

音響特徴を用いたオノマトペの用法分類に関する一考察

A Consideration on Usage Classification of Onomatopoeia using Acoustic Features

浦田大貴^{*1}

Daiki Urata

中村剛士^{*1}

Tsuyoshi Nakamura

加納政芳^{*2}

Masayoshi Kanoh

山田晃嗣^{*3}

Koji Yamada

^{*1}名古屋工業大学

Nagoya Institute of Technology

^{*2}中京大学

Chukyo University

^{*3}情報科学芸術大学院大学

Institute of Advanced Media Arts and Sciences

Onomatopoeias can simply describe sounds or state of things. Therefore it is expected that onomatopoeias can be a useful communication tool between humans and machines as well as between humans. Previous works focused on phonemes of onomatopoeias and quantified them. In our study, we focus on utterances of onomatopoeias and uses acoustic features of the utterance. This paper evaluated an accuracy rate of usage classification of onomatopoeias between a previous work and our work. As the result, our work can mostly achieve higher accuracy than the previous work.

1. はじめに

物事の状態や様子を適切に伝える言語表現として、オノマトペが注目されている。オノマトペとは、擬音語・擬態語の総称であり、実際に発生した音を言語化したものが多く存在する。オノマトペには、音象徴という特性があるとされる。音象徴とは音声や音韻から特定のイメージができる現象である。例えば、「ドンドン」という発話されたオノマトペからは、物体を叩いている動作を感覚的に想像することができる。また、オノマトペには、「何かを表現したいけれどもうまく言語化して表現できなかったモヤモヤとした曖昧な印象」が含まれるとされる[小松 09]。オノマトペには、説明が困難な複雑な事柄を直感的かつ簡便に表現・伝達することができる可能性がある。このことから、オノマトペをインタフェースとして用いることで、例えば Adobe Illustrator に代表されるようなデザインツールの複雑な操作を感覚的に行うことが可能になると期待される。

本研究では、発話されたオノマトペの定量化を試み、発話オノマトペのイメージ伝達の可能性を探る。発話されたオノマトペには、音量・音高・音色・リズム等のテキストには陽に現れない特徴が存在する。これまでの定量化は、テキスト表現されたオノマトペの音韻に注目し、それが与える印象を定量化する試みが主であった。発話オノマトペは、テキスト表現に比べ複雑な情報を持つことから、テキスト以上にイメージを適切に伝達することが可能ではないかと我々は考える。本研究では、発話オノマトペを音響特徴量によって定量化し、その特徴量とイメージ感の関係を明確にすることを目指す。オノマトペをデザインツール等で活用するためには、オノマトペを定量的に扱える必要がある。本報告では、関係調査の一つとして、イメージと関係の深い「用法」に着目しオノマトペの音響特徴量とオノマトペの用法の間の関係を調査した結果について報告する。

2. 関連研究

秋山ら[秋山 11]、[小松 08]はテキスト表現されたオノマトペの音韻から、オノマトペの三つの印象値を算出する重回帰式を作成した。三つの印象値はそれぞれ「キレ・俊敏さ」、「躍動感」、「柔らかさ・丸さ」を表している。以下に、XYXY 型オノマトペ O の三つの印象値 (O_q = キレ・俊敏さ, O_d = 躍動

感, O_s = 柔らかさ・丸さ) を求める重回帰式を式 (1) に示す。式中の $X_a^{(b)}$, $Y_a^{(b)}$ は事前調査により得られたオノマトペの構成母音 $X^{(v)}$, $Y^{(v)}$ および子音 $X^{(c)}$, $Y^{(c)}$ の印象値である。

$$O_q = 0.60X_q^{(c)} + 0.52Y_q^{(c)} \quad (1)$$

$$O_d = 0.59X_d^{(c)} + 0.4Y_d^{(c)}$$

$$O_s = 0.56X_s^{(c)} + 0.46Y_s^{(c)} + 0.22Y_s^{(v)}$$

3. 提案

本研究では、音声の特徴量として MFCC (メル周波数ケプストラム係数) を用いたオノマトペの定量化と用法分類を試みる。短時間フレームごとに MFCC, Δ MFCC を取得することで一つのオノマトペから 2048 次元の特徴量を獲得する。オノマトペの音象徴としての性質を利用し定量化されたオノマトペを用いてオノマトペの持つ用法ごとの分類を試みる。

4. 用法分類実験

本実験では、発話オノマトペとオノマトペが使用される用法の関係性を調べるために、定量化されたオノマトペを実験データとして SVM (サポートベクターマシーン) を用いた二クラス分類を行った。分類対象である発話オノマトペは、Open JTalk^{*1} によって合成した音声をを用い、各音声の発話時間を統一した。以下に、本実験の概略図を示す (図 1)。オノマトペの分類はオノマトペ辞典[小野 07]に従い、「歩く・走る」、「騒ぐ」、「食べる・かむ・なめる」、「飲む・酔う」、「せきをする・むせる」、「言う・話す」の六種類の用法とした。以下に実験にて用いたオノマトペを示す (表 1)。

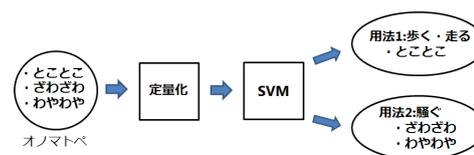


図 1: 用法分類の概要

表 1: 用法別のオノマトベ

用法名	オノマトベ	用法名	オノマトベ
歩く・走る	かっぽかっぽ	飲む・酔う	がぶがぶ
	たたたた		がぼがぼ
	とことこ		ぐびぐび
	ばたばた		こくこく
	ばかばか		ごっごっ
騒ぐ	びたびた	ごぶごぶ	ちゆるちゆる
	ざわざわ	せきをする・むせる	げほげほ
	どたどた		こほこほ
	どやどや		ごほんごほん
	わやわや		ごぼんごぼん
かりかり	ざりざり		
食べる・かむ・なめる	くしゃくしゃ	言う・話す	こそこそ
	くちやくちや		ひそひそ
	ぐちゃぐちゃ		ふがふが
	ぐちゅぐちゅ		ぶつぶつ
	こしこし		ぼそぼそ
	こりこり	れるれる	
	さぶさぶ		
	しゃきしゃき		
	しよりしより		
	つるつる		
	にちやにちや		
	ぱりぱり		
	ぱりぱり		
	べちゃべちゃ		
	ぼりぼり		

