

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 19 日現在

機関番号：32503

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330203

研究課題名(和文)音環境理解のための音カテゴリークラスタリングに関する研究

研究課題名(英文)A study on sound category clustering for auditory scene analysis

研究代表者

大川 茂樹 (OKAWA, Shigeki)

千葉工業大学・先進工学部・教授

研究者番号：40306395

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究は、ロボットの聴覚機構などへの実装を目指して、我々の身の回りに日常的に存在する様々な音に着目し、その物理的および聴覚心理的(意味的)カテゴリーを合理的にクラスタリング(分類)する方法と、様々な音響事象のカテゴリー識別技術を検討することを全体的な目的とし様々な検討を実施した。代表的な成果としては、まず、大規模に収集した様々な音データに対して主観評価に基づく音カテゴリーの表現方法を定義した。次に、識別アルゴリズムとして、GMMやDNN、NMFに基づく方法を検討・実装した。最後に、音環境理解に関連して、小規模空間で発生する音を制御する際にロボティクス技術を応用する手法について検討した。

研究成果の概要(英文)：This study focused on "life-log sound", which is various sounds surrounding our life environments, to determine those physical and semantic categories and to discriminate those acoustic events. Our achievements include: (i) defining a method to express the life-log sound mainly based on the subjective evaluation; (ii) investigating and implementing some algorithms to discriminate sounds based on GMM, DNN, and NMF; (iii) designing a robotic-based method to control the sound, in terms of Auditory Scene Analysis, in a comparatively small space.

研究分野：音響情報処理

キーワード：音環境理解 音クラスタリング 機械学習 音環境デザイン

1. 研究開始当初の背景

音に関する識別技術として過去に最も研究されてきたと思われる自動音声認識の分野は、数理的枠組みやコンピュータ能力が飛躍的に進歩した結果、丁寧な発声であれば実用に供し得るレベルの性能を達成している。現在では、(A)会話のような自由で複雑な発話モード、(B)雑音や反響が含まれるような環境への頑健性、(C)様々な状況で利用可能なアプリケーション実装などに関する検討が、国内外の研究界において盛んに進められている。音声以外の音についても、楽器音の音響的分析に基づく音楽構造の認識、環境音の識別やモデル化、環境雑音や騒音の評価など、様々な見地からの研究が行われている。また、聴覚情景分析として知られるように、人間の聴覚系の分析特性に着目して音源や音の属性を解析する研究も多く行われるようになってきている。

我々人間は、たとえば TV やラジオのスイッチを入れ、そこから流れる音を耳にした途端、放送内容(ニュースか、スポーツ中継か、音楽番組か)をほぼ瞬時に認識できる。放送音声のような「人が聞くことを前提としている音」でなくとも、たとえば屋外や室内で耳に入る様々な音(自動車や電車の走行音、動物や鳥の鳴き声、風や水の音、ドアのロック音など)に対して、我々は時間をかけずにその音源を知ることができる。このことは、言うまでもなく我々の脳内に学習により蓄積された「様々な音に関する知識」があるからに他ならないが(その証拠に、聞いたことのない音の音源を推測することは難しい)、この「知識」は、構造的に明確に分類され定義されているわけではない。

研究代表者は、約 20 年前、博士学位論文として音声認識のための頑健な音響モデルに関する研究をまとめた後、主に音声認識における頑健性の問題(上述の(B))に強い興味を持ち、雑音環境下での音声認識技術に関する研究を進めてきた。その過程で、雑音だけでなく、構造的な特徴を持った音響データ(たとえば音楽や他人の音声)が付加された状況下での音声認識の重要性に気付き、音声とそれ以外の音を効果的に分離する技術、特に音声と楽音(歌声を含む)の識別に関する研究を実施してきた。本研究テーマは、それら一連の研究を発展させ、より一般的な音カテゴリーのクラスタリング手法について包括的に検討しようとして位置づけたものである。

申請の時点で上述の問題に対する決定的な原理や手法が見出されていないことを重視し、関連研究の成果を検証しつつ新たな方法を提案する必要があると考えた。また、関連研究では考慮されることの少ない音カテゴリーが重畳した場合や未知の音カテゴリーを含む場合についても、本研究計画では取り組もうとした。また、音環境理解やロボティクス等への応用も試みることにした。

