

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 17 日現在

機関番号：32692  
研究種目：基盤研究(C)  
研究期間：2012～2014  
課題番号：24500131  
研究課題名(和文) 音声認識にとって未知語となる擬態語を含む音声クエリで動作可能な図柄パターン検索法

研究課題名(英文) Symbolic pattern retrieval using onomatopoeic voice queries including out-of-vocabulary words for automatic speech recognizers

研究代表者  
相川 清明(AIKAWA, Kiyooki)  
東京工科大学・メディア学部・教授

研究者番号：00367202  
交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：言語的な表現が確定していない検索対象を音声で検索する方法を開発することを目的として研究を進めた。音声認識により得られる音素列とその未知語処理による登録表現との類似度、及び音声認識において使用していない大局的音響特徴による登録表現との類似度を併用する検索法を提案した。この方法を、コミュニケーションシステムの1つであるLINEにおいて雰囲気や感情を伝えるために用いられるスタンプの音声検索において検証した。類似度の値に応じて条件分岐を行う方法、および、二種の類似度を同時に用いて判断する方法を開発し、後者の方が高い性能が得られることを示した。開発した方法は擬音語による効果音検索においても検証を行った。

研究成果の概要(英文)：The objective of this research was the development of voice search methods for items which have no fixed linguistic expressions. Two methods were developed. One was the method using Out-Of-Vocabulary processing using phone sequences obtained by speech recognition. The other was the method using super-segmental acoustic features which were not used in speech recognition. The methods were evaluated on the retrieval "STAMPs" for "LINE" communication. Decision-based methods and similarity measure-based method were compared. The latter method outperformed the former method. The search method was also evaluated on sound-effect retrieval using onomatopoeic voice input.

研究分野：音声情報処理・マルチモーダルインタフェース

キーワード：音声検索 擬態語 擬音語 未知語 感性情報 パターン検索 音響特徴

## 1. 研究開始当初の背景

2011年の時点で、音声データを自動音声認識でテキストに変換して検索する「音声ドキュメント検索」の研究が盛んになりつつあった。音声や映像データが大量にインターネット上やコンピュータ内部に存在するようになったからである。研究代表者は2009年度から情報処理学会音声言語情報処理研究会音声ドキュメント処理ワーキンググループ主査を務め、この分野の研究活性化を図っていた。国際会議 Interspeech2010 においては研究代表者をオーガナイザの一人とする音声ドキュメント検索に関するスペシャルセッションが開催された。また、国立情報学研究所の主催する NTCIR ワークショップに上記ワーキンググループから SpokenDoc という検索タスクを提案して採択となった。

音声ドキュメント検索の対象は音声データであったが、今後さまざまなものが検索対象になると考えられた。デザインツールなどで多くの種類がある基本図形や図柄パターン・図案などの画像を、一覧表から視覚的に探すことなく、直接音声で検索できれば便利であると考えた。

ところが、音声入力による図柄パターン検索を試みた結果、検索クエリには音声認識辞書にない即ち未知語の擬態語が数多く現れることがわかった。これらの未知語は誤認識されるが、誤認識結果の部分音素列を用いて最も類似した、正解のわかっている既知クエリを推定する方法を提案し、ある程度の推定精度が得られた。しかし、検索クエリには新造語を含めて様々な擬態語・擬音語が発声されることが多く、根本的な対策が必要であることがわかった。

すなわち、音声入力による図柄パターン検索には問題が2つ存在する。

(1) 図柄パターンは音声と異なり、テキストとの一対一対応がなく、様々に表現され得る。

(2) 検索クエリには自動音声認識にとって未知語の擬態語が頻出する。

ここで、擬態語はそもそも音声・言語で対象を疑似的に表現しようとする試みであり、その音響特徴と図柄パターンの形状の特徴には関連があるはずであるとの考えから、擬態語が正しく音声認識できなくても、擬態語の音響特徴から直接検索対象を検索できるのではないかという着想に至った。

一方、研究代表者はこれまでに、ベクトル空間法による気象情報案内・効果音検索を提案してきた。これらの方法では音声ドキュメント検索のように一旦テキストに変換して検索するのではなく、特徴ベクトルの類似度で検索する。特徴パラメータの類似度を用いる方法では特徴パラメータに元データの情報が残存しているため、不明瞭な特徴に対する検索上の頑健性があることに気付いた。

