

## 科学研究費助成事業 研究成果報告書

平成 30 年 6 月 26 日現在

機関番号：34407

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00255

研究課題名(和文)日常生活音シーン理解をめざした屋外集音とその応用

研究課題名(英文)Outdoor sound collection for auditory scene analysis in a daily life and its applications

研究代表者

高橋 徹 (Takahashi, Toru)

大阪産業大学・デザイン工学部・准教授

研究者番号：30419494

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：日常生活音を理解するための音声信号処理について研究した。まず日常生活環境音中の個々の音源の方向を検知し、音源を分離するための要素技術を開発した。日常音の音源は、多岐に渡るため、あらかじめ録音レベルを最適に設定することが難しい。そこで、従来の機器で録音がどの程度困難であるかを調査した。更に音源方向を検出する課題と、背景に音楽が流れている環境における音声信号強調手法について検討した。その他、日常音環境では、環境中の音源数を事前に知ることができないため、マイクロホン数以上の音源が存在する環境での音源分離課題について、特に音声に特化した目的信号強調手法とそのパラメータ学習手法を開発した。

研究成果の概要(英文)：Sound signal processing to understand daily life sounds is investigated in this project. Sound source localization, sound source separation, and sound source emphasis methods are developed. Since many kinds of sounds are radiated in daily environment, it is hard to optimize recording level in advance. I investigated how difficult for optimizing the level with conventional recording devices. And I also investigated the task of detecting the sound source direction and the speech signal enhancement method in the environment where the music flows in the background. In the everyday sound environment, since the number of sound sources in the environment cannot be known in advance, the sound source separation task in environments where there are sound sources more than the number of microphones, especially the target signal emphasis method specialized for speech and its parameter learning method was developed.

研究分野：音声信号処理, ロボット聴覚

キーワード：音環境理解 ロボット聴覚 音源定位 音源分離 音声強調 環境音

## 1. 研究開始当初の背景

日常生活音のパターンを抽出し、パターンの意味を理解するためには、実環境で安定動作する音処理技術が必要であった。具体的には、(1) 実環境に存在する混合音中から、目的となる音を捉える技術、(2) 捉えた音からパターンを抽出する技術、(3) 抽出される表現を解析する技術の3つが必要であった。

従来これら3つの技術は、個別に研究されてきたため、これらの技術の単純な足し合わせでは、日常生活音のパターン抽出や意味解析は困難であった。何故ならば、個々の技術が動作する前提条件が異なっているため、要素技術を直列に繋いだシステムを構築しても、いずれかの部分で条件の不整合が起こるためである。

実環境に存在する混合音中から目的音を捉える技術と、目的音のパターンを抽出する技術、および、抽出された表現を解析する技術を一体的に開発し、実験室外の実用的な日常場面での音処理手法を確立することが重要な課題であった。

## 2. 研究の目的

本研究の目的は、混合音を許容する環境音認識技術を用いた、日常生活音のパターン抽出と意味解析手法の確立で、音のシーン理解と呼ばれる分野に該当する。音を記号化し意味付けし理解することを最終的な目標としている。屋外を含む日常生活の実音声を対象とすると、それらの音は、時間的・空間的・周波数的に重なり合った混合音である。実験室環境と異なり、音の大きさや周囲に存在する音源の数、音源の種類をあらかじめ想定することができない。従って、研究の第一歩は、これらの音の分離と識別、そこから音の認識と理解に踏み込む必要がある。

音を分離し何らかの表現に落とし込むための收音技術、音源定位技術、音源分離技術を統合的に扱う音処理の研究開発を目的とした。本研究により音を音源要素ごとに分離し、一定のパラメータ表現に落とし込むことが可能となった。

## 3. 研究の方法

日常生活音の收音方法や日常生活音が複数混合した環境での音の取り扱い、および、既存音源分離技術を用いた後処理としての分離歪を抑制する手法の研究を行った。具体的には以下の3種の研究を推進した。

- (1) 屋外での日常音収集における従来録音との差異を明らかにする研究と日常音收音のためのアルゴリズム最適化
- (2) 混合音における音源定位・音源分離・音源検出に関する研究

