

機関番号：32503

研究種目：基盤研究(C)

研究期間：2008～2010

課題番号：20500162

研究課題名(和文) 音響事象カテゴリの分類および識別に関する基礎的研究

研究課題名(英文) A Fundamental Study on Classification and Recognition of Acoustic Event Categories

研究代表者

大川 茂樹 (OKAWA SHIGEKI)

千葉工業大学・工学部・教授

研究者番号：40306395

研究成果の概要(和文)：音声や音楽や環境音などの音響事象カテゴリを効率的に分類し識別するための基礎的研究として、様々な音からの特徴抽出法および特徴量の検討、識別アルゴリズムの開発と検討、得られた知見に基づいたメディア検索やロボティクス、ヒューマンインタフェースへの応用法について研究を実施した。

研究成果の概要(英文)：As a fundamental study to effectively classify and recognize acoustic event categories such as speech, music, and environmental sounds, we have studied feature extraction methods from various sound sources, the development of classification algorithm, and those applications to media search, robotics, and human interface.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,000,000	300,000	1,300,000
2010年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：音声言語処理，ヒューマンインタフェース，コミュニケーションロボティクス

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：パターン認識，音響事象カテゴリ，音声音楽識別，音声認識，信号処理

1. 研究開始当初の背景

音に関する識別技術として最も盛んに研究されてきたと思われる自動音声認識の分野は、数理的枠組みの進歩や近年のコンピュータ能力の向上に伴って理論や技術が飛躍的に発展した結果、丁寧な発声であれば実用に供し得る程度の性能を達成している。現在では、(1)会話音声のような複雑な発話モード、(2)背景に雑音や反響が含まれるような環境への頑健性、(3)音声認識を利用したアプリケーションの実装などに関する検討が、研究界において盛んに進められている。

音声以外の音に関しても、楽器音の音響的分析に基づく音楽構造の認識や、環境音の識

別とモデル化、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分析する研究も近年多く行われるようになってきている。

申請者は、1996年に、博士学位論文におい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(2))に興味を持ち、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。本申請は、それら一連の研究を発展させ、より一般的な音響事象カテゴリーの分類・識別について包括的に検討するものと位置づけている。

2. 研究の目的

本研究は、音声・楽音・環境音・雑音など我々の身の回りに日常的に存在する様々な「音響事象」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分類する方法と、任意の音響事象のカテゴリー識別を行う技術を検討することを全体構想（総合的な目標）とする。

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、どのような内容の放送が行われているのか（ニュースか、スポーツ中継か、音楽番組か、など）をほぼ瞬時に認識できる。放送音声のように「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、ドアのノック音など）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。

従来、音声や楽音など、それぞれの事象カテゴリーを細分化した構造（たとえば音声における音韻性や韻律的特徴、楽音における音高や音色）に関しては非常に多くの研究が行われているが、上に述べたような「日常的に耳に入る様々な音」を包括して扱う研究は少ないと言える。

そこで本研究では、音響事象の分類基準に関する基礎的な知見を得るとともに、カテゴリーの合理的な識別手法およびメディア検索への応用法について検討することを具体的な目的とする。

3. 研究の方法

本研究では、音響事象カテゴリーの分類および識別に関する検討を行った。そのための準備として、まず公平な実験を実施するためのデータ収集と音響事象カテゴリーを表現する合理的な方法に関する検討を行った。次に、分類や識別に有効な音響的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った。最後に、提案する手法の実環境での検証と、メディア情報検索、ロボティクス、ヒューマンインタフェースへの応用に関する

検討を行った。以下にその具体的な方法を述べる。

(1) 音響事象カテゴリー分類・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した。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った。すでに保有しているデータベースが数種類あったのでそれらも利用した。購入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である。また、実放送音源として TV やラジオの音声を収録し、必要に応じて利用許諾を受けた。さらに、日常生活環境音の収録（デジタル機器による録音）を計画的に行った。

(2) 音響事象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の実験として、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した。その結果および関連研究の調査により得られた知見をもとに、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った。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ーや、収録環境、複数音源の重量なども考慮して行った。

(3) 音響事象カテゴリー分類・識別に有効な音響的特徴量の検討

様々な音カテゴリーを分類・識別するのに有効な音響的特徴量として、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した。

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・提案した。同時に、音声認識で用いられるような既存の音響的特徴量との比較も行った。

特に、申請者らが主に雑音下での音声認識のために開発してきた音声のFM（周波数変調）特徴およびそれをを用いた認識アルゴリズムについて、本研究での有用な特徴量として利用できないかについても積極的に検討を行った。

(4) 音カテゴリ分類・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)などの統計的モデルの利用を検討した。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った。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、条件を変えた識別実験を行った。また、モデルのカテゴリを細分化・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った。

(5) 音カテゴリ分類基準の設計と提案

検討した音響的特徴量および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の分類基準について提案した。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考慮した。提案した分類基準の有用性についても比較・評価した。

(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、提案する音響的特徴量および識別アルゴリズムを適用し、自動インデクシング実験およびその精度の評価を行った。また、インデク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った。さらに、収録した生活環境音など未知の音カテゴリを含む場合についての実験も行った。

(7) 得られた知見の応用

提案した音カテゴリ識別手法の応用について検討した。まず、ロボティクスへの応用として、包囲型マイクロホンアレイ内での音響事象を分析・識別し、可視化することにより、音カテゴリ識別結果をロボット動作の周辺動作として利用することを試みた。次に、ヒューマンインタフェースへの応用として、高齢者の音声から健康状態を推定するアルゴリズムを考案し、評価実験を行った。

