一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2017-25,SP2017-1(2017-06)

音韻と人体部位の動きの関係に着目した オノマトペによる歩容の記述に向けて

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††

† 名古屋大学　大学院情報科学研究科
†† 名古屋大学　大学院情報学研究科
††† 中京大学　工学部 〒 470–0393 愛知県豊田市貝津町床立 101
†††† 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: †katoh@murase.is.i.nagoya-u.ac.jp, ††{hirayama,kawanishi,ide,murase}@i.nagoya-u.ac.jp

あらまし　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．また，オノマトペには音象徴性があり，オノマトペから連想されるイメージはその音韻と強い関係があるとされる．このことから，音象徴性に基づく「音韻空間」を歩容の特徴空間と対応付けることができれば，歩容の微妙な違いをオノマトペの音韻の違いで記述することができると考えられる．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ことを踏まえ，本報告では人体部位の相対的な動きに基づく映像特徴を利用し，深層学習を用いた回帰モデルにより両空間を対応付ける手法を提案する．評価実験により，提案手法の有効性を確認するとともに，任意の歩容をオノマトペで記述できる可能性を検討した．
キーワード　オノマトペ，歩容，音韻，人体部位，音象徴性

Toward Description of Gaits by Onomatopoeia Based on the Relationship between Phoneme and Body-Parts Movement

Hirotaka KATO†, Takatsugu HIRAYAMA††, Yasutomo KAWANISHI††, Keisuke DOMAN†††,††, Ichiro IDE††, Daisuke DEGUCHI††††,††, and Hiroshi MURASE††

† Graduate School of Information Science, Nagoya University
†† Graduate School of Informatics, Nagoya University
††† School of Engineering, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470–0393 Japan
†††† Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
E-mail: †katoh@murase.is.i.nagoya-u.ac.jp, ††{hirayama,kawanishi,ide,murase}@i.nagoya-u.ac.jp

Abstract　Gaits are expressed by various onomatopoeias according to their appearance. It is said that onomatopoeia has sound-symbolism and its phoneme is strongly related to its impression. Thus, if the "phonetic-space" based on sound-symbolism can be corresponded with feature-space of gaits, subtle difference of gaits could be expressed as difference in phoneme. Since previous studies imply the relative body-parts movement is associated to the imagination of onomatopoeias when we look at gaits, in this report, we propose a method to convert the relative body-parts movements to onomatopoeias using a deep learning based regression model. Through experiments, we confirmed the effectiveness of the proposed method, and discussed the potential of description of an arbitrary gait by onomatopoeia.
Key words　Onomatopoeia, gait, phoneme, body-parts, sound-symbolism

1. はじめに

事象の様子を直感的に表現する言葉として，口語表現では「のろのろ」,「つるつる」,「しゃかしゃか」など，オノマトペが使用される．このような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．

オノマトペは音象徴性という性質を持ち，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．例えば神原ら [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 [10] 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．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．このうち，音響信号 [5]〜[7] や画像 [8], [9] に関しては工学的な研究例も存在するが，映像とオノマトペとの関係はほとんど検討されてこなかった．このような背景のもと，我々は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 [11], [12]．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．また，歩容をオノマトペという直感的な表現を用いて記述できれば，例えば自動車運転者の注意誘導のためのわかりやすい音声提示などへの応用も期待できる．

関連研究として，鍵谷ら [13] 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 [14] 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，体の部位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．

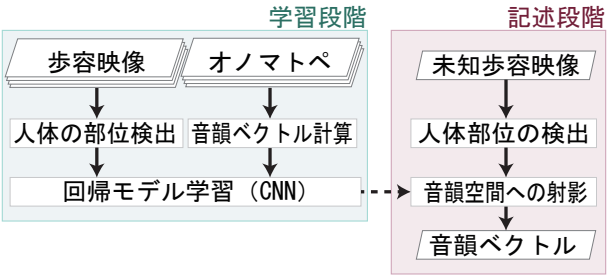


図 1　歩容映像の音韻空間への射影手法の処理手順

これらをふまえ，我々の先行研究 [11] 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とで，歩容映像とオノマトペと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．この際，前述の杉山ら [14] の知見に基いて人体部位の相対的な位置関係を特徴として利用し，その有効性を確認した．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，続く先行研究 [12] で我々はオノマトペを構成する音素が持つ特徴に注目し，この特徴によって張られる特徴空間である「音韻空間」の利用を提案した．この音韻空間上では，音象徴性に基づき連続的にオノマトペの印象が変化していくことが期待できるため，回帰によって音韻空間と映像特徴空間を対応付けることにより，映像とオノマトペをより柔軟に対応付けることができる．しかし先行研究 [12] では評価実験における提案手法の性能が低い問題があった．これは，人体部位の相対的な位置情報に関して，時間方向の統計量を映像特徴として用いており特徴量の表現力が低かったことが原因と考えられる．また，線形 SVR を用いて音韻空間の各次元を個別に回帰したため，音韻が持つ印象の共起を学習することができないという問題点もあった．

