

# 聴覚アウェアネス可視化のための 深度センサとマイクロフォンアレイを用いた 物体認識と音イベント検出

井山 貴裕<sup>1</sup> 杉山 治<sup>2</sup> 坂東 宜昭<sup>1</sup> 糸山 克寿<sup>1</sup> 吉井 和佳<sup>1</sup> 奥乃 博<sup>3</sup>

<sup>1</sup>京都大学 大学院情報学専攻 知能情報学専攻 <sup>2</sup>東京工業大学 先進理工学研究所

<sup>3</sup>早稲田大学 実体情報学プログラム

## 1. 序論

多様なメディアデータが増加する中、効率的なデータの解析・可視化は必要不可欠な技術である。特に音や画像に含まれる情報は環境を理解する上で重要な手がかりとなる。例えば、ロボットには画像を取得するためのカメラや音を取得するためのマイクロフォンが搭載されている。しかしながら、ロボットで取得した膨大なデータを人間によって解析することは時間がかかり負担が大きい。そのため、効率的なデータの解析・可視化のためには聴覚アウェアネスの提示が必要不可欠である。聴覚アウェアネスとは、音源の方向や位置、種類、状態変化など音源に対する総合的な気付きを意味する。そこで、ユーザの着目したい物体の情報や重要なシーンのみを検出しユーザに提示するシステムが必要である。

本稿では、聴覚アウェアネスの可視化システムの要素技術である音の発生区間検出を扱う。音の発生区間検出は音源分離や音のイベント検出のために必要な技術である。しかしながら、実環境における音の発生区間検出では音源数が不明なことが課題である。従来の音源定位手法は音源数を既知とするものが多く、音源数が不明な環境では音源定位を行うことは困難である。そこで、本稿ではレーザーレンジファインダで取得したポイントクラウドを用いてクラスタリングを行い環境中の各物体を推定し、推定した物体数やその方向情報を用いた音の発生区間検出手法について述べる(図1)。

## 2. 関連研究

レーザーレンジファインダとマイクロフォンアレイを用いたマルチモーダル音源定位手法では Even らが開発した手法がある [1]。この手法はレーザーレンジファインダを用いて空間のグリッドを作成し、音源のパワーの高いグリッドを音源の位置として求めている。提案手法では、空間のグリッドではなく、物体の形状を推定することで音源がどのような物体であるかに着目している。物体まで推定することは聴覚アウェアネス可視化システムの実現に必要である。

## 3. 提案手法

提案する音の発生区間検出手法は、各物体の方向情報と音響特徴量を入力とし、音の発生区間を出力する(図2)。本手法では、環境中の各物体をクラスタリングする手段としてレーザーレンジファインダから取得したポイントクラウドを用いる。ポイントクラウドを用いたクラスタリングはカメラから取得した輝度によるクラスタリングと比較し、物体と類似した色が多く含まれる環境下でも安定したクラスタリングが可能である。また、音響特徴量には MUSIC スペクトルを用いる。

