

聴覚アウェアネス可視化モデルに基づく ジェスチャ操作インタフェースの開発

○井山 貴裕¹ 杉山 治² 坂東 宜昭¹ 糸山 克寿¹ 吉井 和佳¹ 奥乃 博³

¹ 京都大学 大学院情報学研究科 ² 東京工業大学 大学院情報理工学研究科

³ 早稲田大学 実体情報学博士プログラム

1. 序論

環境の監視・探索システムの性能・機能向上には画像情報だけでなく音情報を利用することが不可欠である。例えば、監視システムのリモートオペレータにとっては、マイクを通した音だけでは音の到来方向や種類を認識することは困難である。一方、マルチチャンネル信号処理を用いれば様々な音情報を提示することは可能であるが、聴覚アウェアネスが十分に考慮されているとは言いがたい。聴覚アウェアネスとは、音源の方向や音圧だけでなく、音源の位置や音源の状態の変化の気づきのことを意味する。従来の音情報可視化インタフェースは [1][2][3]、音情報を加工せずに直接提示するものであり、操作が直感的ではなかった。

本研究では、実時間で動作し、直感的に操作可能な聴覚アウェアネス可視化システムを設計・実装する。従来の聴覚アウェアネス可視化インタフェース [3] では、多数のパラメータをユーザがスライダやボタンで指定する必要があった。本研究では、このインタフェースを拡張し、多様なパラメータ操作をユーザのジェスチャ操作で解決する。ジェスチャによる操作を用いることで、ユーザは簡易で直感的な操作で要求を満たす結果を得ることができる。本稿ではジェスチャ操作可能な可視化インタフェースを取り上げ、関連研究、提案手法、手法の設計、および可視化結果例について述べる。

2. 関連研究

直感的な聴覚アウェアネス可視化システムの開発のため、音環境の可視化に関する従来法を概観する。

神保ら [1] は、192 個のマイクロフォンアレイと CMOS カメラを使用し、RGB 画像上へ音高の帯域ごとの強さを重畳表示している。この可視化手法は、音源の分布を提示するのみで、音源の位置や時間変化といった聴覚アウェアネスの提示は行っていない。

Evenら [2] は、マイクロフォンアレイとレーザーレンジファインダを使用し、SLAM で作成した地図上に音源の位置を重畳表示している。この可視化手法は、音源の強さと位置を提示するが、音源の時間変化といった聴覚アウェアネスの提示は行っていない。

これらの研究を受け、井山ら [3] は、マイクロフォンアレイと深度センサを使用し、聴覚アウェアネスを三層モデルで定義し、これを可視化するシステムを開発した。この三層モデルは、環境内の音の分布を概観する機能を提供する音源分布レイヤ、着目した音源の位置やパワーを抽出する音源位置レイヤ、新しい音源の出現や音源のパワーの大きな変化を抽出する顕著性レイヤ

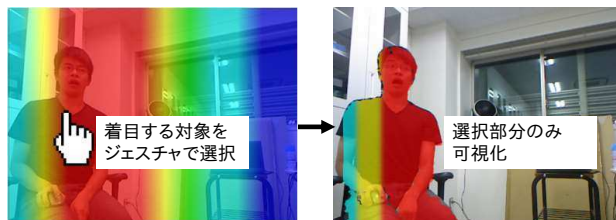


図 1 ジェスチャ操作可能なインタフェース

から構成される。可視化インタフェースは、ユーザが柔軟に環境内の特定の対象をフォーカスできるようにするため、Shneiderman の Visual Information-Seeking Mantra [4] に基づいて設計されている。ユーザは要求に合わせて、ユーザ自身が様々な特徴量をスライダの操作によって可視化結果を変更することができる。このシステムは、設定するパラメータが大量にあり、ユーザの要求を満たす結果を得るための操作が煩雑で直感的ではない。