以上の考察に基づき、複数種類の音響特徴を併用する擬態語・擬音語音声入力による検索法を研究することとした。

## 2. 研究の目的

図柄パターンは多種多様で一覧表から視覚的に探すのは非常に時間を要する作業である。これを音声で検索できれば便利である。ところが、図柄パターンを検索しようとする人は検索クエリとして「ギザギザ」などの擬態語を発声することが多い。その多くは自動音声認識にとって辞書にない未知語であるため、その根本的な対策が必要である。本研究では、擬態語の音響特徴が図柄パターンの形状の特徴と関連があることに注目した。本研究では擬態語音声で与えられる検索クエリから所望の図柄パターンを検索するための方法を研究開発することを目的とした。

## 3. 研究の方法

音声クエリによる図柄パターンの検索について、大きく分けて次の2系統で研究を進めた。

1つは音声認識の音響処理結果を利用して、未知語の擬態語音声クエリに類似した既知クエリを推定して検索する方式である。もう1つは音声認識と同時に、別の音響特徴抽出を行い、直接検索対象を推定する方式である。なお、擬態語のみでなく、擬音語まで拡張し、効果音検索への応用も同時に研究した。

具体的研究課題は以下の5つであった。

(1) 擬態語・擬音語による未知語音声クエリの音響特徴からの既知語クエリの推定法

(2) 音声クエリの擬態語・擬音語と図柄パターン・効果音に共通な特徴パラメータの導出

(3) 特徴パラメータ間の類似度、及び、正解判定のための類似度閾値の設定

(4) 擬態語・擬音語の未知語に対して音響特徴抽出と連携動作する音声認識の音響・言語処理法

(5) 音声対話型で再検索機能のある図柄パターン検索システムの構成法

## 4. 研究成果

(1) 初年度においては、主にデータベース収集を行った。パターンとその擬音語・擬態語の音声表現の組を収集分類整理した。

(2) また、音声による複雑な検索要求が現れることを想定し、音声検索におけるANDの条件の言い回しを分析した。その結果、「と」あるいは単なる空白などの比較的簡単な表現が用いられることが分かった。

(3) 次年度においては、音声認識によって得られる音素列を用いて登録してある擬音語・擬態語との類似性を音素間の類似度を定義することにより求める方式と音声認識において使用していない大局的音響特徴による類似度を使い分けることによるパターン検索法を提案した。この方法は閾値による条件判断の組み合わせにより検索結果を求める方法である。この方法を、コミュニケーションシステムのLINEで雰囲気や感情を伝えるために用いられるスタンプの音声検索にお

いて検証した。51種類の擬態語・擬音語を分析した結果、45種類は音声認識により得られる音素列情報を用いて、6種類は音響特徴を用いて検索できることがわかった。

(4) また、音素列検索を用いる擬音語・擬態語音声による検索法を効果音検索に応用したシステムを作成した。擬音語音声入力による効果音検索実験を行った結果、67%の精度で意図に合った効果音を検索できた。

(5) この他、音声検索の研究の一環として、音声指示による演奏スタイルを制御するシンセサイザの研究も行った。「雄大に」という指示では5段階評価で4.3の高い評価が得られた。

(6) 最終年度においては、主としてLINEのスタンプの音声検索を課題として、擬音語・擬態語によるスタンプ検索における検索性能向上の研究を行った。音素認識結果による類似度と大局的音響特徴分析による類似度をベクトル空間法において同時使用することにより、スタンプ選択性能を向上させることができた。この方法により、論理判断による方法で68%であった正答率を95%にまで向上できた。

同時に、LINEにおいて2ユーザが双方とも音声入力により対話する状況を想定し、音声認識を用いて対話履歴を、また、音声検索を用いてスタンプを表示するシステムを作成した。

(7) この他、感性表現による楽器音と和音の検索を目的として、楽器音の音色と和音の組み合わせによる感情情報表現の研究を行った。特定の楽器と和音の組み合わせにより、「明るい」、「激しい」などの感情を強く表現できることが分かった。

(8) さらに、視覚的イメージと楽曲のイメージの類似性を検討するため、コンサートポスターのデザイン提案システムを作成した。楽曲のイメージ、ポスター全体のイメージとも、60%以上が背景に用いる画像に関係するという結果となった。

(10) 一連の研究を通じて、擬態語・擬音語を用いた検索における個人性の吸収には、複数の特徴の同時並行処理が必要であることが分かった。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

① Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyoaki Aikawa, Yoshiaki Itoh, Tatsuya Kawahara, Seiichi Nakagawa, Hiroaki Nanjo, Yoichi Yamashita, "Overview of the NTCIR-10 SpokenDoc-2 Task", Proceedings of the 10th NTCIR Conference, Vol. 1, pp. 573-587, 2013年6月 (査読有)

