6S-5

深度センサとマイクロホンアレイを用いた音源位置可視化による聴覚アウェアネスの提示

井山 貴裕 † 杉山 治 ‡ 大塚 琢馬 ‡ 糸山 克寿 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

近年，複数のマイクロホンを用い，音源の定位・分離を行う研究・開発が進んできている．マイクロホンアレイを用いることで，各マイクロホンから得られる観測信号間の位相差などを元に音源定位・分離を行うことができ，それらの情報を可視化し，直感的にユーザに提示する研究が行われている．本研究では，音環境をより容易に理解するため，音源の可視化に取り組む．その中でも，特に人が気づきを覚える音源や人が注意を向ける音源に着目する．気付きを覚えるときは，一つは人が話し始めたときなど新しく音を発する音源が現れたときである．もう一つは，話している人が突然大声を上げるときのように既に存在する音源のパワーや音のパターンが変化したときである．

本研究では，これらのような環境内に存在する音源のなかでも人が注意を払う音源を顕著性が高い音源と定義し，顕著性を通じたユーザの聴覚アウェアネスの補佐・提示を行うシステムを提案する．音の顕著性を提示するためには，環境内に音源がいくつあるかを理解する音源分布の理解，音源がどの物体から発せられているのかという空間的な音源位置の理解，そして音源物体の位置や音情報の変化などから求まる音の顕著性の理解の 3 つのプロセスが必要であると考えられる．本研究では，これら 3 つのプロセスをそれぞれレイヤと定義し，聴覚アウェアネスの三層モデルを提案する．そして，提案モデルに基づいた可視化システムを実装した．本稿では，聴覚アウェアネスの三層モデルとその可視化システムの実装について述べる．

2. 関連研究

音環境を可視化する研究は従来から続けられている．例えば，大内らや神保らの研究では，マイクロホンアレイから得られた観測混合音の強さ・音響スペクトルをや到達時間差から推定した音源位置を可視化した [2][3]. これらの研究では，音源定位結果のみを入力としていたため，音源が三次元空間のどこに存在し，音源位置が画像領域内のどの範囲にマッピングされるのかを解析できず，三次元位置を可視化することはできなかった．さらに，このような問題を受け，マイクロホンアレイに深度センサの情報を加え，三次元的に音源物体の位置を推定・可視化する研究が進んでいる. Jani らの研究では，距離センサから得られた距離情報に基づき，音源を発している三次元位置を推定・追跡することができる [4].

しかしながら，これらの研究においても音源を発している対象の範囲までは推定していないので，音源が重なった場合の対応ができておらず，また，その音源が新規のものであるかどうかを推定するまでには至っていない．本研究では，音環境理解の観点から，既存研究で解析されてきた情報を音源分布レイヤと音源位置レイヤの

visualization of acoustic awareness based on sound source positions estimated by depth sensor and microphone array: *Takahiro Iyama, Osamu Sugiyama, Takuma Otsuka, Katsutoshi Itoyama, and Hiroshi G. Okuno (Kyoto Univ.)*

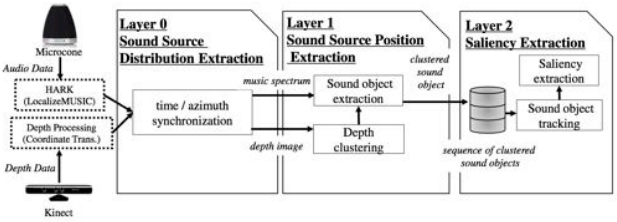


図 1: 聴覚アウェアネスの可視化のための三層モデル

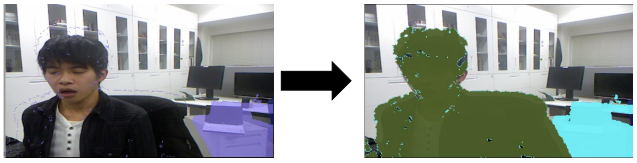
2 つのレイヤに整理し，さらにこれら 2 つの上位レイヤとして時系列の情報を解析する顕著性レイヤを加えることで段階的にユーザに聴覚アウェアネスを提示する可視化システムを提案する．

