

サラウンド映像視聴時の注視点分布の分析

高比良 英朗^{*1} 篠原 未歩^{*2} 野坂 祐介^{*2} 横内 雅也^{*3}
山田 光穂^{*2} 平山 高嗣^{*1*4} 川西 康友^{*1*5}
出口 大輔^{*1} 井手 一郎^{*1} 村瀬 洋^{*1}

Analysis of Gaze Point Distribution while Viewing Surround-Sound Videos

Hideaki Takahira^{*1}, Miho Shinohara^{*2}, Yusuke Nosaka^{*2}, Masaya Yokouchi^{*3},
Mitsuho Yamada^{*2}, Takatsugu Hirayama^{*1*4}, Yasutomo Kawanishi^{*1*5},
Daisuke Deguchi^{*1}, Ichiro Ide^{*1} and Hiroshi Murase^{*1}

Abstract – We have been conducting research focusing on the human gaze as a principal biological reaction while viewing videos. In this paper, we clarify the difference in the area of gaze region under different sound conditions; *monaural* (1.0 ch), *stereo* (2.0 ch), and *surround* (5.1 ch). We measured the gaze of subjects viewing video scenes of natural landscapes and calculated the horizontal and vertical standard deviations of the gaze points and the area of gaze region defined from them. A two-way analysis of variance shows the main effects of each of the sound and scene conditions on the vertical standard deviation of the gaze points and the interaction effect and the main effect of the sound condition on the area of gaze region. In addition, a multiple comparison test shows the significant differences in the vertical standard deviation for the interaction effect between the sound and scene conditions and in the area of gaze region for the sound conditions between *monaural* and *stereo* or *surround*. The hypothesis that the variance of gaze points is expanded by the surround sound is confirmed experimentally.

Keywords : gaze analysis, surround-sound, video viewing

1. はじめに

近年の4K/8K映像表示技術やサラウンド（マルチチャンネル）音響システムの発展、テレビの大画面化やホームシアタの普及により、超高精細映像やサラウンド音響などの新しい視聴覚体験を楽しむことが可能になりつつある。この新しい視聴覚体験の主要要素として、超高精細映像とサラウンド音響が挙げられる。

4K/8K超高精細映像では、高精細な画素構造により、標準観視距離（視力1.0を有する視聴者が映像視聴時に、画素構造が見えなくなる距離）が、従来の高精細映像（フルハイビジョン：2K）より画面に近くなる（Hを画面の高さとする、好ましい観視距離は4K：1.5H、8K：0.75H）。このことから、近距離での視聴が推奨され、視野内で広角の映像を見ることが可能である。

一方、サラウンド音響では、従来のステレオスピー

カ（2.0 ch）に、リアスピーカやバックスピーカなどスピーカの数を追加し（5.1 ch / 7.1 ch / 22.2 ch）、視聴者を包み込むように配置することで、臨場感がある音響空間を構築することが可能である。このうち、5.1 ch、7.1 chは水平など同じ平面内にスピーカを配置するが、22.2 chでは垂直方向にもスピーカを配置し上下方向の音像移動を可能にするもので、4K/8Kで新たに追加された規格である。

このような新しい視聴覚体験を提供する環境において、人が映像視聴時に示す生体反応を計測して分析することは、映像制作や映像視聴環境の設計に有用である^[1]。そこで、我々は映像視聴時における人の主要な生体反応の1つである「視線」に着目した研究を行っている。本論文では、異なる音響条件（音が響く効果の違い）で大画面映像を視聴した際の視線を計測し、注視点の分布からサラウンド音響が映像視聴に与える効果を分析する。

我々は以前に、大画面の超高精細映像の視聴環境における視聴位置の影響を分析する研究を行なった^{[4],[5]}。この研究では、映像視聴時の視聴の距離と方向を変えた際の視線を計測し、その結果から注視点分布の違いと頭部運動の違いについて報告している。しかし、これらの研究では、映像視聴時の視聴位置の検討のみに留まっており、視聴する映像や音響そのものの要因に

*1: 名古屋大学 大学院情報学研究所

*2: 東海大学 大学院情報通信学研究所

*3: 東海大学 情報通信学部

*4: 人間環境大学 人間環境学部

*5: 理化学研究所 ガーディアンロボットプロジェクト

*1: Graduate School of Informatics, Nagoya University

*2: Graduate School of Information and Telecommunication Engineering, Tokai University