2. 研究の目的

本研究で達成したいと考えた大きな目標は、我々が日常的に耳にする様々な音に関して、「日常的」の定義も含めその合理的な力カテゴリーを定義し、任意の音カテゴリーを分類する技術を検討すると共に、その応用方法(音環境理解、ロボットへの実装など)を提案することである。具体的には、音声・楽音・環境音・雑音など各種の音カテゴリーが混在した音響データ(たとえばラジオ音声)に対して、それぞれの音カテゴリーの時間区間を自動的に識別し、その結果をメディア検索などへの応用のためにメタデータ(音データ自体に付加的に与える情報)として付与することである。この際、たとえば音声と楽音が重畳しているような場合についても複数のカテゴリーが重畳した区間として出力できること、また、未知の音源については未知のカテゴリーとして出力できることを目指した。

申請の段階で、音カテゴリーの分類タスクに対する決定的な手法が見出されていないことを鑑みて、(1)音カテゴリーの分類や識別に有用な音響特徴量の検討、(2)クラスタリング手法および識別手法(アルゴリズム)の検討、(3)放送音源や生活環境音などの実データに対する検証、の3つの論点について明らかにしようと考えた。また、適切な分類・識別を可能とした上で、この技術の応用として音環境理解やロボティクス(特にコミュニケーション技術により人間と接する部分)などへの適用についても検討することとした。

3. 研究の方法

本研究では、音カテゴリーのクラスタリングに関する検討を行う。そのための研究計画として、まず公平な実験を実施するためのデータ収集と音カテゴリーを表現する合理的方法に関する検討を行う。次に、分類や識別に有効な音響特徴量の検討および分類・識別のためのアルゴリズムの検討を行う。最後に、提案する手法の実環境での検証と、音環境理解やロボティクスへの応用に関する検討を行う。

(1)音カテゴリークラスタリング・音カテゴリー識別実験のためのデータ収集

実験に用いるデータとして、既存の音声・楽音・環境音・雑音のデータベースを収集する。これには世界のいくつかの研究機関で集められているデータベースを有効に活用し、データ収録における効率化および評価基準の共通化を図る。すでに保有しているデータベースが数種類あるのでそれらも利用する。申請時に購入を計画したデータベースとしては、RWC 研究用音楽データベース、ATR 環境音データベースなどがある。また、実放送音源として TV やラジオの音声を収録し、必要ならば利用許諾を受ける。さらに、日常的な生活環境音の収録(デジタル機器による録音)を計画的に行う。

(2)音カテゴリーの合理的な表現方法に関する検討

分類方法を定義するための予備的実験として、各種音データを被験者に聴取させ、合理的な音カテゴリーの表現方法について検討する。その結果および関連研究の調査により得られる知見をもとに、データの一部に対して目視で情報内容に関するラベル付けを行う。ラベル付けは、音カテゴリーの種類のみならず、音声ならば男声と女声の区別、楽音ならばジャンル(クラシック、ポピュラーなど)の区別のように細分化・階層化されたカテゴリや、収録環境、複数音源の重畳なども考慮して行う。

(3)音カテゴリークラスタリング・音カテゴリー識別に有効な音響特徴量の検討

様々な音カテゴリーを分類・識別するのに有効な音響特徴量として、スペクトル形状に基づく静的特徴と比較的短時間の時間変化を表現するための動的特徴、さらに長時間にわたる変化が表現できるような大域的特徴について検討する。

音声認識の場合、静的特徴により音韻性の違いが、短時間動的特徴により過渡的な変化(音素から音素への渡り部分の特徴など)が表現され、音響モデルに用いられるが、扱う音の対象が広がった場合、たとえば楽音においては音高に対応する特徴的な周波数パターンが、また機械音や一部の雑音においては音源に起因する周波数帯域の偏りや集中が見られる。各々の音源の特徴を熟考しつつ、音響信号中の支配的な成分を検出・追跡することにより、各種の音カテゴリーが重畳した場合にも有用な特徴量を検討・提案する。同時に、音声認識で用いられるような既存の音響特徴量との比較も行う。

(4)音カテゴリークラスタリング・音カテゴリー識別のためのアルゴリズムの検討

分類・識別アルゴリズムとして、混合ガウス分布モデル(GMM)や隠れマルコフモデル(HMM)、ディープニューラルネットワーク(DNN)などの統計的モデルの利用を検討することとした。この際、音の種類による音響的特徴の分散や多様性を考慮した最適なモデル化手法について十分な考察を行う。最初に比較的簡単なタスク(たとえば音声+楽音の2カテゴリ識別問題など)を設定し、条件を変えた分類・識別実験を行う。また、モデルのカテゴリーを細分化・階層化(音楽における楽器やジャンルによるモデルの区別など)した実験も併せて行う。