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

めて来た結果、以下に記す成果を得た。

(1) 音響事象カテゴリ分類・識別実験のためのデータ収集

音響事象カテゴリ識別実験用データとして、既存の音声・楽音・環境音・雑音のデータを収集・整理した。様々な研究機関で集められているデータベースを有効活用し、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った。

(2) 音響事象カテゴリ分類・識別に有効な音響的特徴量の検討

種々の音カテゴリを分類・識別するのに有効な音響的特徴量として、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した。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、各種の音カテゴリが重畳した場合にも有用な特徴量を検討・提案した。これに関連して、楽音中の音声情報を強調する枠組みとして、正弦波トラジェクトリを用いたアルゴリズムについて検証し、日本音響学会誌に論文を発表した。この成果は、音響事象カテゴリの分類に対する新たな特徴量の提案という点で国内外の研究に対してそれなりのインパクトを与えたと考える。

(3) 音カテゴリ分類・識別のためのアルゴリズムの検討

音響事象カテゴリの分類・識別のための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)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、モデルカテゴリを細分化・階層化した実験も併せて行った。

(4) 音カテゴリ分類基準の設計と提案

音響的特徴量および識別アルゴリズムを種々の音データに適用して得られた結果を用いて、適切と思われる音カテゴリの分類基準について検討した。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考慮した。提案した分類基準の有用性について比較・評価した。

(5) 音響事象カテゴリの識別実験

音響事象カテゴリの識別に関して、音声・音楽・環境音データに対して提案したいくつかの特徴量を用いた実験を行った。中でもBlock Cepstrum Fluxと呼ばれる特徴量とSVMを組み合わせた識別実験で良好な性能を得

た。これに関連して、日本音響学会研究発表会において研究成果を発表した。

(6) 包囲型マイクロホンアレイを用いた音響マップの生成

本研究で得られた成果のロボティクスへの応用を考慮して、室内空間に設置した包囲型マイクロホンアレイ内での音響事象を分析・識別し、可視化することにより、音カテゴリ識別結果をロボット動作の周辺情報として利用することを試みた。

(7) 音響事象カテゴリ分類のための長時間特徴量の検討

分類・識別のための特徴量として、音データの時間的および周波数的特徴を組み合わせた3次元の特徴量(長時間特徴量)を提案し、これを混合ガウス分布モデル(GMM)や隠れマルコフモデル(HMM)などの統計的モデルに適用し、カテゴリ分類・識別実験を行った。

この結果は、Speech Communication 誌に発表し、国内外の類似研究にインパクトを与えたものとする。

(8) 高齢者音声からの健康状態の推定

本研究で得られた知見の応用として、高齢者の音声から健康状態を推定するアルゴリズムを考案し、評価実験を行った。この成果は、コミュニケーションRT(ロボット技術)システムおよびインタフェースに実装し、約1か月間に渡り被験者の協力のもとでのモニター試験による評価を行った。この実験結果については、日本音響学会研究発表会や電子情報通信学会総合大会において成果を発表した。

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)

①Yotaro Kubo, Shigeki Okawa, Akira Kurematsu, Katsuhiko Shirai, Temporal AM-FM combination for robust speech recognition, Speech Communication, 査読有, 53, 2011, 716-725

②大川茂樹, 音響事象カテゴリの分類および識別に関する基礎的研究, 千葉工業大学プロジェクト研究年報, 査読無, 6, 2009, 246-248

③谷口徹, 大川茂樹, 白井克彦, 正弦波トラジェクトリに基づく楽器音中の音声強調, 日本音響学会誌, 査読有, 65(5), 2009, 251-261

④Yotaro Kubo, Shigeki Okawa, Akira Kurematsu, Katsuhiko Shirai, Recognizing reverberant speech based on amplitude and frequency modulation, IEICE Transactions on Information and Systems, 査読有, E91-D-3, 2008, 448-456

[学会発表] (計5件)

①藤井敦啓, コミュニケーション RT による高齢者の在宅健康管理・支援システムー音声インタフェースの実装と評価ー, 電子情報通信学会総合大会, 2011/3/14, 東京都市大学(東京都)

②藤井敦啓, 在宅健康管理 RT システムにおける高齢者音声からの健康状態の推定, 日本音響学会研究発表会, 2011/3/10, 早稲田大学理工学部(東京都)

③戸田健吾, コミュニケーション RT による高齢者の在宅健康管理・支援システムーシステム概要とプロトタイプの開発ー, 日本機械学会ロボティクス・メカトロニクス講演会, 2010/6/15, 旭川大雪アリーナ(北海道)

④上田みき, 時間・周波数パターンを利用した音響事象カテゴリの表現, 日本音響学会秋季研究発表会, 2009/9/17, 日本大学郡山キャンパス(福島県)

⑤Yotaro Kubo, A comparative study on AM and FM features, International Conference on Speech Communication and Technology, 2008/9/24, Brisbane, Australia

[図書] (計1件)

白井克彦, 菅田雅彰, 大川茂樹他, コロナ社, 音声言語処理の潮流, 2010, 135-151

[その他]

ホームページ等

<http://www.oklab.it-chiba.ac.jp/>

6. 研究組織

(1) 研究代表者

大川 茂樹 (OKAWA SHIGEKI)
千葉工業大学・工学部・教授
研究者番号: 40306395

(2) 研究分担者

なし

(3) 連携研究者

なし