② 伊藤慶明, 西崎博光, 中川聖一, 秋葉友

良, 河原達也, 胡新輝 5, 南條浩輝, 松井知子, 山下洋一, 相川清明, 音声中の検索語検出のためのテストコレクションの構築と分析, 情報処理学会論文誌 Vol. 54 No. 2 471-483, 2013年2月. (査読有)

③ NISHIZAKI Hiromitsu, AKIBA Tomoyosi, AIKAWA Kiyoaki, Tatsuya Kawahara, Tomoko Matsui, Evaluation Framework Design of Spoken Term Detection Study at the NTCIR-9 IR for Spoken Documents Task, 自然言語処理 Vol. 19, No. 4, pp. 329-350, 2012年12月. (査読有)

④ Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyoaki Aikawa, Tatsuya Kawahara, Tomoko Matsui, "Designing an Evaluation Framework for Spoken Term Detection and Spoken Document Retrieval at the NTCIR-9 SpokenDoc Task", LREC2012, No. 693, 2012年5月. (査読有)

[学会発表] (計 10 件)

① 雨宮 香織, 相川 清明, "擬態語・擬音語による音声検索を用いたスタンプ描画システムの性能を左右する要因", 日本音響学会 2015年春季研究発表会, pp. 73-74, 2015年3月17日. 中央大学後楽園キャンパス、東京都文京区.

② 関 沙友里, 相川 清明, "楽曲の感性表現によるポスターデザイン提案方法", 日本音響学会 2015年春季研究発表会, pp. 457-458, 2015年3月16日. 中央大学後楽園キャンパス、東京都文京区.

③ 雨宮 香織, 相川 清明, "擬態語・擬音語音声入力を用いたLINE風スタンプ描画システムの動作分析", 情報処理学会研究報告, Vol. 2015-SLP-105, No. 12, pp. 1-6, 2015年2月28日、合歓の郷ホテル&リゾート、三重県志摩市.

④ 佐々木司, 相川清明, "音色と和音の組み合わせによる感情表現", 日本音響学会聴覚研究会資料, Vol. 44, No. 9, H-2014-09, pp. 585-590, 2014年12月20日、休暇村志賀島、福岡県福岡市.

⑤ 雨宮香織, 日置優介, 相川清明, "音声入力による描画インターフェースの類似度の検討", 日本音響学会 2014年秋季研究発表会, pp. 151-152, 2014年9月4日. 北海学園大学豊平キャンパス、北海道札幌市.

⑥ 渡邊智之, 相川清明, "擬音語音声入力による効果音検索システム", 日本音響学会

2014 年 春 季 研 究 発 表 会 ， 3-5-3,  
pp.1038-1039, 2014 年 3 月 12 日, 日本大学  
理工学部、東京都千代田区.

⑦ 富田観月、相川清明，“擬態語・擬音語  
音声入力によるスタンプ描画インタフェー  
ス”，日本音響学会 2014 年春季研究発表会，  
2-Q4-12, pp.144-145, 2014 年 3 月 11 日, 日  
本大学理工学部、東京都千代田区.

⑧ 西崎博光，秋葉友良，相川清明，伊藤慶  
明，河原達也，胡新輝，中川聖一，南條浩輝，  
山下洋一，“NTCIR-10 SpokenDoc-2 Spoken  
Term Detection タスクの結果と知見”，日本  
音響学会 2013 年秋季研究発表会，3-8-6，  
pp.107-110, 2013 年 9 月 27 日、豊橋技術科  
学大学、愛知県豊橋市.

⑨ 秋葉友良，西崎博光，相川清明，胡新輝，  
伊藤慶明，河原達也，中川聖一，南條浩輝，  
山下洋一，“NTCIR-10 SpokenDoc-2 Spoken  
Content Retrieval タスクの結果と知見”，  
日本音響学会 2013 年秋季研究発表会，3-8-7，  
pp.111-114, 2013 年 9 月 27 日、豊橋技術科  
学大学、愛知県豊橋市.

⑩ 高田由梨，相川清明，音声検索における  
AND 検索の言い回しの分析，日本音響学会  
2013 年春季研究発表会，3-9-6, pp.103-104，  
2013 年 3 月 15 日，東京工科大学、東京都八  
王子市.

## 6. 研究組織

### (1) 研究代表者

相川 清明 (AIKAWA, Kiyooki)  
東京工科大学・メディア学部・教授  
研究者番号：00367202