そこで，本報告では深層学習を用いた回帰モデルを利用することでこれらの問題の解決を試みる．

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．

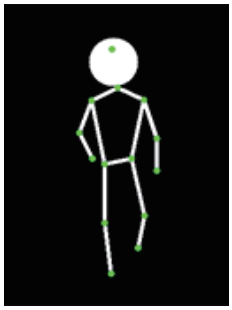
2. 歩容映像の音韻空間への射影手法

提案手法の処理手順を図 1 に示す．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 Convolutional Pose Machines

(a) 元の映像フレーム



(b) 部位検出結果

図 2　CPM による部位検出結果の例

(CPM) [16] を利用する．CPM は，深層学習モデルを用いた姿勢推定手法であり，入力画像に対して，人体の部位 14 か所の位置座標を検出する．入力映像中のあるフレームに対して CPM 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 CPM を適用し，各部位の位置座標系列 $P(p,t)$ を得る．ここで，$p \in \{0,\ldots,13\}$ は各部位の識別子である．$t \in \{1,\ldots,T\}$ はフレーム番号である．ここで映像長を T とした．

得られた位置座標系列 $P(p,t)$ から，14 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 距離を用い，単位は画素とする．

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．

$$L_{p_1,p_2}(t) = \frac{D_{p_1,p_2}(t)}{\bar{H}} \tag{1}$$

$$\bar{H} = \frac{1}{T}\sum_{t=1}^{T} H(t) \tag{2}$$

$p_1 < p_2$ を満たす p_1 と p_2 の組み合わせは ${}_{14}\mathrm{C}_2 = 91$ 通りである．

2.2　回帰モデルの構築

本報告では，音韻空間として [15] で提案されている 32 次元のオノマトペベクトルを利用する．これは，ABAB 型（「すたすた」，「のろのろ」等の同じ 2 音が 2 回繰り返される形）のオノマトペを構成する 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ている．各属性が持つ値は，音韻がその属性に与える影響の大きさによって 2，1，0，−1，−2 のいずれかの値をとる．なお，これらの値はすべて経験的に定義されている．2.1 節で得た正規化された部位の相対距離系列を入力として回帰し，32 次元の音韻空間上に射影する．本報告では回帰モデルとして，深層学習モデルの一種である 1 次元の CNN（Convolutional Neural Network）を用いる．入力層で相対距離系列それぞれをチャンネルとみなしてチャンネル数 91，ユニット数 T の入力を受け付け，出力層は上述の 32 次元のオノマトペベクトルを出力する．

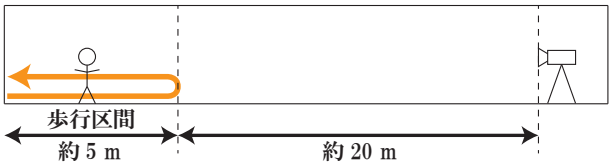


図 3　歩容映像の撮影状況の模式図

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ペ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ペ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(注1) 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．これらのオノマトペは，歩行に関するオノマトペとしてオノマトペ辞典 [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

(注1)：https://www.ptgrey.com/flea3-usb3-vision-cameras/ [2017/5/29 参照]

図 4　主観評価に用いたインタフェース

観的に表現するように指示したが，撮影実験協力者がイメージ通りに体を動かせるとは限らないため，得られた歩容は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協力者が含まれるが，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 名から回答を得て，その過半数である 4 名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．ここで，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 4.1 節で，実験標本の作成方法について述べる．次に 4.2 節で，音韻空間上での多クラス分類実験について述べる．更に 4.3 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 4.4 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 s フレームずつずらしながら 100 フレーム分

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	0	0	0	6	0	0	0	0	11
すた	11	0	0	0	4	2	0	0	0	0	17
のろ	0	5	0	0	0	0	2	0	0	2	9
よろ	0	0	7	0	0	0	0	0	2	2	11
どっし	0	0	0	7	0	1	0	1	0	0	9
せか	1	0	0	0	3	0	0	0	0	0	4
てく	1	0	0	0	1	1	0	0	0	0	3
とぼ	0	0	0	0	0	0	3	0	0	0	3
のし	0	1	0	0	0	0	2	0	1	0	4
よた	0	0	2	0	0	0	2	0	0	0	4
ぶら	0	0	1	0	0	0	0	0	0	2	3
合計	18	6	10	7	8	10	9	1	3	6	78