*3: School of Information and Telecommunication Engineering, Tokai University

*4: Faculty of Human Environment, University of Human Environments

*5: Guardian Robot Project, RIKEN

については未検証であった。超高精細映像の視聴における眼球運動と頭部運動の関係性の研究では、頭部を左/右（または上/下）に向けると、眼球も頭部に対して左/右（または上/下）を向く傾向が報告^[6]されているが、音響の違いに着目した研究は行なわれていない。

視聴者を包み込むようにリアスピーカやバックスピーカなどを配置したサラウンド音響では、視聴者後方から鳴る音に対して注意が引き寄せられることにより、音の広がりを感じるようになると考えられる。この音の広がりにより、視聴者は従来のモノラルやステレオ音響と比較して、広く視線を動かすことが予想される。

そこで我々はサラウンド音響により、注視点の分布が広がるという仮説を立てた。異なる音響での視線を計測し、注視点の分布を分析することで、人の視覚に対する聴覚の作用のメカニズムを解明するための基礎知識となる情報を得る。その結果が映像制作者のための音響システムを用いた視聴環境構築に有用となる可能性がある。

本論文における貢献は、モノラルやステレオ音響に比べて、サラウンド音響により視聴者の注意が広がることによる注視点分布の違いの分析から音と視覚の関係性を明らかにすることである。

以降、2章で関連研究を紹介する。次に、3章で実験設定の詳細を説明する。4章では、実験結果と分析結果を報告する。5章では、実験結果を考察する。最後に、6章で結論を述べる。

2. 関連研究

音と視覚の注意に着目した研究では、聴覚と視覚が相互に影響しあうことが報告されている^{[7]~[9]}。

音像定位と視覚の関係の研究では、腹話術効果（音が鳴っている音像位置と視覚刺激の位置が多少ずれていても視覚刺激から音が鳴っていると知覚する現象）が明らかになっている^{[10],[11]}。この効果は、音源と視覚刺激の空間的距離が近く、提示が同期している場合に生じることが報告されている。Komiyamaは、音像と視覚刺激の方位差が20°以内であれば、音像定位が視覚刺激の方向に引き寄せられることを報告している^[12]。

サラウンド音響では、音はあらゆる方向から聞こえてくるため、方位差は前述の20°を大きく上回る可能性がある。このような状況下での音像と視覚刺激の関係性を調べた研究はなく、基礎的な知見が少ないことが課題の1つである。そこで、本論文では、サラウンド音響で、音と視覚の関係性を明らかにすることを試みる。

サラウンド音響に着目した研究では、特定のタスク

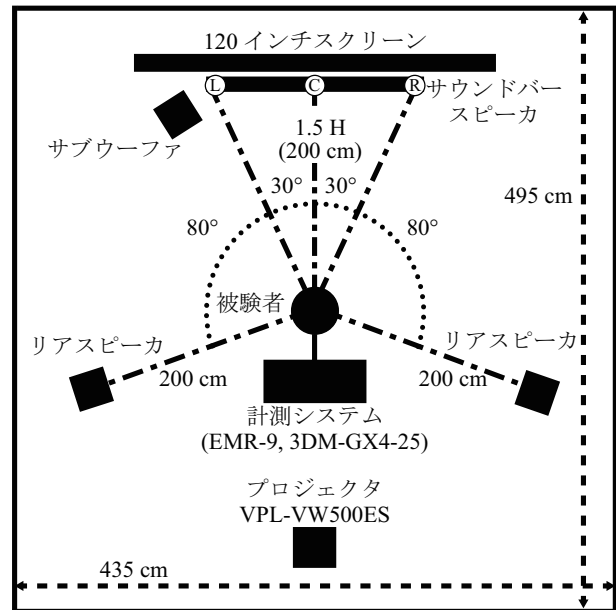


図1 実験室（シールドルーム）配置
Fig.1 Laboratory (shield room) layout.

におけるサラウンド音響と視覚的注意の関係性に着目したものが^{[13],[14]}。実験参加者へ人の数を数えるタスクを課した際に、視野外からのサラウンド音によって視覚的注意が引き寄せられ、視覚的注意領域が広がることが報告されている。これらの研究では、特定のタスクにおける音と注意行動の関係性を明らかにしている。一方で、本論文では、タスクを課さずに自然な条件下での映像視聴時の検証を行ない、サラウンド音響の効果を明らかにする。

3. 実験設定

3.1 実験目的

視聴者の周囲から聞こえるサラウンド音が視線の動きに与える影響を分析する。サラウンド音響と視覚的注意の関係性に着目した先行研究^{[13],[14]}では、サラウンド音により注意領域が広がることが報告されている。そこで、各音響条件間の注視点の分布の広がり方の違いを検証する。