3. 聴覚アウェアネス可視化のための三層モデル

本研究では，マイクロホンアレイと距離画像を用いて得た音源の情報を，音源分布・音源位置・顕著性の 3 つのレイヤで表現し，ユーザに聴覚アウェアネスを提示する．これらのレイヤに分けて可視化することで，ユーザは RAW データから，空間情報を統合し推定された音源物体の提示，さらにその音源物体を追跡することで検出される顕著性の提示までの様々な情報を切り替えながら音環境を解析することができる．

レイヤー 0: 音源分布レイヤ　音源分布レイヤでは，同一時刻に得られた全方位カメラで得られた画像とマイクロホンアレイから得られた聴覚情報を統合し，1 つの画像 (フレーム) として可視化する (図 1 左部参照)．具体的には，ロボット聴覚ソフトウェア HARK[1] によって得られた MUSIC スペクトルのパワー情報を方向情報を揃えた上で，全方位カメラで得られた画像上にマッピングする．異なる 2 つの情報を 1 つの画面上に描画することによって，ユーザは音が画像中のどの部分から発生しているのかを直感的に理解することができる．

レイヤー 1: 音源位置レイヤ　音源位置レイヤでは，下位の音源分布レイヤの情報に加えて，深度センサの距離画像の情報を利用し，音源物体の三次元位置を推定する (図 1 中央部参照)．具体的には，Kinect で得られた距離画像データから，三次元座標の点群を求め，Kmeans 法を用いてそれらの点群をクラスタリングし N 個のクラスタを得る (本研究では N=10 とした)．これら得られたクラスタと音源分布レイヤから得られた音源位置を統合することで，音源位置にあるクラスタを音源物体として保持する．そして，これらの情報を元に距離画像からクラスタ上に MUSIC スペクトルのパワー情報をマッピングし可視化する．これら一連の処理により，物体の範囲を考慮した音情報のマッピングを行うことができ，ユーザは画像中の物体と音源との対応を直観的に解析することができる．



a. 時刻t-1では画像中の人物の顕著性は低い状態

b. 時刻tでは画像中の人物が音を発したため顕著性が高くなっている

図 2: 時刻 t で画像中の人物が音を発したため顕著性が高くなっている

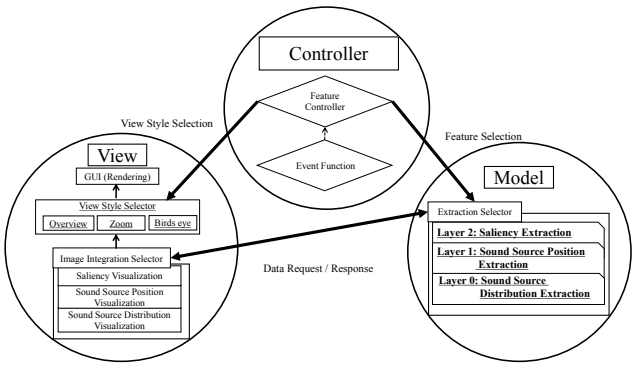


図 3: システム図

レイヤー 2: 顕著性レイヤ　顕著性レイヤでは，音源位置レイヤで得られた音源物体を追跡し，顕著性のある音源の推定と可視化を行う．音源の顕著性を考えたとき，顕著性が高い物体とは，我々が最も着目する状態は新しく出現した音源物体とこれまで存在した音源の音のパワーやパターン・位置が変化した物体の 2 つであると考えられる．逆に，顕著性が低い物体とは音を発していない物体・音のパワー・パターン・位置が変化しない物体であると考えられる．顕著性レイヤでは，音源物体を追跡し，これら音源の顕著性に基づいた可視化を行う．顕著性の検出は，音源物体の追跡と顕著性フィルタの 2 つの処理から成り立つ．音源物体の追跡では，時刻 t と時刻 t-1 の音源物体 (クラスタ) の位置．分散，クラスタ内の MUSIC スペクトルの平均と分散から仮想的距離を算出し，物体の同一判定を行う．その結果を受けて，顕著性フィルタではそれぞれの物体にエントロピーを与え，その値に基づき，顕著性の有無を判定する．結果，出現してから時間が立っていない，もしくは音のパワー・パターンが変化して間もない物体の顕著性が高くなり，時間とともに減衰していく．このように得られた顕著性に基づいて，顕著性レイヤは音源物体のマスク処理を行う．具体的には，顕著性の高いものほど透過性が低く，顕著性の低いものは透過度が高くなるように画像上に音源情報をマッピングする．これら一連の処理により，ユーザは音源物体の存在とともに，その顕著性有無を直感的に解析することができる．顕著性レイヤの可視化の一例を図 2 に示す．

