

## オノマトペにより歩容を記述するための 音韻空間と人体部位の動きの関係性

加藤 大貴<sup>†</sup> 平山 高嗣<sup>†,†</sup> 川西 康友<sup>†</sup> 道満 恵介<sup>†,†,†</sup> 井手 一郎<sup>†,†</sup>  
出口 大輔<sup>†,†,†,†</sup> 村瀬 洋<sup>†,†</sup>

<sup>†</sup> 名古屋大学大学院情報科学研究科

<sup>††</sup> 名古屋大学大学院実世界データ循環学リーダー人材養成プログラム

〒464-8601 愛知県名古屋市千種区不老町

<sup>†††</sup> 中京大学工学部 〒470-0393 愛知県豊田市貝津町床立101

<sup>††††</sup> 名古屋大学情報戦略室 〒464-8601 愛知県名古屋市千種区不老町

E-mail: <sup>†</sup>katoh@murase.m.is.nagoya-u.ac.jp, <sup>††</sup>{hirayama,kawanishi,ide,murase}@is.nagoya-u.ac.jp

あらまし 日本では、人間の歩行動作は、その見た目に応じて多様なオノマトペで表現される。人間が歩容を見るときには、部位間の相対的な動きからオノマトペを連想する可能性が従来研究で示唆されている。一方、オノマトペには音象徴性があり、音韻と連想するイメージには関係があるとされる。そこで、本報告では音韻と部位間の相対的な動きの関係性を利用し、任意の歩容をオノマトペで記述することを目指す。

キーワード オノマトペ、歩容、音韻、人体部位、音象徴性

## Relationship between Phonetic-Space and Body-Parts Movement for Description of Gaits by Onomatopoeia

Hiroataka KATO<sup>†</sup>, Takatsugu HIRAYAMA<sup>†,†</sup>, Yasutomo KAWANISHI<sup>†</sup>, Keisuke DOMAN<sup>†,†,†</sup>,  
Ichiro IDE<sup>†,†</sup>, Daisuke DEGUCHI<sup>†,†,†,†</sup>, and Hiroshi MURASE<sup>†,†</sup>

<sup>†</sup> Graduate School of Information Science, Nagoya University

<sup>††</sup> Graduate Program for Real-World Data Circulation Leaders, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

<sup>†††</sup> School of Engineering, Chukyo University

101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan

<sup>††††</sup> Information Strategy Office, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

E-mail: <sup>†</sup>katoh@murase.m.is.nagoya-u.ac.jp, <sup>††</sup>{hirayama,kawanishi,ide,murase}@is.nagoya-u.ac.jp

**Abstract** In Japanese, Gaits are expressed by various onomatopoeias according to their appearance. Previous study implied relative body-parts movement imagined onomatopoeias when we look gaits. On the other hand, it is thought that onomatopoeia has sound-symbolism and its phoneme has relation to an impression imagined. Thus, in this report, we aim to describe an ability gait by onomatopoeias using the relationship between phoneme and relative body-parts movement.

**Key words** Onomatopoeia, gait, phoneme, body-parts, sound-symbolism

## 1. はじめに

口語表現では「のろのろ」、「つるつる」、「しゃかしゃか」など、事象の様子を直感的に表現する言葉として、オノマトペが使用される。オノマトペは擬音語及び擬態語と呼ばれている言語表現の総称である [1]。他の言語と比べ、日本語はオノマトペの種類が圧倒的に多く、その使用範囲も多岐にわたり、多用されることが知られている。

オノマトペは、その音響的印象が事象の様態と対応しているため、人間はオノマトペに対して共通のイメージを想起するとされている [2]。そのため、オノマトペは論理的な表現が容易ではない直感的な印象を端的に他者に伝えるための有効な手段であると考えられている。