本実験では、自然風景のシーンを用い、モノラル（1.0 ch）、ステレオ（2.0 ch）、サラウンド（5.1 ch）の音響条件で提示した際の視線を計測し、注視点分布の分析を行なう。その結果から、注視点分布に対する音チャンネル数の増加の効果を分析する。

3.2 実験室環境

実験は東海大学所有のシールドルームで行なった。このシールドルームは吸音材の壁面で構成されており、435 cm × 495 cm × 260 cmの空間である。この空間内に、120インチ大画面スクリーンおよびプロジェクタ VPL-VW500ES（SONY製）、サラウンドスピーカ HT-RT5（SONY製）、視線計測システム^[2]を配置

した(図1)。

実験参加者の視距離は4K解像度の1.5Hを参考に200cmとして、サラウンドスピーカ配置はITU-R BS.775-1^[15]に従い、視距離200cmの円周上に配置した。前面のサウンドバースピーカおよび、サブウーファはスクリーンの真下(床上30cm)に設置した。リアスピーカはスクリーン中心から左右110°の方向、実験参加者の耳の高さ(床上100cm)になるように三脚に固定した。

実験時は室内の照明を消灯し、その際のスクリーン輝度は最小:0.04 cd/m²、最大:104 cd/m²であった。

3.3 実験映像・音響設定

実験映像は、市販のブルーレイディスクである「森林浴サラウンドブルーレイ・エディション [映像遺産・ジャパントリビュート]」(シンフォレスト)を使用した。画面全体を満遍なく見せたい映像や注意を引く可能性がある対象が画面全体に広く分布する映像では、サラウンド音による視線移動が促されやすいと仮説を立て、自然の風景を中心とした俯瞰的に撮影されたシーンが多い映像を選出した。この映像から、リアスピーカから音が鳴る箇所を含むカットを選別し、森林(Forest A, Forest B, Forest C)と渓谷(Valley A, Valley B, Valley C)を各1分間のシーンとして、編集した。森林は木々の揺らめきを主とした緑が多い映像、渓谷は溪流を主とした水が多い映像であり、サラウンド音によって視線移動が生じやすくなると仮定して選出した。両映像は、この観点から、分析対象として同質と考える。シーンが持つ固有の映像特徴による視線移動への影響を検証するために、シーン条件として森林、渓谷それぞれでバリエーションを増やし、異なるカットを3つずつ用意した。映像の音声は自然の風景で発生する音で自然の中に溶け込む音である。森林では木々のざわめきや鳥の鳴き声、渓谷では水の流れる音などが収録され、風景とは関係ない音は含まないものである。

音響条件は、モノラル(1.0ch)、ステレオ(2.0ch)、サラウンド(5.1ch)とした。ブルーレイディスクに収録されているリニアPCM 5.1chサラウンドから、音声編集ソフトDaVinci Resolve(Blackmagic Design製)を用い、モノラルとステレオをリミックスした。

ラウドネス補正はITU-R BS.1770^[16]に従って行ない、音圧レベル値として、映画館を参考に最小:55db、最大:85dbとなるように設定した。

3.4 実験参加者

実験参加者は20代の人物15名(男性8名、女性7名)であり、映像視聴に支障がない視力(矯正視力を含む)と聴力を有していた。

本実験は、東海大学「人を対象とする研究」に関する倫理委員会規定に従い実施した。

3.5 視線計測方法

アイマークレコーダEMR-9(ナックイメーテックテクノロジー製)と加速度センサ3DM-GX4-25(LORD MicroStrain製)を用い、眼球運動と頭部運動を計測した。サンプリング周波数は60Hzであった。

実験参加者は帽子モデルのEMR-9を被り、その後頭部に3DM-GX4-25を固定バンドで装着した。

3.6 映像視聴方法

実験参加者には、「自由に眼球と頭部を動かして見てよい」と指示をして映像を視聴させた。

3つの音響条件と6つのシーンからなる計18シーンを無作為に提示し、各シーンの間に10秒間の視線固定画像を挿入した。18シーンの視聴は、約5分間の休憩時間を挟んで2部に分けて、実施した。1シーンの長さは1分間で、各シーンの先頭に10秒間の視線固定映像を挿入した。そのため、前半と後半の9シーンの視聴時間は各々10分30秒である。この際、全実験参加者が異なる順序で各シーンを視聴した。