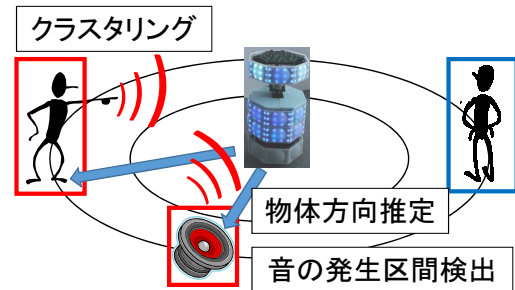


図1: クラスタリングと音の発生区間検出

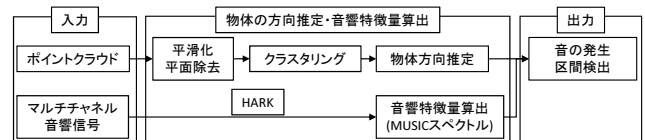


図2: 音の発生区間検出の処理の流れ

まず、各物体方向の推定法について述べ、その後音の発生区間検出手法について述べる。

### 3.1 物体の方向推定

各物体の方向は深度情報をクラスタリングし推定した各クラスターの重心を物体方向として推定する。入力はレーザーレンジファインダから取得したポイントクラウドである。

まずポイントクラウドからクラスタリングを行う事前処理として平滑化と平面除去を行う。平面除去を行うことでクラスタリングの精度を向上させることができる。次に事前処理を施したポイントクラウドから kd-tree を構成する。kd-tree を構成することで、クラスタリングを行う際に、ポイントクラウドを全探索するのではなく、バイナリサーチとなるため計算コストが少なくすることができる。そして、ポイントクラウド間のユークリッド距離から K 近傍法を用いてクラスタリングを行う。その後、各クラスターの重心を求め、各重心の方向を物体の方向とする。以上のようにして深度情報からクラスタリングを行い、物体の方向推定を行う。

### 3.2 音の発生区間検出

音の発生区間検出は 3.1 章で推定した方向情報と音響特徴量を用いて、推定した方向の音響特徴量の閾値判定により行う。入力は、マイクロフォンアレイから取得したマルチチャンネル音響信号である。

まず、マルチチャンネル音響信号をロボット聴覚ソフトウェア HARK [2] を用いてイベント検出に用いる音響特徴量である MUSIC スペクトルを算出する。次に、推定した物体の方向情報と MUSIC スペクトルを用いて、MUSIC スペクトルから物体の方向の MUSIC スペクトルの値を取得する。最後に取得した MUSIC スペクトルを閾値判定し、閾値を超えた区間を音の発生区間とする。以上の

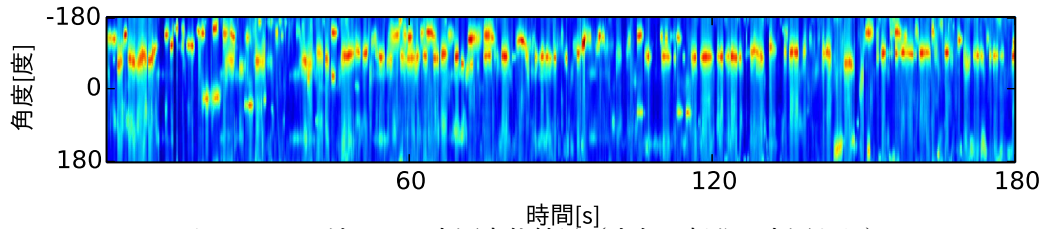


図 4: MUSIC 法による音源定位結果 (赤色の部分に音源あり)

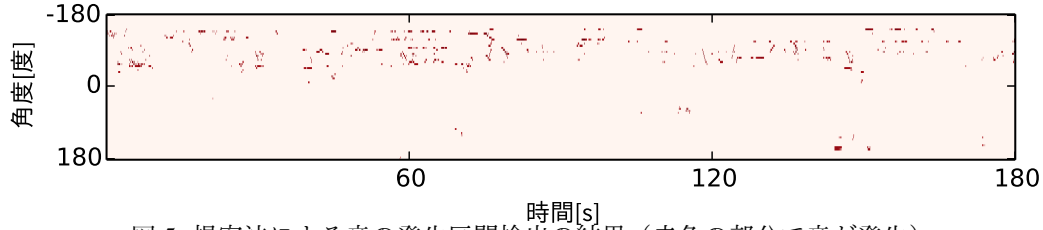


図 5: 提案法による音の発生区間検出の結果 (赤色の部分で音が発生)

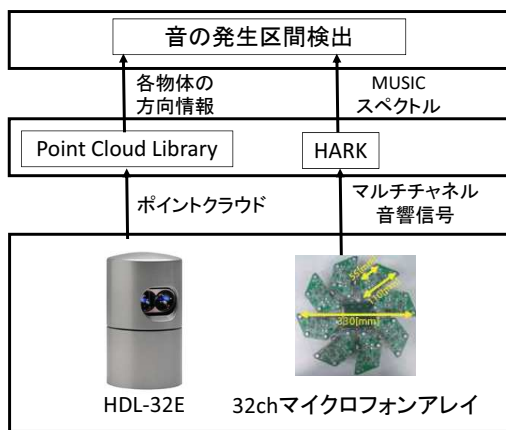


図 3: 音の発生区間のシステム構成

ようにして方向情報と音響特徴量を用いて音の発生区間検出を行う。

#### 4. 動作検証

開発したクラスタリングと音の発生区間検出手法を用いた動作検証を行った。また、本手法と比較するためマルチチャンネル音響信号の音源定位手法である MUSIC 法による音検出を行った。