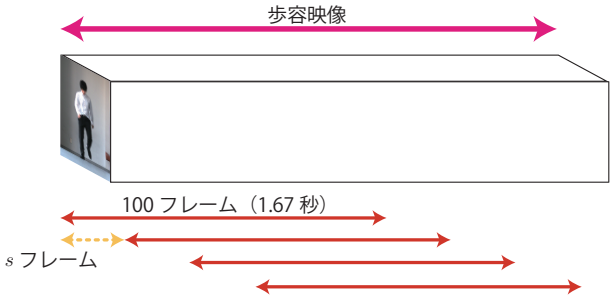


図 5　標本の切り出し方法

（1.67 秒）の映像を順次切り出した．すなわち $T = 100$ とした．この際，各クラスの標本数を大まかに揃えるため，標本数が最も少なくなる（延べ映像長が最も短い）クラスの標本を $s = 5$ として切り出し，それ以外のクラスは適宜 s の値を調整することで標本を間引いた．映像長を 100 フレームとしたのは，歩容の 1 周期（2 歩）が十分収まる長さであるためである．

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最近傍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「のしのし」及び「よたよた」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 名中 6 名の歩容を学習用データとし，1 名の歩容を評価用データとする試行を 7 回行なった．評価指標には正解率（Accuracy）を用いた．比較手法として，CNN の代わりに LSTM を利用した手法と，我々の先行研究 [12] の手法を用いる．LSTM は深層学習において時系列データを用いる際に利用する一般的なアーキテクチャである．実験で用いた CNN および LSTM の構造をそれぞれ表 2，表 3 に示す．これらの深層学習モデルの実装

表 2　実験に用いた CNN のアーキテクチャ

Input	Units: 100, Channel: 91
Convolution 1	Kernel: 10, Channel: 128, Maxpooling: 10
Convolution 2	Kernel: 10, Channel: 128, Maxpooling: 10
Output	Units: 32

表 3　実験に用いた LSTM のアーキテクチャ

Input	Length: 100, Units: 91
Fully-connect	Units: 100
LSTM	Units: 100
Output	Units: 32

表 4　音韻空間上での多クラス分類結果

手法	客観定義	主観定義
提案手法（CNN）	0.474	0.384
比較手法（LSTM）	0.334	0.312
比較手法（先行研究 [12]）	0.338	0.278

には Keras(注2) を用いた．先行研究 [12] の手法は，正規化された部位の相対距離系列 $L_{p_1,p_2}(t)$ の時間方向の分散および尖度を特徴量とし，音韻空間の各次元について線形 SVR を用いて回帰する手法である．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．実験の結果を表 4 に示す．なお，8 クラス分類であるため chance rate は 0.125 である．

主観定義，客観定義の双方で提案手法である CNN が最も良い結果となり，提案手法の有効性を確認した．また，すべての手法で客観定義が主観定義を上回る結果となった．事象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であると考えられるので，この実験結果は妥当である．

4.3　未知の歩容映像への任意のオノマトペ付与

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に，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，音韻ごとに最近傍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．

本実験でも 4.2 節と同じく「のしのし」及び「よたよた」を除いた 8 種類のオノマトペのデータセットを用いた．評価はオノマトペ別の leave-one-onomatopoeia-out 交差検定で行なった．すなわち，オノマトペ 8 種中 7 種を学習用データとし，1 種を評価用データとしてオノマトペを付与する試行を 8 回行なった．付与結果の例と，その射影から真値までの音韻空間上の距離を表 5 に示す．

4.4　考　　察

音韻空間を用いた分類は，学習に用いたオノマトペの単純な多クラス分類問題を，学習に用いないオノマトペも認識できるよう拡張したものと考えることができる．ここで音韻空間を用いず単純な多クラス分類を行なった場合との比較を表 6 に示す．具体的には，表 2 に示した CNN の出力層を 8 とし，8 クラス分類問題として学習を行なった．客観定義では音韻空間を用いた場合でも単純な多クラス分類の場合とほぼ同等の分類性能となっており，音韻空間上での分類は単純な多クラス分類の妥当な拡張であることを示唆している．一方，主観定義では音韻空間を利用することで若干分類性能が下がっている．これは，主観定義が音韻空間にあまり適合しない，すなわち音象徴性に基づく直感的な印象と乖離していることを示している可能性がある．