## (3) 音環境理解技術を用いた応用システムの研究開発

## 4. 研究成果

### 日常生活音の録音条件に関する研究

整備された録音環境では、音声録音に用いる機器の収録条件にサンプリングレート 16000Hz、量子化ビット 16 ビットを用いるのが一般的である。一方、日常生活音の収録環境では、音源の数・音源までの距離・音源の出力エネルギーを予め想定することが困難である。従って、適切な入力レベルを固定的に設定することは不可能である。最大音圧レベルを見積り、最大音圧レベルが、記録可能範囲を越えない設定にせざるを得ない。この設定で目的音と非目的音の混合音を録音すると、非目的音のエネルギーに依存して、目的音の実効的量子化ビット数が低下する問題に着目した。

目的音と非目的音のエネルギー差が 40dB 未満で、この問題が顕著に表れることを示した。実行的量子化ビット数の低下による目的信号の歪みが記録されることが確認できた。この歪は、回復困難であり音声認識も難しいため、従来の録音条件では 16bit 量子化で十分と考えられてきたが、日常生活音の収録には 16bit 以上の量子化が有効であることが示唆された。

現在音声録音機器は、量子化ビット数 24bit に対応するものも増え、専用の機材を用いれば 32bit 精度で収録も可能になってきた。これら機器の性能は、1ch 音声の録音には、オーバースペックと考えられることが多いが、音環境理解には、可能な限り高精度の録音パラメータを用いることが望ましいことを確認できた。

### MUSIC 法のパラメータ最適化

音源到来方向推定に MUSIC 法を用いてきた。この手法は、時間波形信号に短時間フーリエ変換を適用し、マイクロホン素子間の観測値の相関から方向を推定する。この時、観測値の相関の期待値が用いられる。相関の期待値は、直近のフーリエスペクトルから推定される。このため、分析フレーム長や直近何フレームから相関の期待値を推定するかによって、音源到来方向推定精度に影響を与える。音源が音声である場合について、分析フレーム長を評価した結果、音声スペクトル分析に用いる 20~40ms より短い分析フレーム長が適当であることを実験的に確認した。

## 音源方向推定と音源検出の統合

複数音源を同時に観測する音環境理解において、従来の目的音源の分離は、音源方向推定と、推定方向音源の收音と、收音結果から音声区間を検出する音源検出を組み合わせて実現されていた。

音源到来方向推定と音源検出を同時に行う手法を検討した。MUSIC法は、マイクロホンアレイによる観測信号を多次元空間で表現する手法である。信号と雑音の空間を異なる次元に分離する部分空間法と呼ばれる手法である。信号を表す部分空間表現は、音源信号のスペクトログラムと類似した表現が得られることに着目し、信号空間上で音声区間検出を行う手法を開発した。同時に音源定位・音源分離・分離音声の音声区間検出をする従来手法と併用し、音源到来方向推定と音声区間検出を一体的に扱う枠組を提案できた。

## 日常生活の発話の重なるの調査

日常生活音中の音声発話は、話者が1名であるとは限らない。複数話者が同時に発話する割り込み発話が起こる。通常の会話において、発話が時間的重なりを生じる頻度についての調査例がほとんどなかった。原因の一つに、一般の日常生活音を調査に用いることはプライバシーの問題から困難であったためである。

日常生活中における会話の代替として、テレビ番組の中からバラエティー番組の複数出演者が会話するシーンを収集した。これらの中で同時に発話する頻度を調査した。笑い声や相槌、感嘆表現などが重なる傾向が強いことを確認できた。更に、出演者の発話一文全体に、他の話者の発話が一切重ならないケースは、30分当たり数文に限られることが明らかになった。これは、同時発話が頻繁に起きていることを示している。

この調査結果から、複数発話者が同時に発話することを考慮しなければ、日常生活音の認識が、困難であることが窺える。同時発話認識の重要性を迫ってきた。

次にプライバシーの問題を避けつつ自然な複数会話音声を効率的に収集し、日常音声に相当する評価データを集める方法を構築することが必要となった。アナログカードゲームに着目し、プレイ中のプレイヤー会話を収集する方法を提案した。

人狼をプレイする4名の発話を収集し分析した。人狼は、プレイヤーが人チームと狼チームに分かれ、他人の役割(人または狼)を推理するゲームである。推理するために雑談することから、自然な会話を得られやすい特徴を利用した。プ