用法別に分けられた六種類のオノマトベ群の中から二種類を選択し、定量化した後、SVMによる学習を行った。例えば、「歩く・走る」または「騒ぐ」の用法を持つ 11 種のオノマトベを定量化した後、SVMにより「歩く・走る」または「騒ぐ」の二クラスに分類した。学習結果の評価については、leave-one-out法を用いて、二クラス分類における正解率を算出した。上記の実験をテキスト表現されたオノマトベに対する定量化手法 [小松 08] と本研究による音声表現されたオノマトベに対する定量化手法でそれぞれ行った。

5. 実験結果

分類の組み合わせは 15 通りあるのに対して、本研究の手法の方が文献 [小松 08] の手法より正解率が高かった組み合わせは 13 通りであった。以下に、SVMによる二クラス分類の正解率を表 2, 3 に示す。表 2, 3 はそれぞれ本研究の手法による分類結果と、文献 [小松 08] の手法による分類結果である。便宜上、オノマトベの意味をそれぞれ「歩く・走る」を「歩く」、「言う・話す」を「言う」、「せきをする・むせる」を「せき」、「食べる・かむ・なめる」を「食べる」、「飲む・酔う」を「飲む」と表記する。

表 2: 本研究の分類結果 (SVM による正解率)

	歩く	言う	せき	騒ぐ	食べる	飲む
歩く						
言う	0.6429					
せき	0.7043	0.8571				
騒ぐ	0.6364	0.8182	0.9091			
食べる	0.8846	0.7692	0.8846	0.9565		
飲む	0.7500	0.7500	0.8125	1.0000	0.7857	

表 3: 文献 [小松 08] の分類結果 (SVM による正解率)

	歩く	言う	せき	騒ぐ	食べる	飲む
歩く						
言う	0.5000					
せき	0.2857	0.5714				
騒ぐ	0.6364	0.5455	0.7273			
食べる	0.7308	0.7308	0.7308	0.8261		
飲む	0.8125	0.6250	0.6875	0.6923	0.8214	

6. 考察

実験結果から、本研究の手法の方が文献 [小松 08] の手法より分類精度が高いことが確認された。本研究の手法の方が分類精度が高かった理由として、特徴量の次元数の違いがあげられる。文献 [小松 08] の手法では一つのオノマトベに対して、3次元の特徴量が取得可能であるが、本研究の手法では 2048次元の特徴量を得ることができる、これは、短時間フレームごとに特徴量が取得できる音声データの利点であると言える。しかし、本実験で用いたオノマトベは、擬音語の動作に関する XYXY 型のオノマトベのみであり、任意のオノマトベに対して同様の結果が得られるか調査する必要がある。

本研究の手法の場合、正解率が一番低かったのは「騒ぐ」と「歩く・走る」の分類であるが、これは「騒ぐ」のオノマトベ群の中に「歩く・走る」の用法で使われる可能性があるオノマトベが存在していたからだと考えられる。「騒ぐ」のオノマトベ群には「どたどた」というオノマトベが含まれているが、その意味は「乱暴に音を立てて動き回っている様子」である。このことから、「足音を立てて動いている様子」を表す「歩く・走る」の用法と類似していると考えられる。本研究では用法分類というオノマトベの粗い分類をしたため、このように複数の用法を併せ持つオノマトベが存在した可能性がある。今後はオノマトベの意味ごとの分類を行うことで、より細かな分類を検討していきたい。

7. まとめ

本研究では、音声表現されたオノマトベに着目し、音響特徴である MFCC を用いた定量化とオノマトベが持つ用法ごとの自動分類を行った。音声から MFCC と Δ MFCC を抽出することで 2048次元の特徴量でオノマトベを表した。SVMを用いた用法分類の実験では、テキスト表現されたオノマトベを定量化する従来の手法に比べて、音声表現されたオノマトベを定量化する手法を用いた方が分類精度が高いことが確認された。また、一部の正解率が低くなった原因として複数の用法を併せ持つオノマトベが存在している可能性が考えられる。

今回は、擬音語の動作に関する XYXY 型のオノマトベのみを対象にしたため、任意のオノマトベに対して同様の結果が得られるか調査をする必要がある。また、多種類のオノマトベ間の関係性を調べるために多クラス分類によるオノマトベの意味ごとの分類を検討していきたい。

本研究は JSPS 科研費 15K12106 の助成を受けたものである。

参考文献

- [小松 09] 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ベ表現システム, 電子情報通信学会論文誌 A, J92-A(11), pp. 752-763, 2009.
- [秋山 11] 秋山広美, 小松孝徳, 清河幸子: オノマトベから感じる印象の客観的数値化手法の提案, 情報処理学会研究報告書, Vol. 2011-HCI-142. No23 pp. 1-7 2011
- [小松 08] 小松孝徳, 秋山広美: ユーザの直観的表現を支援するオノマトベ意図理解システムの開発, HAI シンポジウム, 2A-4, 2008
- [小野 07] 小野正弘 (編者): 擬音語・擬態語 4500 日本オノマトベ辞典, 小学館, 2007