本稿では、聴覚アウェアネスの三層モデルを拡張し、ジェスチャ操作可能な可視化システムを開発する。リモートオペレータによる効率的な探索システムでは、探索システムから得られる情報は瞬時に提示される必要があり、可視化インタフェースの操作は簡易で直感的であることが重要である。そこで、多様なパラメータ操作をユーザのジェスチャ操作で解決する。大量のパラメータ操作をユーザのジェスチャ操作のメタファを用いることで、より直感的なシステムの操作やパラメータの変更を実現する。

3. ジェスチャ操作可能な可視化システム

ユーザが要求する可視化結果を得るためのインタフェースの操作を簡易で直感的にできるよう設計する (図 1)。そのため、GUI の操作やパラメータの変更はスライダやボタンでなく、可視化画像に対するジェスチャを用いて行う。例えば、ユーザが着目したい音源対象を選択する操作は、画面内の画像を直接選択できるほうがより直感的である。本インタフェースはジェスチャの入力としてマウスの左クリック、右クリック、中クリック、マウスホイールを使用する。以下では、レイヤの概要と各レイヤにおける操作について述べる。

3.1 インタフェース設計

開発したシステムのインタフェースはユーザが環境内の音源情報を観測し、直感的に操作できるよう設計した。GUI は図 2 に示すように、画像表示部、ステータス部から成る。画像表示部に、三層モデルで生成した画像を組み合わせたものが表示される。ユーザは画



図2 インタフェースのデザイン

表1 インタフェースのパラメータ

レイヤ	パラメータ	概要	範囲
1	透明度 t	MUSIC スペクトルのパワー画像の透明度	$0 < t < 1$
	可視化の最小値 p	可視化する MUSIC スペクトルの最小値	$0 < p$
2	領域成長法の閾値 d	領域成長法で統合するかの閾値	$0 < d$
3	重み α	顕著性の重みパラメータ	$0 < \alpha < 1$

画像表示部上で三層モデルで使用するパラメータを変更することによって、表示する画像を変更できる。表1はユーザが変更できるパラメータの一覧である。レイヤごとにパラメータは存在し、ユーザは三層モデルにおける特徴量を柔軟に組み合わせることができ、自由に画像表示部に表示される描画結果を変更することができる。ステータス部は現在のレイヤやパラメータの変更内容などを表示する部分である。これによりユーザは現在のレイヤ状態やパラメータ状態を確認しながら操作することができる。

3.2 音源分布レイヤ

ユーザは音源分布レイヤで環境内の音源の分布を概観する。ユーザに提示される可視化内容は環境内の音源の分布を MUSIC スペクトル [5] の色画像を RGB 画像に重畳した画像である。本レイヤでは、ユーザは重畳される MUSIC スペクトルの色画像の濃淡と可視化する MUSIC スペクトルの帯域を変更できる (図3)。これらの機能によって、ユーザは環境や要求に応じた可視化結果を得ることができる。例えば、音の分布を鮮明に観察するときは濃く、音の弱い部分が必要ないときは帯域の最小値を上昇させることができる。

これらのパラメータの変更は、いずれも増減であるため、マウスホイールの操作により変更する。マウスホイールを上に戻ると、色画像の透明度や可視化する MUSIC スペクトルの最小値が増大し、下に回ると減少する。これら2つのパラメータのどちらを変更するかの切り替えは、右クリックで行えるようにする。



図3 マウスホイールによる透明度の変更



図4 クリックによる音源の選択

3.3 音源位置レイヤ

ユーザは音源位置レイヤで着目した音源の情報を観察する。ユーザに提示される可視化内容は、着目した音源の RGB 画像上にも MUSIC スペクトルを重畳した画像である。本レイヤでは、ユーザは着目する音源対象の選択と領域成長法 [6] の類似度の閾値パラメータの変更できる (図4)。これらの機能によって、ユーザは着目したい音源を選択できる。