レイ中の会話の語彙が限られる問題があったものの、プレイとプレイの間の雑談がより日常生活の自然な会話に近い傾向があった。これらの結果から、アナログゲームのプレイ中音声を日常生活の会話として収集することが、疑似日常会話音声コーパスの作成に役立てられる可能性を示すことができた。

## 音声と音楽の混合音の認識

日常生活音の中で音声と音楽が混合するシーンは多い。ショッピングセンター・デパート・レストランなど店内の背景に音楽が流れていることが多い。音声と音楽が混合する音環境での音声認識には、混合音から音声を認識する技術が必要である。音声と音楽の混合音では、音声を認識することと、音楽を認識することは、完全な対称性を持たない。音声の認識には、音楽を抑圧することが必要である。一方、音楽を認識することは、音声の抑圧が必要になる。音声は、任意の発話が可能であるのに対して、音楽は、音楽のある時刻の断片が得られると、その近傍では、当該楽曲の近傍が観測されている可能性が高く、音声に比べて近傍音響特徴の推定が容易である。推定の難易度が音声と音楽では非対称である特徴がある。音声と音楽の混合音では、音楽の音響特徴を先に推定し、スペクトル特徴を除去することで音声の特徴を強調する枠組みを提案した。

音楽はCDなどデータが再生されることが多い。従って、どのCDの何トラックのどの時刻が再生されているか認識できれば、CDデータベースから現在付近の音楽波形を知ることができる。データベースの情報を利用して、観測信号から音楽特徴を除去し音声強調が可能である。

高精度な楽曲認識を実現するため、雑音に耐性のあるバイナリクロマスペクトルを用い楽曲認識実験を行った。音源との距離や、音声と音楽音の相対レベルの違いに頑健に楽曲を高精度に認識可能であることが明らかになった。楽曲を推定可能な他、楽曲の再生時刻を推定することも可能であることを確認した。

## 音声と音声の混合音の認識

複数のマイクロホンを使った音源分離手法が多数研究されている。これまでMUSIC法による音源方向推定とビームフォーミングを用いた音源分離手法、更に音源検出手法を直列に接続して音源分離を実現してきた。この手法の利点は、音声と音声の混合音の場合にも音源分離可能という特徴である。欠点は、

音環境中の音源数が、マイクロホンアレイを構成するマイクロホン素子数より少ないことが動作条件になっている点である。日常環境ではこの動作条件を保証することは困難である。

日常生活音の認識では、音源数を事前に行うことができないため、音源数に関する条件を緩和する必要があった。マイクロホンアレイを用いた音源分離手法やその他の音源分離手法を用いて音源分離しても、分離歪が残る。必要な技術は、他の音源分離手法を適用した後に得られる、1ch音声を対象とした信号処理技術であり、この音声信号に含まれる非目的音の抑圧である。これは同時に目的音声の強調でもある。

1ch音声強調かつ非目的音の抑圧を同時に実現する方法に非負値行列因子分解があった。当初、音声と音声の混合音に適用すると、音声スペクトルのピーク一つ一つが独立に基底化され、音声のスペクトル構造を保ったままの基底を獲得させることが困難であった。この基底は音声のスペクトルも非音声のスペクトルも表現可能となる。これは、通常非負値行列因子分解では、基底行列表現の自由が高過ぎることを意味した。

音声スペクトルだけを表現可能な基底行列を学習可能な新しいアルゴリズムを設計した。非負値行列因子分解に音声スペクトルモデルを組み込み、板倉斎藤ダイバージェンス基準で統一的に最適化する手法を開発した。非負値行列因子分解の中で板倉斎藤ダイバージェンス基準に基づく学習アルゴリズムに、音声スペクトルモデルを導入した。これは、音声スペクトルモデルに広く用いられる基準も板倉斎藤ダイバージェンスであり、親和性が高いためである。

特に、スペクトルモデルとして、音声分析や音声合成で広く用いられる全極モデルと指数スペクトルモデルを非負値行列因子分解に組み込むことができた。この拡張により、音声スペクトル包絡形状を直接的に基底行列に学習可能となった点が、音声強調や音源分離に重要である。