4. 聴覚アウェアネス可視化システムの構成

図 3 に本研究で実装した聴覚アウェアネスの三層モデルに基づいた音環境の可視化システムの概要を示す．システムは，MVC モデルで設計されており，モデル部（図 3 右下部），ビュー部（図 3 左下部），コントローラ部（図 3 上部）で構成される．モデル部ではデータの処理

a) 音源分布レイヤの可視化結果　b) 音源位置レイヤの可視化結果　c) 顕著性レイヤの可視化結果

図 4: 各レイヤの可視化結果

を，ビュー部では，モデル部で解析された情報に基づいた可視化画像の合成を，コントローラでは各機能の切り替えをそれぞれ行う．以下，それぞれの機能について詳細に述べる．

モデル部　モデル部では，マイクロホンアレイとロボット聴覚システム HARK から得られるミュージックスペクトラムと Kinect から得られる深度画像を用いて，3 章で述べた三層モデルに基づき，音源分布，三次元の音源位置・分布，顕著性の有無をそれぞれ検出し，ビュー部に伝達する．また，コントローラ部からの入力に基づき，どのレイヤの処理を行うかを選択的に決定する．

ビュー部　ビュー部では，モデル部から送られてきた三層モデルの各レイヤのデータ（色画像，音源分布，三次元音源位置，顕著性の有無）に従って，音源の可視化画像を合成する．可視化画像にはオーバービュー・ズーム・鳥瞰図の形式が存在し，コントローラ部からの入力に基づき，その切り替えを行う．各レイヤの可視化結果の例を図 4 に示す．

コントローラ部　コントローラ部では，キーボードから入力される各機能に割り当てられたショートカットキー，もしくは GUI からの入力に基づき，モデル部に処理の切り替え，ビュー部の可視化形式の切り替えの指示を行う．

以上の設計により，ユーザはそれぞれのレイヤの可視化画像を切り替えながら，音環境の解析を行うことができる．

5. おわりに

本研究では，段階的なユーザへの聴覚アウェアネスの提示を実現するため，聴覚アウェアネスの三層モデルに着目した．三層モデルとして，音源分布・音源位置・顕著性の 3 つのレイヤからなるモデルを提案し，各レイヤの構成と可視化データの作成手順について述べた．

今後の展望としては，1. kmeans クラスタリングのパラメータが現在，ヒューリスティックのため，被験者実験を通じてパラメータを学習し決定する，2. 仮想距離の内部パラメータや同一音源物体内でのアルファ値の減少度のパラメータを適宜ユーザが変更できるようにする，3. 実時間処理を目指すなどが考えられる．

謝辞

本研究は科研費基盤研究 (S) No.24220006 の支援を受けた．

参考文献

[1] 中臺一博, et al. :”ロボット聴覚オープンソースソフトウェア.” 日本ロボット学会誌, (2010)
[2] 大内康裕, et al. :”音響テレビを用いた音場の可視化.” 日本音響学会アコースティックイメージング研究会資料, AI2009-2-06 (2009).
[3] 神保直史, et al. :”多チャンネル音場計測システムを用いた音環境の可視化” 日本音響学会講演論文集,(2008)
[4] Even, Jani, et al. :”Combining Steered Response Power with 3D LIDAR scans for building sound maps.” 人工知能学会研究会資料, SIG-Challenge-B302-08(2013)