3.7 視線分析手法

本実験では、各音響条件間で注視点の分布の広がりの違いを検証する。

この注視点の分布の広がりを表す指標として、眼球の動かしやすさ及び、視野範囲が水平方向と垂直方向で異なることから^[17]、それぞれの標準偏差を分離して検証する。さらに、Yamadaら^[18]が注視点の広がりを楕円として表現したことになり、水平と垂直の標準偏差および視距離に基づいて計算される楕円の面積を評価する。

最初に、計測した視線データから、移動速度に基づいて注視点定義し、抽出した^[19]。注視点抽出の基準として山田ら、福田らの研究^{[19],[20]}を参考に30°/秒以下のデータが0.15秒以上続いた区間を注視点候補として抽出した。そして、基準を満たす視線座標位置の平均値をその注視点候補区間の代表点として定義し、1つの注視点とした。計測装置の仕様上、瞳孔を検出できない瞬きなどの状態の区間はエラー値として除外した。

次に、実験参加者がある音響条件でシーンを視聴している間に抽出された注視点の平均的な位置を注視点の平均として算出し、その注視点の標準偏差を水平方向と垂直方向のそれぞれについて算出した。この水平と垂直方向の注視点の各標準偏差が長軸と短軸の半径となる楕円を注視点領域面積 a とし、その面積を式(1)で算出した。

$$a = \pi d^2 \tan(2x) \tan(2y). \quad (1)$$

ここで、水平方向と垂直方向の注視点の標準偏差をそれぞれ x [°]と y [°]とし、視距離を d [cm]、 π は円周率を表す。

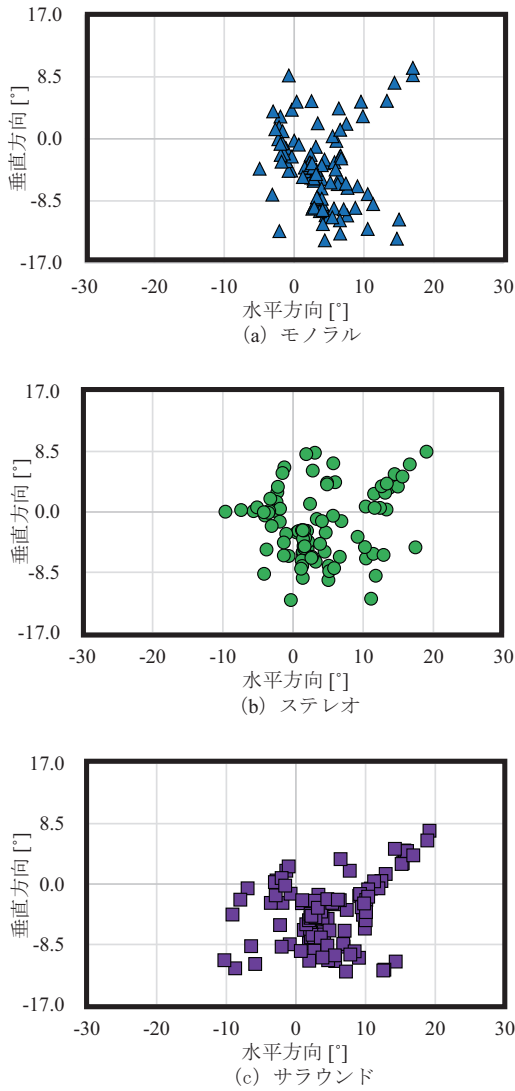


図2 各音響条件における注視点分布 (Valley C)
Fig.2 Gaze point distribution in each sound condition (Valley C).

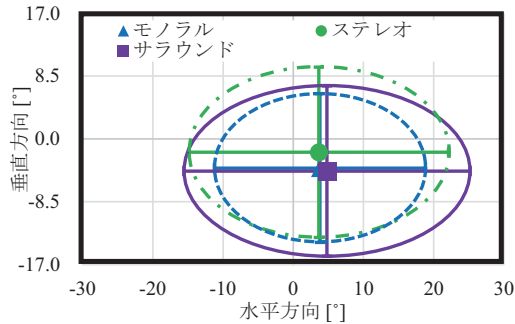


図3 各音響条件における平均注視点と垂直・水平方向の標準偏差 (2σ) (Valley C)
Fig.3 Average gaze point and horizontal and vertical standard deviations (2σ) in each sound condition (Valley C).