着目する音源の選択は、可視化領域内の着目する画像をマウスの左クリックにより行う。領域成長法の閾値パラメータが大きいと、より広い範囲に存在する複数物体を同一領域とみなし、小さいと、領域をより細かく分割する。このパラメータの変更は増減であるため、マウスホイールの操作により変更する。マウスホイールを上に戻ると、閾値パラメータは増大し、下に回ると減少する。

3.4 顕著性レイヤ

ユーザは顕著性レイヤで着目した音源の顕著性を観察する。音源の顕著性とは、音響情報や音源の位置・形状の時間変化によって定義される。音を発していない音源が音を発し始める場合や、環境内に新しく音源が出現した場合は顕著性が大きくなる。一方、音源が音を発生していない場合や、音源が発生している音に変化がない場合は顕著性が小さくなる。ユーザに提示される可視化内容は着目した音源の顕著性の大きさに対応した色枠を RGB 画像に重畳した画像である。本レイヤでは、ユーザは顕著性を算出する際の音響情報と深度情報の重みの変更できる。音響情報の重みが大きくなるほど、着目した音源の音響情報の変化が顕著性に大きく影響し、深度情報の重みが大きくなるほど、着目した音源の移動量や形状の変化が顕著性に大きく影響するようになる。

これらのパラメータの変更は、いずれも増減であるため、マウスホイールの操作により変更する。また、音響情報と深度情報のどちらを変更したいかは、ユーザの要求に応じて変わるので、重みを変更する情報をマウスの右クリックで変更する。

各レイヤ間の移動については以下のように設計する。音源分布レイヤから音源位置レイヤへの移動は、音源位置レイヤで着目する音源を左クリックしたときに移動する。音源位置レイヤから顕著性レイヤへの移動は、音源位置レイヤで着目している音源の領域内を左クリックしたときに移動する。顕著性レイヤから音源位置レ



図5 聴覚アウェアネスの可視化システムの詳細

イヤ、音源位置レイヤから音源分布レイヤへの移動はマウスの中クリックを行うことで移動する。

4. 可視化システムの設計

聴覚アウェアネスの三層モデルを用いて、聴覚アウェアネスの可視化システムを開発する。本システムへの入力データはRGB画像、深度画像、および多チャンネル音響信号である。動作例のシステムはKinectを用いてこれらを取得する。図5に示すように、入力は、Kinectから得られる。RGB画像と深度画像はOpenNIライブラリ[7]を、多チャンネル音はHARK(Honda Research Institute Japan Audition for Robots with Kyoto University)[8]を通じてそれぞれ取得され、システムに渡される。各レイヤのデータ処理やGUIへの様々な画像の描画にはProcessingを用い、三層モデルの特徴量を柔軟に変化させることができるようシステムを設計する。

聴覚アウェアネスの可視化システムにおいて、データは異なる時間間隔で得られるが並行して処理され、低次のレイヤから高次のレイヤに送られる。図6はデータ処理の概略を示している。マルチメディアデータを扱うため各レイヤは、データを解析するバックエンド処理と可視化画像を生成するフロントエンド処理の2つから成る。

音源分布レイヤにおけるバックエンド処理は、Kinectから取得される深度データ、RGB画像、MUSICスペクトルの時間同期である。各データは異なるフレームレートで取得されるので、音源分布レイヤでは、各データのタイムスタンプを比較することで各データの時間を同期する。また、MUSICスペクトルはKinectから取得したマルチチャンネル音響信号をHARKを用いて算出する。音源分布レイヤにおけるフロントエンド処理は、MUSICスペクトルのパワーの大きいものを赤、小

さいものを青と対応させた色画像の生成である。時間同期されたデータは音源位置レイヤに送られ、生成された画像はビュー部の音源分布画像生成部に送られる。

音源位置レイヤにおけるバックエンド処理は、深度画像のクラスタリングである。クラスタは、ユーザの指定した領域に該当し、深度画像に領域成長法を用いてユーザの指定した領域を生成する。音源位置レイヤにおけるフロントエンド処理は、生成したクラスタ上のみMUSICスペクトルの色画像を重畳した画像の生成である。そして、クラスタ情報は顕著性レイヤに送られ、生成された画像は音源位置画像生成部に送られる。