音声と音声の混合音であっても、大人と子供などスペクトル傾斜が異なる2音声を分離できる可能性がある。また、基底行列に、音声スペクトル包絡のみを学習可能なため、基底行列の学習段階で、狭帯域雑音を抑圧できる。音声スペクトルの線形結合で表現不可能な雑音は、アクティベーション行列の計算段階で、選択的に抑圧される仕組みになっているためである。

音環境理解技術の応用例

日常生活音処理の応用例として、音源到来方向を可視化する方法を検討した。デジタルビデオカメラで撮影した画像上に音源方向を提示するには、音源到来方向を画像上にマッピングする必要がある。カメラの画角内にスピーカーを設置し、測定信号を用いて音源方向を推定する方法を提案した。方向の推定値と画像内のスピーカー部分の画素位置を対応づけることで、音源到来方向を画像中に表示可能なことを示した。

音楽の生演奏音を音環境理解する応用について検討した。音源到来方向推定や、音源分離手法を生演奏音に適用すると、新しい問題を考えることができる。オーケストラを指揮者位置で聴取すると楽器の種類毎に異なる方向から音が到来する。演奏者位置で演奏音を聴取すると、楽器の発音体の位置の違いを区別することも可能である。例えば、一つの楽器音が、複数音源を持つ構造から奏でられることがある。音源と収音位置の関係により、音源をマクロに1つと扱うか、ミクロに個別に扱うかという問題を提起した。

## 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 1 件)

高橋 徹, 赤塚 俊洋, "テレビ番組音声の自発的発話音声コーパスとしての活用可能性について", 大阪産業大学論文集, 自然科学編 128, 研究論文, 査読あり, pp.21-30, 2017, ISSN 0287-1394.

[学会発表](計 10 件)

高橋 徹, "MUSIC 法に基づく音源到来方向推定のための分析帯域幅の検討", 日本音響学会 2016 年春季研究発表会, 桐蔭横浜大学, 1-P-18, pp.1-2, March 2016.

高橋 徹, "屋外利用を目指した音到来方向の可視化", 電子情報通信学会, 福祉情報工学研究会, 産業技術総合研究所, 臨海副都心センター, vol.116, no.360, pp.17-22, December 2016.

高橋 徹, "背景音として音楽音～背景音認識に向けて～", 日本音響学会, 音楽音響研究会資料, MA2017-05, pp.1-6, July 2017.

高橋 徹, "MUSIC 法に基づく DOA と VAD の検討", 日本音響学会 2017 年秋季研究発表会, 愛媛大学, 1-P-39, pp.1-2,

September 2017.

河野 秀祐, 高橋 徹, "人狼プレイ録音音声の NMF ベース任意話者音声強調", 電子情報通信学会技術報告, 音声研究会, vol.117, no.251, pp.7-12, October 2017.

高橋 徹, "定常雑音下での AD 変換における量子化ビット数と SNR の関係", 日本音響学会 2017 年春季研究発表会, 明治大学, 1-P-31, pp.1-2, March 2017.

乾 聡志, 高橋 徹, "実環境下におけるバイナリクロマスペクトルを用いた楽曲検索の検討", 日本音響学会 2017 年春季研究発表会, 明治大学, 3-4-6, pp.1-4, March 2017.

高橋 徹, 乾 聡志, 森田 貴志, "音楽検索における音源からの距離減衰とパターン一致度合いの関係", 日本音響学会 2017 年春季研究発表会, 明治大学, 3-4-8, pp.1-2, March 2017.

乾 聡志, 高橋 徹, "音楽データベースに基づく音楽と音声の混合音源からの演奏位置同定", 電子情報通信学会技術報告, 音声研究会, 沖縄産業支援センター, vol.116, no.477, pp.129-134, March 2017.

高橋 徹, "複数スペクトルモデルの混合表現可能な NMF に基づく音声スペクトル強調", 日本音響学会 2018 年春季研究発表会, 日本工業大学, 1-Q-45, pp.1-4, March 2018.

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 件)

名称：  
発明者：  
権利者：  
種類：

番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

#### 6. 研究組織

##### (1) 研究代表者

高橋 徹 (TAKAHASHI, Toru)  
大阪産業大学・デザイン工学部・准教授  
研究者番号：30419494

##### (2) 研究分担者

( )

研究者番号：

##### (3) 連携研究者

( )

研究者番号：

##### (4) 研究協力者

( )