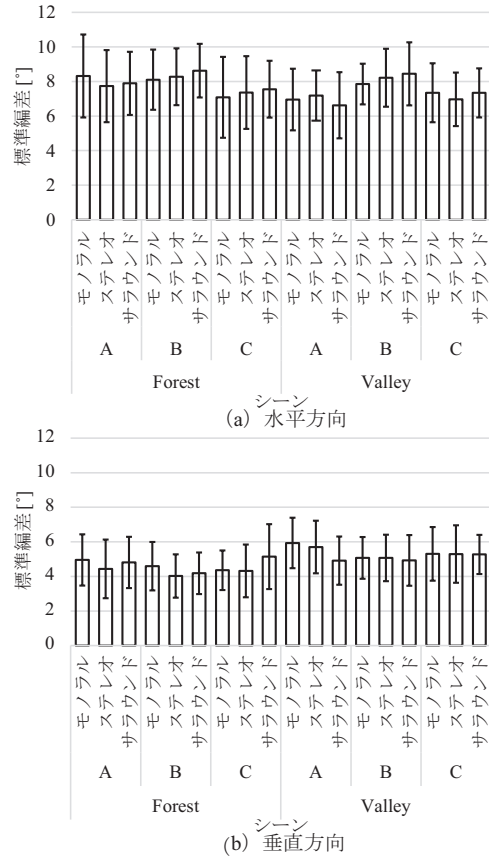


図4 標準偏差の平均

Fig.4 Average standard deviations.

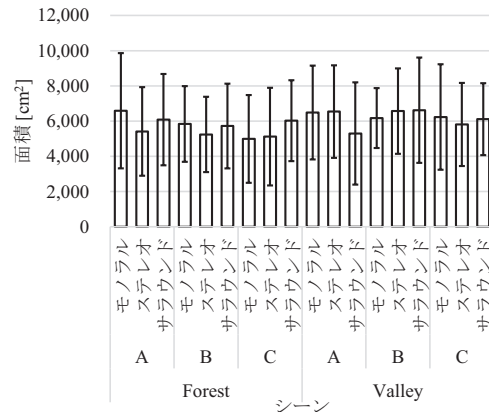


図5 注視領域の平均面積

Fig.5 Average area of gaze region.

4. 実験結果と分析

4.1 注視点分布

図2に、各音響条件で Valley C のシーンを視聴した際の実験参加者の注視点の分布を例として示す。図中の外枠 (水平方向 $\pm 30^\circ$, 垂直方向 $\pm 17^\circ$) はスクリーン中心座標を (0,0) とした際の大画面スクリーンの観視画角 $[\circ]$ を示している。

この実験参加者については、スクリーンの対角方向に広い分布となり、サラウンドにおいて、他の条件よ

りも対角方向に広く分布する結果が示された。

4.2 標準偏差と注視領域面積

前節に示した結果から、音響条件による違いをさらに分析するために、図2に示した実験参加者データの注視点の平均と水平および垂直方向の標準偏差 (2σ) を算出する(図3)。注視点分布を比較する指標として、標準偏差と注視領域の面積 a に着目する。

具体的には注視点の標準偏差から、図中の破線(モノラル)、一点鎖線(ステレオ)、実線(サラウンド)で示す楕円を注視領域と定義し、その面積を算出する。

この実験参加者の Valley C における注視領域面積は、モノラル: 3,694 cm²、ステレオ: 4,625 cm²、サラウンド: 4,586 cm² となった。ステレオとサラウンドにおいて、モノラルと比較して面積が大きくなった一方で、ステレオとサラウンドの音響条件間では面積に大きな違いが見られなかった。

一方、他の実験参加者について分析すると、実験参加者によって標準偏差と注視領域面積の傾向は異なり、ステレオとサラウンドでは仮説に反して注視点分布を狭める場合もあった。図4と図5に、全実験参加者で平均したシーン別の水平と垂直方向の標準偏差、注視領域面積を示す。

4.2.1 2要因分散分析

注視点の標準偏差と注視領域面積についてさらに分析するため、シーン条件と音響条件を要因とする2要因分散分析を行なった。シーンや音響の効果及び交互作用を明らかにすることによって、音と視覚の関係性を検証する。

本実験ではシーンが持つ固有の映像特徴の影響を考慮し、シーン条件として森林、溪谷それぞれでバリエーションを増やし、カットが異なるシーンを3つずつ用意している。シーンに起因する影響を確認するため、用意した映像のバリエーションである6水準での分析を補助的に行なう。その結果を表1に示す。