また、オノマトペは直感的な印象を計算機に伝える手段としても有効であると考えられており、近年オノマトペを直感的なインタフェースの入力手段として利用する研究が盛んになりつつある [3], [4]。神原ら [3] は、「オノマトペン」という描画システムを開発している。これは例えば、「ぎざぎざ」と発話しながら線を描くことで、「ぎざぎざ」な線を描くことができるインタフェースであり、利用者に直感的な操作環境の提供を実現している。また、小松ら [4] は、利用者がオノマトペを入力すると、そのオノマトペに合致したロボットの動作記述作業を自動で行なうシステムを開発し、一般人には敷居が高い作業を直感的に行なえるようにしている。このように、オノマトペをインタフェースの入力手段として用いることの有用性が示されつつある。

一方で、藤野ら [5] が運動感覚の学習などにオノマトペの利用が効果的であると指摘しているように、入力とオノマトペを対応付け、オノマトペを出力するシステムの需要も高まっていると思われる。しかし、これらの対応付けはその多くが人手で行なわれており、対応の獲得にコストを要するという問題点がある。このような対応付けの自動化を実現するためには、計算機が理解できるように入出力メディアとオノマトペとの対応関係を定量的に記述する必要がある。我々は、そのような入出力メディアの中でも映像に注目し、歩容を撮影した映像とオノマトペとの関係性を利用した歩容の分類研究に取り組んできた [6]。人間の歩容は、多様なオノマトペで表現されることが知られており [1]、その動きの差異が、オノマトペの差異と密接に関係していると考えたためである。

関連研究として、鍵谷ら [7] は、液体の粘性に注目し、CG映像作成ソフトウェアを用いて、映像作成時の動粘度パラメータと、作成された映像から想起されるオノマトペを構成する音韻の種類に関連性があることを明らかにしている。また、杉山ら [8] は、犬型ロボットの歩行シミュレータを用いて、被験者にオノマトペを表現したロボットの歩行パターンを設計させる実験を行ない、動きに対応したオノマトペの種類を人間が判別するためには、肩と足、右足と左足など、体の相対的な運動に着目することが重要であると示唆している。

これらをふまえ、我々の先行研究 [6] では、歩容映像が特定のオノマトペに対応するか否かを識別する識別器を学習するこ

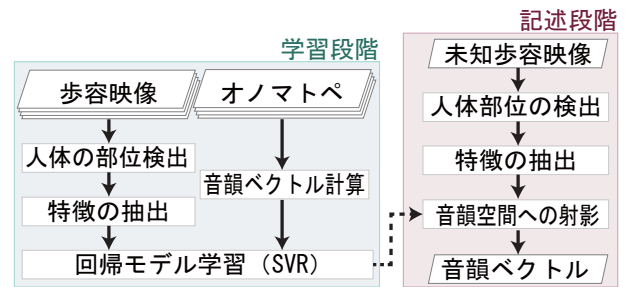
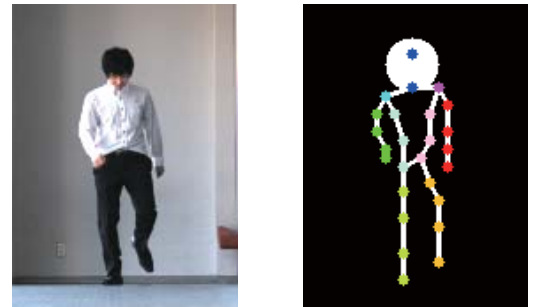


図1 歩容映像の音韻空間への射影手法の処理の流れ



(a) 元の映像フレーム

(b) 部位検出結果

図2 FMPによる部位検出結果の例

とで、歩容映像とオノマトペの間には関係性があり、機械的に識別可能であることを確認した。しかし、先行研究のような手法では、学習に用いなかったオノマトペを識別することは困難である。膨大な種類が存在するオノマトペ全てに対し、このような学習を個別に行なうのは現実的ではないため、より抽象的な対応付けの枠組みが必要とされる。

そのために、本研究ではオノマトペを構成する音素が持つ特徴 [2], [4] に注目する。本報告では、この特徴によって張られる特徴空間を音韻空間と呼ぶ。そして、オノマトペが音象徴性に基づいて音韻空間上で連続的に印象が変化すると仮定し、回帰と射影により音韻空間と映像特徴空間を対応付けることにより、映像・オノマトペ間を対応付けるためのより柔軟な枠組みの確立を目指す。