また，4.2 節の実験においては CNN を用いた方が LSTM を用いた場合よりも良い結果が得られた．しかし，LSTM は CNN と比較して構造が複雑であるため，LSTM が良い結果を得られなかったのは学習標本が足りないためであり，学習標本が十分多ければ CNN よりも良い結果が得られる可能性も考えられる．これを検証するために追加で実験を行なった．標本数を増やすことは難しいため，逆に減らして実験することによりその変化を調べた．具体的には，4.2 節では標本切り出し時の最小ずらし幅が $s = 5$ であったものを $s = 10$ に変更した．これにより学習標本数は客観定義で平均で 627 から 347 に，主観定義では 673 から 360 に減少した．そのほかの実験条件は 4.2 節に準ずる．結果を表 7 に示す．標本を減らすことにより，CNN，LSTM は双方とも概ね同じ程度悪化していることが分

表 5　歩容に付与されたオノマトペの例

客観定義ラベル	生成オノマトペ	距離
すたすた	せかせか	30.5
すたすた	すかすか	31.6
すたすた	せこせこ	33.7
のろのろ	よろよろ	14.8
のろのろ	よらよら	18.6
のろのろ	とろとろ	22.7
よろよろ	のろのろ	17.2
よろよろ	のらのら	21.3
よろよろ	ぬらぬら	24.0
どっしどっし	とっのとっの	69.6
どっしどっし	つこつこ	69.8
どっしどっし	とことこ	70.0
せかせか	せっつせっつ	22.2
せかせか	てっつてっつ	22.6
せかせか	てくてく	22.8
てくてく	つかつか	50.0
てくてく	すかすか	50.6
てくてく	とかとか	51.7
とぼとぼ	のろのろ	51.6
とぼとぼ	のそのそ	51.6
とぼとぼ	ろろろろ	53.6
ぶらぶら	のろのろ	36.6
ぶらぶら	とろとろ	38.3
ぶらぶら	ろろろろ	38.4

表 6　単純な多クラス分類との比較

	客観定義	主観定義
単純な多クラス分類	0.471	0.442
音韻空間上での分類	0.474	0.384

(注2)：https://keras.io./ [2017/5/29 参照]

表 7　学習標本数を減らした場合の比較

	客観定義		主観定義	
ずらし幅	CNN	LSTM	CNN	LSTM
$s = 5$	0.474	0.334	0.384	0.312
$s = 10$	0.429	0.299	0.373	0.283

かる．これは LSTM のみが学習標本不足で不当に低い評価を得ているわけではなく，提案手法である CNN が LSTM と比較して優れていることを示唆している．

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．

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，手法の有効性を検証した．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，ネットワークアーキテクチャの改良，オノマトペの意味が音象徴性に与える影響の調査などが考えられる．4.2 節の実験において提案手法の有効性を確認することができたが，未だ十分な分類性能が得られているとは言い難い．本報告で用いた CNN は単純な構造であり，まだ改良の余地が多分に存在すると考えられる．同時に，データ不足に対応するため，転移学習の利用も検討する必要がある．また，本研究では音象徴性の理論に基づいて，オノマトペの印象はオノマトペを構成する音韻の種類のみによって決まるという仮定をおいているが，「すたすた」などの長年使われてきた一般的なオノマトペは，その辞書的な意味が印象に影響を与えるため，この仮定が厳密には成り立たない可能性が高い．本手法を発展させることで，その影響を定量的に分析することができると考えられる．

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，本研究の一部は栢森情報科学振興財団，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助成を受けて実施された．

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―形態と意味―”，くろしお出版，1999.
[3] 神原啓介，塚田浩二，“オノマトペン”，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] 石原一志，坪田康，奥乃博，“日本語の音節構造に着目した環境音の擬音語への変換”，信学技報，SP2003-38，June 2003.
[6] 比屋根一雄，澤部直太，飯尾淳，“単発音のスペクトル構造とその擬音語表現に関する検討”，信学技報，SP97-125，Mar. 1998.
[7] S. Sundaram and S. Narayanan. “Classification of sound clips by two schemes: Using onomatopoeia and semantic labels”, Proc. 2008 IEEE Int. Conf. on Multimedia and Expo, pp.1341–1344, June 2008.
[8] W. Shimoda and K. Yanai, “A visual analysis on recognizability and discriminability of onomatopoeia words with DCNN features”, Proc. 2015 IEEE Int. Conf. on Multimedia and Expo, pp.1–6, July 2015.
[9] 權眞煥，川嶋卓也，下田和，坂本真樹，“DCNN を用いた画像の質感認知―音象徴性からのアプローチ―”，第 31 回人工知能学全大，2L3-OS-09b-1，May 2017.
[10] 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.
[11] 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.
[12] 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，“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”，HCG シンポジウム，HCG2016-A-3-5，Dec. 2016.
[13] 鍵谷龍樹，白川由貴，土斐崎龍一，渡邊淳司，丸谷和史，河邉隆寛，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.
[14] 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.
[15] 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.
[16] S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, “Convolutional Pose Machines”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.4724–4732, June 2016.