注視点の水平方向の標準偏差ではシーン条件の主効果、垂直方向の標準偏差ではシーン条件と音響条件の主効果と交互作用、注視領域面積では音響条件の主効果が有意であった。この表のF値は、群間および群内の平均二乗の比として計算された。この値に基づいて、F分布表から有意水準5%で有意差があるかを判定した。

4.2.2 多重比較

前項の結果から、有意であった項目に対してHolm法による多重比較を行なった。

初めに、水平方向の標準偏差において主効果があったシーン条件の組を表2に示す。有意差があった組のうち有意に大きかった組を太字で示す。Forest B と Valley B の標準偏差が8.0を超えており、これらのシーン条件を含む一部の組で有意差がみられた。

次に、注視領域面積において主効果があった音響条件の組を表3に示す。有意差があった組のうち有意に大きかった組を太字で示す。モノラルとステレオ、モノラルとサラウンドの音響条件間で有意差があり、チャンネル数の増加により有意に分布が広がるという結果が示された。

垂直方向の標準偏差では、シーン条件と音響条件の主効果と交互作用があったため、まず単純主効果の結果を表4に示す。各シーン条件における音響条件に対して、すべてのシーン条件における音響条件間で有意差があった。各音響条件におけるシーン条件に対しては、ステレオにおいてのみシーン条件間で有意差があった。

続いて、Holm法における多重比較を、有意差があった組について行なった結果を表5に示す。有意差があった組のうち有意に大きかったものを太字で示す。各シーン条件における音響条件では、モノラルとステレオの音響条件間とモノラルとサラウンドの音響条件間で有意差があったものが、全6シーン中3シーンであった。ステレオとサラウンドの音響条件間で有意差がみられたのは、全6シーン中2シーンで、そのうち1つはステレオの方が大きいという結果であった。ステレオにおけるシーン条件では、全15組中5組で有意差がみられ、そのうち3組でForest Bが有意に小さいことを示す結果となった。

5. 考察

4章では、実験で計測した視線データから注視点を抽出し、その水平方向と垂直方向の標準偏差とその標準偏差から定義される注視領域面積を算出した。その結果、音響条件における注視領域面積の有意差を示した。

ここでは、2要因分散分析の結果から、各音響条件における標準偏差および注視領域面積の違いについて議論する。

初めに、水平方向の標準偏差、次に垂直方向の標準偏差、最後に注視領域面積について述べる。

5.1 水平方向の標準偏差

水平方向の標準偏差では、シーン条件間で主効果が認められたが、音響条件間では主効果が認められなかった(表1: 標準偏差(水平方向))。

サラウンドスピーカは、ITU-R BS.775-1^[15]に従って配置した。その内、リアスピーカは耳の高さに設置されており、目線の位置とほぼ同じ高さである。このため、注意が引きつけられる場合には、垂直方向よりも水平方向に眼球運動が多く生じる可能性があったが、本実験において、音響条件間では主効果が認められなかった。

この要因として、配置されたリアスピーカから強く

表1 標準偏差（水平・垂直）と注視領域面積の2要因（シーン・音響）分散分析結果
Table 1 Two factor (scene and sound) variance analysis results for standard deviation (horizontal and vertical) and area of gaze region.

| | | F 値 | 有意差 ($p < 0.05$) |
|----------------|-----------|----------------------|-----------------------|
| 標準偏差 (水平方向) | 主効果 (シーン) | $F(5, 70) = 8.673$ | あり |
| | 主効果 (音響) | $F(2, 28) = 2.018$ | なし |
| | 交互作用 | $F(10, 140) = 1.565$ | なし |
| 標準偏差 (垂直方向) | 主効果 (シーン) | $F(5, 70) = 3.449$ | あり |
| | 主効果 (音響) | $F(2, 28) = 20.306$ | あり |
| | 交互作用 | $F(10, 140) = 3.966$ | あり |
| 注視領域面積 | 主効果 (シーン) | $F(5, 70) = 1.565$ | なし |
| | 主効果 (音響) | $F(2, 28) = 6.406$ | あり |
| | 交互作用 | $F(10, 140) = 1.115$ | なし |

表2 シーン条件の組における多重比較結果（水平方向の標準偏差）
Table 2 Multiple comparison test results for pairs of scene conditions (Horizontal standard deviation).

