子音」との組み合わせは、「ka」となり、図示するように、「か」という音となる。

【0083】

このように、ステップS13では、情報処理システム1は、音韻の要素の出現確率OUTから、図における縦軸に示す音韻の要素と、横軸に示す音韻の要素とを組み合わせ音を生成し、音象徴語を生成する。

【0084】

<学習と類似度の精度評価実験結果>

比較対象として、被験者に、第2画像IMG2となる「The KTH-TIPS and KTH-TIPS2 image databases」(<http://www.nada.kth.se/cvap/databases/kth-tips/>)にある100枚の画像を表示して、被験者が想起するオノマトペを回答してもらった。

【0085】

実験では、上記の比較対象と、図3に示す学習処理及び図6に示す実行処理とで得られるそれぞれの音韻の要素の出現確率を以下に示す数式によって、KL情報量(カルバック・ライブラー情報量、Kullback-Leibler divergence)と、cos類似度とを計算し、評価した。

【0086】

図10は、精度評価に用いた数式を示す図である。図示する(2)式は、KL情報量を計算するための数式である。一方で、(3)式は、cos類似度を計算するための数式である。

【0087】

図示する(2)式を用いると、2つの確率分布の差異が計算できる。なお、(2)式による計算結果は、数値が「0」に近づくほど類似度が高いことを示す。また、(2)式は、計算対象となる離散確率分布を「P」及び「Q」で示し、「P(i)」及び「Q(i)

10

20

30

40

50

)」は、「i」の場合の確率を示す。

【0088】

図示する(3)式を用いると、2つのベクトルの類似度が計算できる。なお、(3)式による計算結果は、数値が「1」に近づくほど類似度が高いことを示す。また、(3)式では、計算対象となるベクトルを「p」及び「q」で示す。

【0089】

実験では、(2)式に基づいて、KL情報量が「2.89」程度となり、cos類似度が「0.907」程度となった。以上のような実験結果から、情報処理システム1が、出現確率を用いて、画像の質感に適した音韻の要素を高精度に出力できることが確認された。

10

【0090】

<他の実施形態>

本発明に係る実施形態は、上記の情報処理方法によって生成される音象徴語を用いて、質感を発話するロボット等に適用されてもよい。

【0091】

なお、全体処理は、図示する順序に限られない。例えば、各処理の一部又は全部は、並列又は図示する順序とは異なる順序で行われてもよい。また、各処理は、複数の情報処理装置によって、冗長、分散、並列、仮想化又はこれらを組み合わせて実行されてもよい。

【0092】

なお、本発明に係る情報処理方法は、プログラムによって実現されてもよい。すなわち、プログラムは、1以上の情報処理装置を含む情報処理システム等であるコンピュータに各処理を実行させるためのコンピュータプログラムである。

20

【0093】

したがって、プログラムに基づいて情報処理方法が実行されると、コンピュータが有する演算装置及び制御装置は、各処理を実行するため、プログラムに基づいて演算及び制御を行う。また、コンピュータが有する記憶装置は、各処理を実行するため、プログラムに基づいて、処理に用いられるデータを記憶する。

【0094】

また、プログラムは、コンピュータが読み取り可能な記録媒体に記録されて頒布することができる。なお、記録媒体は、磁気テープ、フラッシュメモリ、光ディスク、光磁気ディスク又は磁気ディスク等のメディアである。また、記録媒体は、補助記憶装置等でもよい。さらに、プログラムは、電気通信回線を通じて頒布することができる。

30

【0095】

以上、本発明の好ましい実施形態の具体例について詳述したが、本発明は、図面に記載された実施形態等に限定されない。すなわち、特許請求の範囲に記載された本発明の広範な要旨の範囲内において、種々の変形又は変更が可能である。

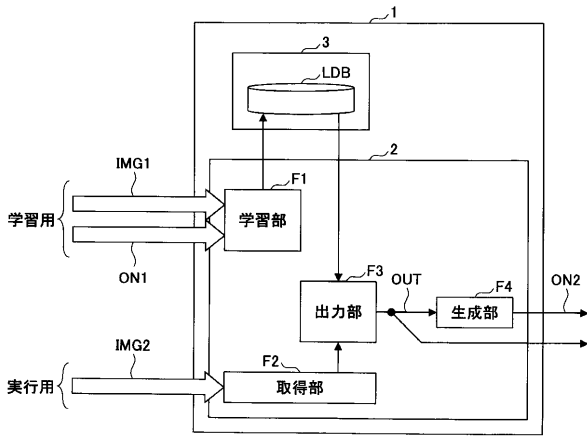
【符号の説明】

【0096】

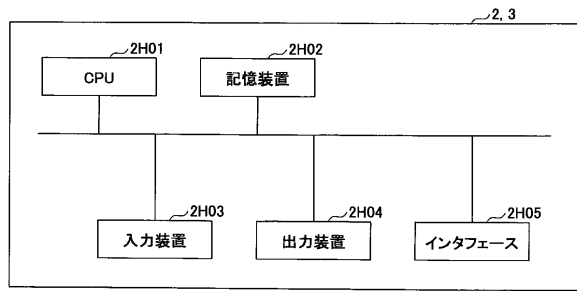
1	情報処理システム
ON 1、ON 2	オノマトペ
OUT	音韻の要素の出現確率
F 1	学習部
F 2	取得部
F 3	出力部
F 4	生成部
IMG 1	第1画像
IMG 2	第2画像

40

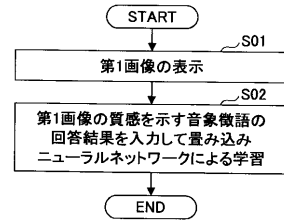
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 】