本報告では、まず2.で歩容映像の音韻空間への射影手法について述べる。次に3.で実験に用いるデータセットの作成方法について述べる。更に4.で、提案する枠組みの妥当性を検証するための実験及び結果について述べる。最後に5.で今後の課題について検討する。

## 2. 歩容映像の音韻空間への射影手法

提案手法の処理の流れを図1に示す。歩容映像に対して行なう特徴抽出までの処理は先行研究 [6] に基づく。以下、各手順について詳述する。

### 2.1 人体部位の検出

本手法では、人体部位の相対的位置関係に基づく特徴を用いる。そのために、事前処理として映像から人体の部位を検出する。ここでは、部位検出処理にFlexible Mixture of Parts (FMP) [9] を利用する。FMPは、人検出と姿勢推定を同時に

行なう手法であり、入力画像に対して、人体の部位 26 か所の位置座標を得られる。入力映像中のあるフレームに対して FMP を実行した結果を図 2 に示す。図 2(a) が元の映像フレームであり、図 2(b) が検出結果をグラフ表現で可視化したものである。図 2(b) 中のノードは、それぞれ検出された部位の位置を示す。

事前処理では、入力された歩容映像の全フレームに対して FMP を適用し、各部位の位置座標系列  $P(p, t)$  を得る。ここで、 $p$  は各部位の識別子であり、 $p \in \{0, \dots, 25\}$  である。 $t$  はフレーム番号であり、映像長を  $T$  とすると、 $t \in \{1, \dots, T\}$  である。

得られた各部位の位置座標系列  $P(p, t)$  にはノイズが多く含まれるため、時間方向に平滑化することでその影響を軽減する。平滑化された各部位の位置座標系列  $P'(p, t)$  は、平滑化窓幅を  $S$  とすると次式で表される。

$$P'(p, t) = \begin{cases} \frac{1}{S} \sum_{\tau=t-S+1}^t P(p, \tau) & (t > S) \\ \frac{1}{t} \sum_{\tau=1}^t P(p, \tau) & (\text{otherwise}) \end{cases} \quad (1)$$

得られた位置座標系列  $P'(p, t)$  から、26 部位のうち全ての 2 部位  $p_1, p_2$  の組み合わせにおける部位の相対距離系列  $D_{p_1, p_2}(t)$  を計算する。相対距離の計算には Euclidean distance を用い、単位は画素 (pixel) とする。

また、各フレームにおける頭の  $y$  座標と足の  $y$  座標の差  $H(t)$  を計算し、映像全体での  $H(t)$  の平均  $\bar{H}$  を求める。そして、すべての  $D_{p_1, p_2}(t)$  を  $\bar{H}$  で除することにより、正規化された部位の相対距離系列  $L_{p_1, p_2}(t)$  を得る。

$$L_{p_1, p_2}(t) = \frac{D_{p_1, p_2}(t)}{\bar{H}} \quad (2)$$

$$\bar{H} = \frac{1}{T} \sum_{t=1}^T H(t) \quad (3)$$

## 2.2 人体部位の相対的位置関係に基づく特徴の抽出

2.1 節で導入した部位の相対距離系列  $L_{p_1, p_2}(t)$  を用いて、部位の相対的位置関係に基づく特徴を抽出する。これにより、大きく動く部位など、局所的な特徴を捉えられると想定する。まず、全ての  $L_{p_1, p_2}(t)$  について、時間方向の分散  $V_{p_1, p_2}$  と尖度  $K_{p_1, p_2}$  を計算する。そして、得られた全ての分散  $V_{p_1, p_2}$  と尖度  $K_{p_1, p_2}$  を連結したものを特徴量とする。 $p_1 < p_2$  を満たす  $p_1$  と  $p_2$  の組み合わせは  ${}_{26}C_2 = 325$  通りである。よって、部位の相対的位置関係に基づく特徴量の次元数は、分散、尖度各々について 325 次元で、合計 650 次元となる。