| シーン条件の組 (平均値) | t 値 | 有意差 ($p < 0.05$) |
|--|-----------------|-----------------------|
| Forest A (7.799) - Forest B (8.316) | $t(14) = 1.970$ | なし |
| Forest A (7.799) - Forest C (7.202) | $t(14) = 2.259$ | なし |
| Forest A (7.799) - Valley A (7.104) | $t(14) = 2.671$ | なし |
| Forest A (7.799) - Valley B (8.189) | $t(14) = 1.571$ | なし |
| Forest A (7.799) - Valley C (7.350) | $t(14) = 2.094$ | なし |
| Forest B (8.316) - Forest C (7.202) | $t(14) = 4.344$ | あり |
| Forest B (8.316) - Valley A (7.104) | $t(14) = 4.276$ | あり |
| Forest B (8.316) - Valley B (8.189) | $t(14) = 0.465$ | なし |
| Forest B (8.316) - Valley C (7.350) | $t(14) = 4.705$ | あり |
| Forest C (7.202) - Valley A (7.104) | $t(14) = 0.396$ | なし |
| Forest C (7.202) - Valley B (8.189) | $t(14) = 4.951$ | あり |
| Forest C (7.202) - Valley C (7.350) | $t(14) = 0.868$ | なし |
| Valley A (7.104) - Valley B (8.189) | $t(14) = 4.395$ | あり |
| Valley A (7.104) - Valley C (7.350) | $t(14) = 0.868$ | なし |
| Valley B (8.189) - Valley C (7.350) | $t(14) = 3.875$ | あり |

表3 音響条件の組における多重比較結果（注視領域面積）
Table 3 Multiple comparison test results for pairs of sound conditions (Area of gaze region).

| 音響条件の組 (平均値) | t 値 | 有意差 ($p < 0.05$) |
|---------------------------------|-----------------|-----------------------|
| モノラル - ステレオ (5,530) (6,118) | $t(14) = 3.179$ | あり |
| モノラル - サラウンド (5,530) (6,155) | $t(14) = 2.570$ | あり |
| ステレオ - サラウンド (6,118) (6,155) | $t(14) = 0.251$ | なし |

表4 各要因における条件間の単純主効果検定の結果（垂直方向の標準偏差）
Table 4 Simple main effect test results between conditions for each factor (Vertical standard deviation).

| 要因 | 条件 | F 値 | 有意差 ($p < 0.05$) |
|--|-----|--------------------|-----------------------|
| Forest A Forest B Forest C Valley A Valley B Valley C | 音響 | $F(2, 28) = 7.971$ | あり |
| | 音響 | $F(2, 28) = 3.652$ | あり |
| | 音響 | $F(2, 28) = 4.573$ | あり |
| | 音響 | $F(2, 28) = 9.062$ | あり |
| | 音響 | $F(2, 28) = 7.178$ | あり |
| | 音響 | $F(2, 28) = 8.919$ | あり |
| モノラル ステレオ サラウンド | シーン | $F(5, 70) = 2.263$ | なし |
| | シーン | $F(5, 70) = 6.484$ | あり |
| | シーン | $F(5, 70) = 2.092$ | なし |

注意を引くような音が発生しなかったことが考えられる。リアスピーカの左右 110° の位置はスクリーンの外側の視野外となる。このため、そのリアスピーカの方へ視線移動を生じさせるためには、視対象が見えなくとも強く注意を引く必要があることが予想される。しかし本実験で使用した風景シーン中には、そのような音の発生がなかった可能性が考えられる。このことから、リアスピーカから鳴る音に対して注意を引きつけられることによる水平方向の眼球運動があまり生じず、視線移動する必要が生じなかったと考えられる。

カメラの視点が固定された俯瞰風景シーンを抽出したことから、すべてのシーン条件で同じような視線行動が生じる可能性が高いと考えていた。しかし、多重比較の結果から、特定のシーン条件 (Forest B, Valley B) において、水平方向の標準偏差が有意に広がっていることが確認されたため、実際にはシーン条件の影響があったことが考えられる。

サラウンド映像視聴時の注視点分布の分析

表 5 各要因における単純主効果が有意であった条件内での多重比較の結果（垂直方向の標準偏差）

Table 5 Multiple comparison test results within conditions in which the simple main effect of each factor was significant (Vertical standard deviation).