音韻の要素

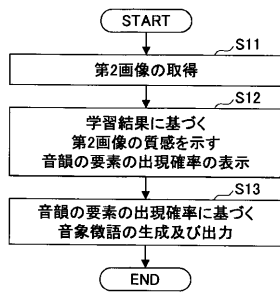
番号(要素)	0(Re)	1(noRe)	2(a)	3(i)	4(u)	5(e)	6(o)	7(emp)	8(k)	9(ky)	10(φ)	72(L)
こわこわ	1	0	0	0	0	0	1	0	0	0	1	0
こわこわ	1	0	0	0	0	0	1	0	0	0	1	0
ざらざら	1	0	1	0	0	0	0	0	0	0	0	0
ちくちく	1	0	0	1	0	0	0	0	0	0	0	0
...
出現確率	1	0	0.25	0.06	0	0	0.68	0	0	0	0.50	0

ON1

【 図 5 】

番号	要素	番号	要素	番号	要素
0	Re	25	by(1)	50	zy(2)
1	noRe	26	p(1)	51	t(2)
2	a(1)	27	py(1)	52	ty(2)
3	i(1)	28	m(1)	53	d(2)
4	u(1)	29	my(1)	54	dy(2)
5	e(1)	30	y(1)	55	n(2)
6	o(1)	31	r(1)	56	ny(2)
7	emp(1)	32	ry(1)	57	h(2)
8	k(1)	33	w(1)	58	hy(2)
9	ky(1)	34	N(1)	59	b(2)
10	g(1)	35	Q(1)	60	by(2)
11	gy(1)	36	R(1)	61	p(2)
12	s(1)	37	a(2)	62	py(2)
13	sy(1)	38	i(2)	63	m(2)
14	z(1)	39	u(2)	64	my(2)
15	zy(1)	40	e(2)	65	y(2)
16	t(1)	41	o(2)	66	r(2)
17	ty(1)	42	emp(2)	67	ry(2)
18	d(1)	43	k(2)	68	w(2)
19	dy(1)	44	ky(2)	69	N(2)
20	n(1)	45	g(2)	70	Q(2)
21	ny(1)	46	gy(2)	71	R(2)
22	h(1)	47	s(2)	72	L(2)
23	hy(1)	48	sy(2)		
24	b(1)	49	z(2)		

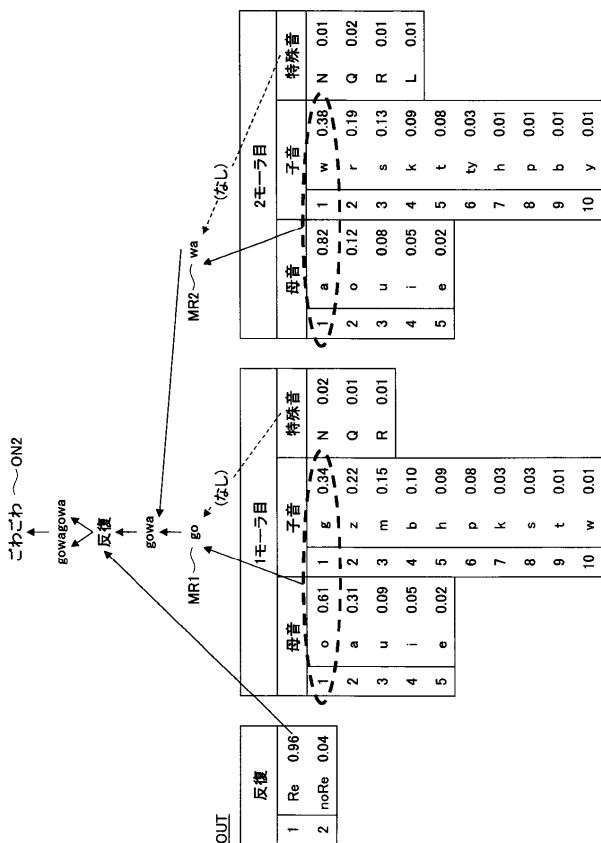
【図6】



【図7】

反復		出現確率	
1	Re	0.96	
2	noRe	0.04	
OUT			
1モーラ目			
母音		子音	
出現確率		出現確率	特殊音
1	o	0.61	N
2	a	0.31	Q
3	u	0.09	R
4	i	0.05	
5	e	0.02	
2モーラ目			
母音		子音	
出現確率		出現確率	特殊音
1	a	0.82	N
2	o	0.12	Q
3	u	0.08	R
4	i	0.05	
5	e	0.02	
6	ty	0.08	
7	h	0.03	
8	p	0.01	
9	b	0.01	
10	y	0.01	

【図8】



【図9】

母音		子音	
emp	a	i	yo
k	か	き	ゆ
g	が	ぎ	よ
s	さ	し	しよ
z	ざ	じ	じよ
t	た	ち	ちよ
d	だ	ぢ	ぢよ
n	な	に	にゆ
h	は	ひ	ひゆ
b	ば	び	びゆ
p	ぱ	ぴ	ぴゆ
y	や	-	-
r	ら	り	りゆ
w	わ	-	-

【 図 1 0 】

$$D_{KL}(P\|Q) = \sum_i P(i) \log_{10} \frac{P(i)}{Q(i)} \quad (2)$$

$$\cos(\vec{p}, \vec{q}) = \frac{\sum_i p_i q_i}{\sqrt{\sum_i p_i} \cdot \sqrt{\sum_i q_i}} \quad (3)$$

フロントページの続き

(72)発明者 権 眞煥

東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内

Fターム(参考) 5L096 BA16 CA25 EA35 GA30 HA11 KA04