## 2.3 回帰モデルの構築

本報告では、音韻空間として [10] で提案されている 32 次元のオノマトベクトルを利用する。これは、ABAB 型（「すたすた」、「のろのろ」等同じ 2 音が 2 回繰り返される形）のオノマトベを構成する 4 つの音韻（「すたすた」であれば「s」、「u」、「t」、「a」の 4 つ）それぞれに対し、[4] で提案されている 8 次

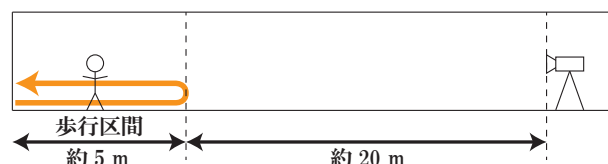


図 3 歩容映像の撮影状況の模式図

元属性ベクトルを割り当てたものである。8 次元属性ベクトルは、日本語の音韻の構成要素であるすべての母音、子音に対して「硬さ」、「強さ」、「湿度」、「滑らかさ」、「丸さ」、「弾性」、「速さ」、「温かさ」の 8 項目の属性で構成され、各属性が持つ値は経験的に定義されている。

2.2 節で得た 650 次元の特徴量を回帰し、32 次元の音韻空間上に射影する。回帰には線形 SVR を用いる。オノマトベとの対応関係が既知の歩容映像から前述の特徴を抽出し、その特徴ベクトルを説明変数、対応するオノマトベから決定される音韻ベクトルを目的変数とする。

## 2.4 音韻空間への歩容特徴の射影

オノマトベとの対応が未知の歩容映像から特徴を抽出し、2.3 節で構築した回帰モデルを用いて音韻空間に射影する。このようにして得られた音韻空間上の射影が、その歩容の音韻的記述であると考えられる。

## 3. データセットの作成

本節では、評価実験で用いるオノマトベがラベルとして付与された歩容映像データセットの作成方法について述べる。まず、3.1 節で歩容映像の撮影方法について述べる。次に、3.2 節で第 3 者による主観評価に基づいて歩容映像にオノマトベを付与する方法について述べる。

### 3.1 撮影方法

歩容映像として、歩行者の前面及び背面を撮影した。奥行き方向の移動による歩行者の大きさの変化を最小限に抑えるために、歩行者から十分離れた位置にカメラを設置した。撮影には Point Gray Research 社のカメラ Flea3<sup>(注1)</sup> を用いた。カメラレンズの焦点距離は 35 mm、センサの大きさは 2/3 inch であり、35 mm 判換算焦点距離は約 138 mm であった。歩容映像の撮影状況の模式図を図 3 に示す。歩行区間は約 5 m、歩行区間とカメラとの距離は約 20 m とした。

撮影実験協力者に対して、通常の歩行、「すたすた」、「のろのろ」、「よろよろ」、「どっしどっし」、「せかせか」、「てくてく」、「とぼとぼ」、「のし のし」、「よたよた」、「ぶらぶら」の 11 種類のうち、指定した数種類を表現するように指示した。これらのオノマトベは、歩行に関するオノマトベとしてオノマトベ辞典 [1] に掲載されているもののうち、ABAB 型であるものの中から、構成する音韻の多様性を考慮しながら著者らが選択した。歩行者は日本語を母語とする 20 代の男性 7 名であった。

図 3 に示すように、歩行者はまずカメラに近づく向きに歩き、歩行区間の端に達したところで一旦静止し、180 度向きを変え

(注1) : <https://www.ptgrey.com/flea3-usb3-vision-cameras> [2016/10/3 参照]



図 4 主観評価に用いたインタフェース

てカメラから離れる向きに歩いた。映像はすべて 527 × 708 画素, 60 fps で撮影した。この撮影実験により, 通常の歩行, 「すたすた」, 「のろのろ」, 「よろよろ」, 「どっしどっし」を表現した歩容の映像を各 22 本, 「せかせか」, 「てくてく」, 「とぼとぼ」, 「のしのし」, 「よたよた」, 「ぶらぶら」を表現した歩容の映像を各 8 本ずつ, 合計 158 本の映像を得た。