(5)音カテゴリークラスタリング基準の設計と提案

検討した音響特徴量および分類・識別アルゴリズムを種々の音データに適用して得られる結果より、適切と思われる音カテゴリーの

分類基準について提案する。1つのデータに対して複数あるいは階層的な分類基準を利用することも考慮する。提案した分類基準の有用性についても比較・評価する。

(6)実音源に対する提案法の検証

TVやラジオの音源から収集した放送音源データに対して、提案する音響特徴量および識別アルゴリズムを適用し、自動インデキシング実験およびその精度の評価を行う。また、インデキシングを行う時間単位(何秒間の音データに対して識別するか)を変えた実験を行い、人間による聴取実験との比較も行う。さらに、収録した生活環境音など未知の音カテゴリーを含む場合についての実験も行う。

(7)音環境理解、コミュニケーションロボティクスへの応用可能性の検討

提案した音カテゴリー分類・識別手法を応用し、多様なメディア情報から目的の音カテゴリーを検索する技術、広義の音環境理解への応用について検討する。具体的には、音情報検索タスクを設定し、大量のデータ中から目的の音カテゴリーを含む時区間を見つけるシステムを設計・実装する。この際、「女声」「ピアノの音」「電車が走る音」などの比較的単純な場合から、「電車の中での会話」「弦楽器と打楽器が含まれる音楽」「機械音が重畳したサイレン音」など複雑な場合を設定し、どの程度の精度で検索が可能かを検証する。

また、より具体的な応用場面として、人間とのコミュニケーションを主たる用途としたロボット(コミュニケーションロボット)への適用について、その方法と組み合わせるべき他のモード(視覚や触覚)やインタフェースについて検討する。

4. 研究成果

上に述べた当初の研究目的および研究方法にしたがって3年間にわたって研究を進めた結果、主に以下に記す成果を得た。

(1)音カテゴリークラスタリング・音カテゴリー識別実験のためのデータ収集

識別実験用データとして、本研究開始前から保有していた音声・楽音・環境音・雑音のデータを整理すると共に、様々な研究機関で作成されたデータベースを入手した。また、いくつかの生活環境音および用途を研究に限定した放送音源を収録した。次に、音の分類方法を定義する準備として、各種音データを複数の被験者に聴取させ、合理的な音カテゴリーの表現方法について検討した。その結果をもとに、データの一部に目視で情報内容に関するラベル付けを行なった。

(2)音カテゴリークラスタリング・音カテゴリー識別に有効な音響特徴量の検討

様々な音カテゴリーを有機的に分類・識別するための音響特徴量として、パワースペクト

ルの形状に基づく静的特徴と短時間の時間変化を表現できる動的特徴、さらに長時間に渡る変化が表現できる大域的特徴について検討した。各々の音源の特徴を考慮しつつ、音響信号中の支配的な成分を検出・追跡することにより、各種の音カテゴリーが重畳した場合にも有用な特徴量を検討・提案した。

また、零交差数およびスペクトルエントロピーと呼ばれる特徴量を用いて、比較的簡単な音クラスタリング実験を行い、非音声と音声、発音の異なる音源が分離できることを確認した。

(3)音カテゴリークラスタリング・音カテゴリー識別のためのアルゴリズムの検討

アルゴリズムとして、混合ガウス分布(GMM)モデルや隠れマルコフモデル(HMM)、ディープニューラルネットワーク(DNN)モデルなどの統計的モデルの利用法を検討した。この際、音の種類による音響的特徴の多様性を考慮した最適なモデル化法についても考察した。最初に比較的簡単なタスクを設定し、条件を変えつつ実験を行った。また、モデルカテゴリーを細分化・階層化した実験も併せて行った。これに関連して、非負値行列因子分解(NMF)に基づく特徴量に関する検討も実施し、識別器にSVMを用いた実験の結果、最も良い条件下で80%以上の音カテゴリー識別率を得た。タスクに様々なバリエーションが存在し得るため、当初計画時点での目標を達成できたかを判断するのは難しいが、具体的な音カテゴリーとして「車の走行音・音声・電車の走行音・水の流れる音」という条件を設定し、それぞれ300個の学習データを与えた条件下では一定レベルの成果に到達できたと考えている。この研究成果は、第16回情報科学技術フォーラム(2017年9月開催)において報告する予定である。