顕著性レイヤにおけるバックエンド処理は、音源位置レイヤから受け取ったクラスタの追跡と顕著性の算出である。顕著性レイヤのフロントエンド処理は、顕著性に基づく画像の生成である。顕著性の大きなクラスタは赤枠、顕著性の小さなクラスタは青枠で囲まれる画像を生成する。各レイヤのフロントエンド処理を実行するかどうかはコントローラ部からの命令によって決定するため、モデル部ではそれに合わせることで必要なデータ処理を適切に行うことができる。

5. システムの可視化結果

本章では、開発した可視化システムの動作例について述べる。使用したデータは、センサの近くにいる話者とセンサから少し離れた位置にあるラップトップが音楽を再生している環境から取得したものである。ユーザは話者の音の様子を観測することを目的としているとする。

5.1 音環境の概観 (レイヤ1)

音源分布レイヤを用いることでユーザは、環境内の音の分布を概観することができる(図7, 上段)。ユーザは話者やラップトップの周辺から音が発生していることがわかる(図7, 上段a)。重ねるMUSICスペクトルの透明度を調整することで、ユーザは音源がより観察しやすくなる可視化結果を得ることができる(図7, 上段b)。可視化するMUSICスペクトルの最小値を調整することで、ユーザはある閾値より大きなMUSICスペクトルをもつ領域のみを観察することができる(図7, 上段c)。

5.2 着目する音源の選択 (レイヤ2)

音源位置レイヤを用いることで、ユーザは指定した位置にある音源のみを観察することができる(図7, 中段)。図7, 中段aは話者を選択した場合、図7, 中段bはラップトップを選択した場合である。

5.3 着目した音源の顕著性の観察 (レイヤ3)

顕著性レイヤを用いることで、ユーザは着目した顕著性を観察することができる。ユーザは話者の時間変化の様子を観察することができる(図7, 下段)。図7, 下段aは、話者が発話を始めた場面である。そのため、話者の顕著性は増加し、枠の色は赤に変化する。一方、図7, 下段bは、話者が発話をしていない状態である。そのため、話者の顕著性は減少し、枠の色は青に変化する。