### 3.2 第 3 者評価に基づく歩容とオノマトペの対応付け

データセットの撮影時に, 歩行者には特定のオノマトペを主観的に表現するように指示したが, そのようにして得られた歩容が客観的に見てそのオノマトペを表現できているとは限らない。そこで, 第 3 者による主観評価に基づいて, 改めて歩容とオノマトペの対応付けを行なった。

主観評価実験には 3.1 節で得られた映像のうち, 歩行者の前面を撮影した 79 本を用いた。日本語を母語とする 20 代の男女 14 名に対して映像を提示し, その映像に対応すると思うオノマトペを, 前述の 10 種の中から, 複数回答を許して回答を得た。また, 主観評価実験の被験者には撮影実験の被験者が含まれる。ただし, 主観評価実験は撮影実験から十分間隔をあけて実施した。主観評価実験で用いたインタフェースを図 4 に示す。映像 1 本あたり 7 人から回答を得て, その過半数である 4 人以上が対応づいていると回答したオノマトペを映像にラベル付けした。各オノマトペに対応付いた映像の数を表 1 に示す。表 1 の各行が, 歩行者の主観表現によるクラス (以下, 主観定義と呼ぶ), 各列が第 3 者の主観評価実験によってラベル付けされたクラス (以下, 客観定義と呼ぶ) である。なお, 主観評価の結果, 複数のオノマトペが過半数票を得る場合や, いずれのオノマトペも過半数票を得ない場合が存在するため, 撮影した映像数 (79 本) と表 1 の合計値は一致しない。表 1 を見ると, 主観定義と客観定義にはあまり相関が無いことがわかる。

なお, 歩行者の背面を撮影した映像については, 対になる前面を撮影した映像と同じオノマトペをラベル付けするものとした。

## 4. 評価実験

本節では, 2. で提案した映像とオノマトペ間を対応付ける枠組みの妥当性を検証するための評価実験について述べる。まず 3.1 節で, 実験標本の作成方法について述べる。次に 3.2 節

表 1 主観評価によるラベル付け結果

映像種別	ラベル付けされた映像数										合計
	すた	のろ	よろ	どっし	せか	てく	とぼ	のし	よた	ぶら	
通常	5	0	6	0	0	0	0	0	0	0	11
すた	11	4	2	0	0	0	0	0	0	0	17
のろ	0	0	0	0	2	0	5	2	0	0	9
よろ	0	0	0	0	0	0	0	2	2	7	11
どっし	0	0	1	7	0	1	0	0	0	0	9
せか	1	3	0	0	0	0	0	0	0	0	4
てく	1	1	1	0	0	0	0	0	0	0	3
とぼ	0	0	0	0	3	0	0	0	0	0	3
のし	0	0	0	0	2	0	1	0	1	0	4
よた	0	0	0	0	2	0	0	0	0	2	4
ぶら	0	0	0	0	0	0	0	2	0	1	3
合計	18	8	10	7	9	1	6	6	3	10	78

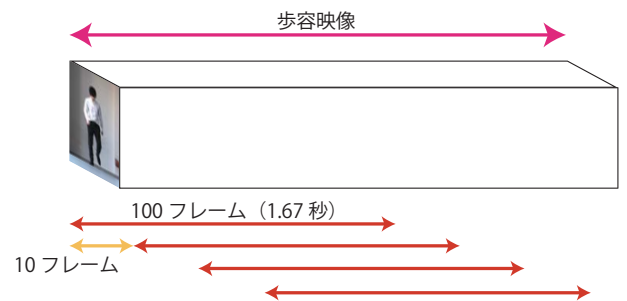


図 5 標本の切り出し方法

で, 音韻空間上での多クラス分類実験について述べる。更に 3.3 節で, 学習に用いないオノマトペを主観表現した歩容に対して, 任意のオノマトペを生成する実験について述べる。最後に 3.4 節で, 実験結果について考察する。