(4)音カテゴリークラスタリング基準の設計と提案

種々の音響特徴量および識別アルゴリズムを実際の音データに適用して得られた結果を用いて、最も適切と思われる音カテゴリーのクラスタリング基準について検討した。また、1つのデータに対して複数あるいは階層的な分類基準を利用することも検討した。提案した分類基準の有用性について比較・評価実験を行った。

(5)音環境理解への応用

当初計画時点では具体的な目標設定をしていなかったが、研究の進行により得た基礎的な実験結果および知見により、音環境理解の一環として小規模空間で発生する様々な環境音を制御する際にロボティクス技術を応用する手法について検討を行なった。具体的には、小型のロボットに調音パネル(残響可変装置)を取り付けて自律的な走行をさせることで、会議室等の空間における音環境を

制御するための実験的検討を実施した。この成果は、日本音響学会研究発表会(2015年3月, 2016年3月, 2016年9月)や日米音響学会ジョイントミーティング(2016年11月)において発表し、本研究が目指した音環境理解への応用に加えて音響学とロボティクスの融合という観点からも、国内外の研究界に対してそれなりのインパクトを与えたと考える。

(6)ロボティクスへの応用

この項目も、当初計画時点では具体的な目標設定をしていなかったが、本研究で得られた知見の応用として、コミュニケーションRT(ロボットシステム)におけるインタフェースに音の識別機構を実装した。具体的には、ロボットの内部状態表出部(出力部分)として眼球インタフェースを考え、その駆動を周囲の環境音に基づいて行った。この研究成果は、インタラクションシンポジウム(2016年3月)およびIEEE RO-MAN(2016年8月)において発表し、一定の評価を得た。

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3件)

Masashi Uehara, Shuhei Kawai, Shigeki Okawa, Manabu Fukushima, Autonomous optimal location of articulation materials using robotics, Proc. ASA/ASJ Joint Meeting, 3pAAc2, 2016/11/28-12/2, Honolulu, Hawaii, USA, 査読無し

Shuhei Kawai, Masashi Uehara, Shigeki Okawa, Manabu Fukushima, Autonomous mobile robot to improve sound environment for speech conversation, Proc. ASA/ASJ Joint Meeting, 3pAAc3, 2016/11/28-12/2, Honolulu, Hawaii, USA, 査読無し

Masaki Kato, Shigeki Okawa, "i-eye" - a wearable eyeball interface, Proc. IEEE International Symposium on Robot and Human Interactive Communication, TuIVS.30, 2016/8/26-31, New York, USA, 査読有り

[学会発表](計7件)

秋庭裕, 大川茂樹, 音ライフログに向けたNMFに基づく特徴量による環境音識別, 第16回情報科学技術フォーラム, 2017/9/12~14, 東京大学(東京都文京区)(発表確定)

上原正志, 河合修平, 大川茂樹, 福島学, ロボティクスを用いた調音材の最適配置の決定手法, 日本音響学会秋季研究発表会, 2016/9/15, 富山大学(富山県・富山市)

河合修平, 上原正志, 大川茂樹, 福島学,

会話に適した音環境を作る自律移動ロボット, 日本音響学会秋季研究発表会, 2016/9/15, 富山大学 (富山県・富山市)

上原正志, 河合修平, 大川茂樹, 残響可変装置とロボティクスの組み合わせによる小規模音場へのアプローチ, 日本音響学会春季研究発表会, 2016/3/10, 桐蔭横浜大学 (神奈川県・横浜市)

加藤真規, 大川茂樹, 装着型眼球インタフェース「i-eye」, インタラクションシンポジウム, 2016/3/3, 科学技術館 (東京都・千代田区)

上原正志, 大川茂樹, 小規模空間のための残響可変装置の開発, 日本音響学会春季研究発表会, 2015/3/16, 中央大学 (東京都・文京区)

加藤真規, 大川茂樹, スマートフォン装着型眼球インタフェースの開発, 日本感性工学会春季大会, 2015/3/28-29, 京都女子大学 (京都府・京都市)

〔図書〕(計1件)

羽田陽一, 大川茂樹, 木谷俊介 (監修), コロナ社, 音響学入門ペディア, 2017, 計 200 ページ, 第 0 章の執筆および全体の監修を担当

6. 研究組織

(1) 研究代表者

大川 茂樹 (OKAWA SHIGEKI)

千葉工業大学・先進工学部・教授

研究者番号: 4 0 3 0 6 3 9 5