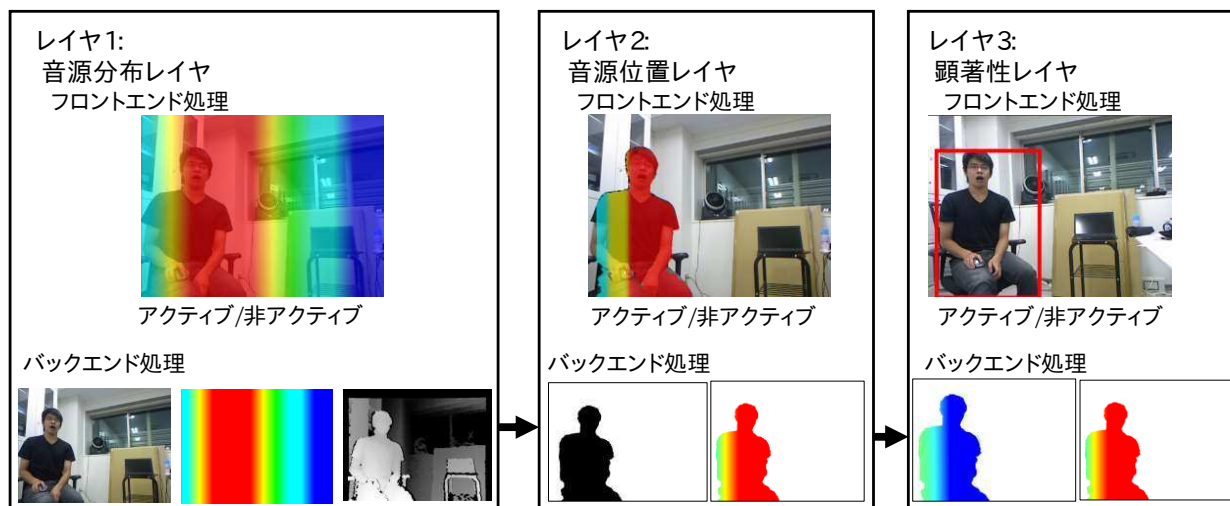


図6 三層モデルのデータ処理

レイヤ1: 音源分布レイヤ



レイヤ2: 音源位置レイヤ



レイヤ3: 顕著性レイヤ

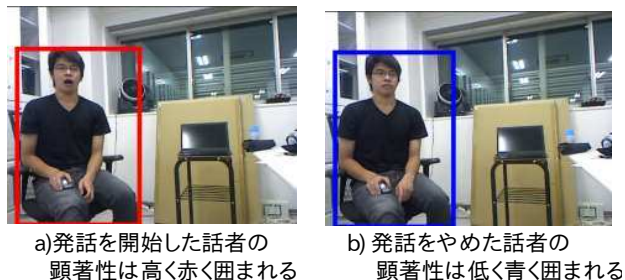


図7 可視化システムの可視化結果例

各レイヤの特徴を調整することで、ユーザは柔軟に着目したい音源を選択することができ、その様子を観察できる。このように設計した聴覚アウェアネスの可視化システムは音環境を観察する際のユーザの聴覚アウェアネスを改善する。

6. 結論

本研究では、音源分布レイヤ、音源位置レイヤ、顕著性レイヤから構成される聴覚アウェアネス可視化の三層モデルを設計し、Kinectを用いた聴覚アウェアネス可視化システムを実装した。音源分布レイヤは環境内の音の分布を概観する機能を、音源位置レイヤは着

目した音源情報を抽出する機能を、顕著性レイヤは音情報の時間変化、すなわち、新しい音源の出現や音源のパワーの大きな変化といった顕著性を抽出する機能を提供する。三層モデルに基づくデータ処理や可視化を行い、各レイヤのパラメータをジェスチャを用いて変更することで、音環境を分析するための直感的な操作が可能なインタフェースを開発した。今後、開発したインタフェースの有効性を検証するため、本インタフェースと他のインタフェースを用いて探索タスクを行い、操作に要した時間の比較やインタフェースの使いやすさなどの評価を行なっていく予定である。

謝辞 本研究の一部は科研費 No.24220006 と No.24700168 の支援を受けた。

参考文献

- [1] J. Naoshi, et al. Visualization of sound environment using multi channel acoustic measurement system. In *Acoustic Society Symposium, 2008*, pp. 1509-1510, Sep 2008.
- [2] J. Even, et al. Creation of radiated sound intensity maps using multi-modal measurements onboard an autonomous mobile platform. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 3433-3438, Nov 2013.
- [3] 井山他. 深度センサとマイクロフォ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. 情報処理学会第76回全国大会(2014), 第2巻, pp. 489-490, March 2014.
- [4] B. Shneiderman, et al. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company, USA, 5th edition, 2009.
- [5] F. Asano, et al. Real-time sound source localization and separation system and its application to automatic speech recognition. In *INTERSPEECH*, pp. 1013-1016, 2001.
- [6] D.H. Ballard, et al. *Computer Vision*. Prentice Hall, 1982.
- [7] simple-opensi - opensi library for processing. <https://code.google.com/p/simple-opensi/>.
- [8] K. Nakadai, et al. Design and Implementation of Robot Audition System 'HARK' - Open Source Software for Listening to Three Simultaneous Speakers. *Advanced Robotics*, Vol. 24, No. 5-6, pp. 739-761, 2010.