### 4.1 実験に用いる標本の作成

3. で作成したデータセットは映像長が一定ではないため, これを実験で扱いやすくするために, 元の映像から固定長の部分映像を標本として切り出した。具体的には, 図 5 に示すように, 開始フレームを 10 フレームずつずらしながら 100 フレーム分 (1.67 秒) の映像を順次切り出した。映像長を 100 フレームとしたのは, 歩容の 1 周期 (2 歩) が十分取まる長さであるためである。以下の実験では, 各標本は同じ映像から切り出されたものであっても別物として扱う。

また, FMP の処理速度の都合により, 映像の解像度を 128 × 172 画素に縮小した。

### 4.2 音韻空間上での多クラス分類

提案手法で得られる射影の妥当性を検証するため, 音韻空間上での歩容の多クラス分類実験を行なった。データセットを学習用と評価用に分け, 学習用データで回帰モデルを学習し, それを用いて評価用データを射影する。そして, 射影した音韻空間上で Nearest Neighbor 法により多クラス分類を行なう。

本実験では, 分類のクラスとして, データセットに含まれる 10 種類のオノマトペのうち, 主観評価実験によりラベル付けされた映像数が少なかった「てくてく」及び「よたよた」

表 2 客観定義による多クラス分類結果の Confusion Matrix

	すた	のろ	よろ	どっし	せか	とぼ	のし	ぶら	再現率
すた	76	19	46	0	6	19	13	31	0.36
のろ	26	3	20	2	7	4	1	7	0.04
よろ	58	65	45	3	7	20	6	24	0.20
どっし	19	28	3	98	9	5	9	9	0.54
せか	63	23	57	4	54	72	19	140	0.13
とぼ	34	56	32	1	59	27	21	35	0.10
のし	33	70	47	5	35	45	26	68	0.08
ぶら	81	26	18	7	74	14	21	160	0.40
適合率	0.19	0.01	0.17	0.82	0.22	0.13	0.22	0.34	

表 3 主観定義による多クラス分類結果の Confusion Matrix

	すた	のろ	よろ	どっし	せか	とぼ	のし	ぶら	再現率
すた	56	19	5	4	0	1	10	11	0.53
のろ	2	318	64	4	0	24	28	33	0.67
よろ	14	248	162	14	0	18	15	44	0.31
どっし	25	51	16	184	1	8	27	11	0.57
せか	18	13	17	3	0	0	0	1	0.00
とぼ	2	70	88	4	0	4	3	3	0.02
のし	7	61	96	11	0	8	4	9	0.02
ぶら	7	89	78	4	0	3	9	2	0.01
適合率	0.43	0.37	0.31	0.81	0.00	0.06	0.04	0.02	

を除いた 8 種類のオノマトペを用いた。評価は、歩行者別の leave-one-person-out 交差検定で行なった。すなわち、歩行者 7 人中 6 人の歩容を学習用データとし、1 人の歩容を評価用データとする試行を 7 回行なった。評価指標には次式の平均マクロ F 値を用いた。ここで、 $C$  はクラス数、 $c$  はクラス番号、 $PRE_c$  はクラス  $c$  の適合率、 $REC_c$  はクラス  $c$  の再現率である。本実験では  $C = 8$  である。

$$F = \frac{1}{C} \sum_{c=1}^C \frac{2 \cdot PRE_c \cdot REC_c}{(PRE_c + REC_c)} \quad (4)$$

比較のため、主観定義によるデータセットを用いた実験結果も併記する。実験の結果、客観定義の平均マクロ F 値は 0.23 となった。一方、主観定義の平均マクロ F 値は 0.25 となった。客観定義による分類結果の Confusion Matrix を表 2 に、主観定義による分類結果の Confusion Matrix を表 3 に示す。

客観定義の結果は主観定義の結果よりも劣るという結果になった。この理由については 4.4 節で考察する。しかし、Chance Rate (0.12) を上回る分類性能が得られていることから、提案手法によって得られた射影にある程度の妥当性があると言える。