##### 4.1 システム構成

システム構成は図 3 の通りである。本システムへの入力データはレーザーレンジファインダ HDL-32E で取得したポイントクラウドと佐々木らが開発した 32ch マイクロフォンアレイ [3] で取得したマルチチャンネル音響信号である。これらセンサは加賀美らが開発した Peacock[4] に搭載されているものを用いた。ポイントクラウドを Point Cloud Library<sup>†</sup> を用いて各物体の方向情報を、マルチチャンネル音響信号を 32ch まで扱えるように拡張した HARK を用いて MUSIC スペクトルを算出し、音の発生区間検出を行う。MUSIC スペクトル算出に必要な伝達関数は各マイクの位置関係から幾何的に算出したものを用いた。

##### 4.2 動作結果と考察

動作検証に用いたデータは 2014 年 11 月に日本科学未来館で行われた G 空間 EXPO で Peacock を用いて収録され、時間同期されたデータを用いた。本稿では、3 分

間のデータを用いて提案する音の発生区間検出手法の動作検証を行った。MUSIC 法による音源定位結果は図 4、提案法による音の発生区間検出の結果は図 5 の通り。

提案法により検出した音の発生区間は MUSIC 法による音源定位結果と一致する部分は多い。また、MUSIC 法による結果で観察できる 30 度付近の音源は提案法では推定されずクラスタリングできない遠方の雑音であることが推測される。このように背景雑音となる音源を提案法では検出しないことができると考えられる。一方、0 度から -180 度の間で MUSIC 法による推定音源数よりも提案法による推定音源数が多くなってしまふ部分が多く見られる。2 つ音源の間のもう 1 つの物体の MUSIC スペクトルも音が発生していない状況でも閾値を超えてしまったことが原因である。今後、音の発生区間検出の適切な閾値を設定するとともにより有効な音響特徴量の選択や音の発生区間検出手法の改良が必要である。

#### 5. おわりに

本稿では、深度センサとマイクロフォンアレイを用いた物体検出や音イベント検出を用いてユーザの着目したい物体の情報や重要なシーンのみを検出し、ユーザに提示するシステムの要素技術であるクラスタリングを用いた音の発生区間検出手法を開発した。また、レーザーレンジファインダ HDL-32E と 32ch マイクロフォンアレイで取得したデータを用い、提案手法の動作検証を行った。今後は音の発生区間を人手でラベリングした正解データを元に提案手法の評価実験や音の発生区間検出に用いる特徴量の変更や音イベント検出手法の改善が考えられる。謝辞 本研究は科研費 No.24220006, No.24700168, IMPACT「タフ・ロボティクス・チャレンジ」の支援を受けた。Peacock の収録データを提供頂いた産業技術総合研究所の佐々木洋子博士に感謝する。

#### 参考文献

- [1] J. Even *et al.* MULTI-MODAL SOUND LOCALIZATION FROM A MOBILE PLATFORM. *AI チャレンジ研究会* (第 36 回), 58-63, 2012.
- [2] K. Nakadai *et al.* Design and Implementation of Robot Audition System 'HARK' - Open Source Software for Listening to Three Simultaneous Speakers. *Advanced Robotics*, 24(5-6):739-761, 2010.
- [3] Y. Sasaki *et al.* 32-Channel Omni-Directional Microphone Array Design and Implementation. *Journal of Robotics and Mechatronics*, 23(3):378-385, 2011.
- [4] S. Kagami *et al.* Autonomous Robot Navigation Functions with Auditory Environmental Mapping. *AI チャレンジ研究会* (第 38 回), 1-2, 2013.

<sup>†</sup><http://pointclouds.org/>