#### 4.3 未知の歩容映像への任意のオノマトペ生成

音韻空間が連続性を持つと仮定し、未知の歩容に対して適当なオノマトペを付与できる可能性を検証するため、歩容映像標本に対するオノマトペ生成実験を行なった。音韻空間上に射影された 32 次元のベクトルを、各 8 次元で構成される音韻 4 つに分解し、音韻ごとに Nearest Neighbor 法で多クラス分類を行なうことで、オノマトペを生成した。

表 4 「てくてく」歩容映像標本へのオノマトペ生成結果

「てくてく」から近い例		「てくてく」から遠い例	
誤差	生成オノマトペ	誤差	生成オノマトペ
58.81	とじよとじよ	156.47	のしのし
69.30	そじゃそじゃ	141.93	のしのし
73.79	とそとそ	140.86	どしどし
77.33	そそそそ	130.87	のしのし
80.76	とさとさ	128.16	としとし
83.03	とさとさ	126.98	としとし
95.94	とそとそ	125.72	としとし

表 5 「よたよた」歩容映像標本へのオノマトペ生成結果

「よたよた」から近い例		「よたよた」から遠い例	
誤差	生成オノマトペ	誤差	生成オノマトペ
38.64	ぶらぶら	149.06	づびゃづびゃ
39.06	ぶさぶさ	126.98	づびよづびよ
39.69	ぶらぶら	120.19	つびよつびよ
43.33	すらすら	113.13	づびよづびよ
44.17	ろろろろ	110.78	ぶっちゃぶっちゃ
44.66	ぶさぶさ	110.28	ぶびゃぶびゃ
46.81	ぶらぶら	110.10	つばつば

表 6 両データセットにおける特徴量のクラス内分散の比較

定義	すた	のろ	よろ	どっし	せか	とぼ	のし	ぶら	平均
客観	0.73	0.77	0.90	1.04	1.16	0.84	1.01	1.42	0.98
主観	0.90	1.17	1.04	1.19	0.91	1.43	0.66	1.02	1.04

本実験では、客観定義によって、4.2 節で用いなかった「てくてく」及び「よたよた」のラベルが付けられた映像標本を未知歩容映像標本として用いた。まず、客観定義により「てくてく」及び「よたよた」以外のオノマトペがラベル付けされた標本をすべて用いて回帰モデルを学習した。そして、「てくてく」及び「よたよた」のラベルが付けられた標本を射影し、各々についてオノマトペを生成した。

射影された音韻空間上の点と真値(てくてく、よたよた)との音韻空間上での距離(誤差)と、生成されたオノマトペの例を表 4 及び表 5 に示す。なお、「てくてく」歩容映像標本の射影の真値からの距離の平均は 106.13、「よたよた」歩容映像標本の射影の真値からの距離の平均は 67.66 であった。

#### 4.4 考察

4.2 節において、客観定義によるデータセットを用いた実験結果よりも、主観定義によるデータセットを用いた実験結果の方が良い結果が得られた。物事の様子を表すというオノマトペの言語的特性をふまえれば、主観定義よりも客観定義の方が妥当な定義であると考えられるので、この実験結果は望ましくない。

ここで、主観定義及び客観定義による両データセットの、映像特徴のクラス内分散の比較を表 6 に示す。これを見ると、客観定義のクラス内分散が主観定義のクラス内分散よりも大きくなっているクラスが複数存在していることがわかる。客観定義が主観定義よりも妥当であれば、客観定義のクラス内分散は主観定義のクラス内分散よりも小さくなるべきである。必ずしも

そうならなかった理由として、本報告で用いた特徴量がオノマトペ記述に適した歩容の特徴を捉えきれていない可能性が考えられる。これは、今後特徴量をさらに検討することによって、客観定義の分類性能が向上する可能性があることを示唆している。

また、オノマトペ生成実験では、「よたよた」よりも「てくてく」の方が真値からの距離（誤差）が大きい傾向が見られた。本実験では、学習に「すたすた」、「のろのろ」、「よろよろ」、「どっしどっし」、「せかせか」、「とぼとぼ」、「のしのし」、「ぶらぶら」の8種類を用いたため、「よたよた」を構成するすべての音韻は学習標本に含まれているが、「てくてく」に関しては、第2モーラの母音「u」が学習標本に含まれていなかった。そのため、「よたよた」の方がより正確な射影が得られた可能性がある。

## 5. む す び

本報告では、人体部位間の相対的な動きと音韻との関係性を利用し、任意の歩容をオノマトペで記述する手法について検討した。提案手法では、映像から人体部位の座標を検出した結果に基づいて、人体部位の相対的位置関係に基づく特徴を抽出し、それらの特徴量をSVRによる回帰を利用して音韻空間へと射影した。

実験では、射影結果を利用して多クラス分類を行なうことで、射影の妥当性を検証した。また、射影結果からオノマトペを生成することで、任意の歩容に対して対応するオノマトペを付与できる可能性について検討した。

今後の課題として、射影方法の改良、特徴抽出における深層学習の利用などが挙げられる。4.2節の実験で得られた分類結果は良い結果とは言えない。本報告では回帰に線形SVRを利用したため、音韻空間の各次元は別々の回帰モデルで計算されており、空間の軸同士の関係性が考慮されていない。回帰手法を検討することによって、より良い分類結果が得られる可能性がある。また、近年注目を集めている深層学習を特徴抽出に利用することで、より音韻空間と親和性の高い特徴が抽出できる可能性がある。これらの方策を用いて、対応付け精度を向上させるのが当面の課題である。

また、本報告では歩容からオノマトペを生成する手法について検討したが、映像特徴空間と音韻空間の対応を利用すれば、逆にオノマトペから歩容を生成することも可能であると考えられる。

謝辞 データセットの撮影、及び被験者実験にご協力頂いた諸氏に感謝する。また、本研究の一部は科研費及び名古屋大学大学院実世界データ循環学リーダー人材養成プログラムの支援による。

## 文 献

- [1] 小野正弘, “擬音語・擬態語日本語 4500 オノマトペ辞典”, 小学館, 2007.
- [2] 田守育啓, ローレンス スコウラップ, “オノマトペー形態と意味”, くろしお出版, 1999.
- [3] 神原啓介, 塚田浩二, “オノマトペ”, インタラクティブシステムとソフトウェアに関するワークショップ (WISS) 2008 予稿集, pp.79-84, Nov. 2008.
- [4] 小松孝徳, 秋山広美, “ユーザの直感的表現を支援するオノマトペ

表現システム”, 信学論 (A), Vol.J92-A, No.11, pp.752-763, Nov. 2009.

- [5] 藤野良孝, 井上康生, 吉川政夫, 仁科エミ, 山田恒夫, “運動学習のためのスポーツオノマトペデータベース”, 日本教育工学会論, Vol.29, pp.5-8, Mar. 2005
- [6] 加藤大貴, 平山高嗣, 川西康友, 道満恵介, 井手一郎, 出口大輔, 村瀬洋, “人体部位の相対的位置関係を利用したオノマトペ歩容映像の識別に関する検討”, 情処研報, 2016-CVIM-202, May 2016.
- [7] 鍵谷龍樹, 白川由貴, 土斐崎龍一, 渡邊淳司, 丸谷和史, 河邊隆寛, 坂本真樹, “動画と静止画から受ける粘性印象に関する音象徴性の検討”, 人工知能学論, Vol.30, No.1, pp.237-245, Jan. 2015.
- [8] 杉山雄紀, 近藤敏之, “ロボットの歩行動作設計によるオノマトペ情報表現の共通理解”, 第25回人工知能学全大, 1C1-OS4a-4, June 2011.
- [9] Y. Yang and D. Ramanan, “Articulated human detection with flexible mixtures-of-parts”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 35, No. 12, pp. 2878-2890, Dec. 2013.
- [10] 戸本裕太郎, 中村剛士, 加納政芳, 小松孝徳, “音素特徴に基づくオノマトペの可視化”, 日本感性工学論, Vol.11, No.4, pp.545-552, July 2012.