| 要因 | 条件の組（平均値） | t 値 | 有意差 ($p < 0.05$) |
|----------|--|-----------------|-----------------------|
| Forest A | モノラル (4.953) - ステレオ (4.808) | $t(14) = 0.619$ | なし |
| | モノラル (4.953) - サラウンド (5.701) | $t(14) = 3.139$ | あり |
| | ステレオ (4.808) - サラウンド (5.701) | $t(14) = 3.616$ | あり |
| Forest B | モノラル (4.592) - ステレオ (4.181) | $t(14) = 1.259$ | なし |
| | モノラル (4.592) - サラウンド (5.069) | $t(14) = 2.496$ | なし |
| | ステレオ (4.181) - サラウンド (5.069) | $t(14) = 2.605$ | なし |
| Forest C | モノラル (4.356) - ステレオ (5.145) | $t(14) = 2.306$ | なし |
| | モノラル (4.356) - サラウンド (5.295) | $t(14) = 2.701$ | なし |
| | ステレオ (5.145) - サラウンド (5.295) | $t(14) = 0.485$ | なし |
| Valley A | モノラル (4.435) - ステレオ (5.935) | $t(14) = 4.418$ | あり |
| | モノラル (4.435) - サラウンド (4.912) | $t(14) = 1.217$ | なし |
| | ステレオ (5.935) - サラウンド (4.912) | $t(14) = 2.956$ | あり |
| Valley B | モノラル (4.024) - ステレオ (5.071) | $t(14) = 3.307$ | あり |
| | モノラル (4.024) - サラウンド (4.926) | $t(14) = 3.490$ | あり |
| | ステレオ (5.071) - サラウンド (4.926) | $t(14) = 0.453$ | なし |
| Valley C | モノラル (4.318) - ステレオ (5.304) | $t(14) = 4.570$ | あり |
| | モノラル (4.318) - サラウンド (5.271) | $t(14) = 2.973$ | あり |
| | ステレオ (5.304) - サラウンド (5.271) | $t(14) = 0.134$ | なし |
| ステレオ | Forest A (4.808) - Forest B (4.181) | $t(14) = 2.129$ | なし |
| | Forest A (4.808) - Forest C (5.145) | $t(14) = 0.899$ | なし |
| | Forest A (4.808) - Valley A (5.935) | $t(14) = 3.620$ | あり |
| | Forest A (4.808) - Valley B (5.071) | $t(14) = 0.837$ | なし |
| | Forest A (4.808) - Valley C (5.304) | $t(14) = 1.985$ | なし |
| | Forest B (4.181) - Forest C (5.145) | $t(14) = 2.400$ | なし |
| | Forest B (4.181) - Valley A (5.935) | $t(14) = 4.624$ | あり |
| | Forest B (4.181) - Valley B (5.071) | $t(14) = 3.664$ | あり |
| | Forest B (4.181) - Valley C (5.304) | $t(14) = 5.038$ | あり |
| | Forest C (5.145) - Valley A (5.935) | $t(14) = 2.911$ | なし |
| | Forest C (5.145) - Valley B (5.071) | $t(14) = 0.242$ | なし |
| | Forest C (5.145) - Valley C (5.304) | $t(14) = 0.398$ | なし |
| | Valley A (5.935) - Valley B (5.071) | $t(14) = 3.664$ | あり |
| | Valley A (5.935) - Valley C (5.304) | $t(14) = 1.576$ | なし |
| | Valley B (5.071) - Valley C (5.304) | $t(14) = 0.748$ | なし |

その要因の1つとして、シーンの画像特徴の影響が音よりも強く出ていた可能性を示唆している。その影響の例として、Forest（森林）の映像では木の太い幹や葉っぱの揺らめき、Valley（溪谷）の映像では川のの流れや落ち方などが挙げられ、視線が誘導されやすい画像特徴が存在することが考えられる。

このシーン条件の違いによる視線の違いを検証するためには、映像そのものの顕著性や視対象の数など、映像自体のパラメータを詳しく分析する必要がある。音響条件に着目した本論文ではその要因を明らかにすることが目的ではないため、分析しなかったが、この問題の検証は今後の課題の1つである。

5.2 垂直方向の標準偏差

垂直方向の標準偏差では、音響条件間で主効果が認められた（表1：標準偏差（垂直方向））。

5.1節でも示したように、サラウンドスピーカの配置はITU-R BS.775-1^[15]に従って配置した。その内、フロントスピーカは、実験室のスクリーンが非透過型

であるためスクリーンの真下（実験参加者の膝の高さ付近）とした。このため、注意が引きつけられる場合には、水平方向よりも垂直方向に眼球運動が多く生じる可能性があった。本実験の結果より、リアスピーカに対しては発生しなかった強く注意を引くような音がスクリーン真下のフロントスピーカから流れ、実験参加者の視線を垂直方向に移動を促した可能性が考えられる。また、フロントスピーカは映像と近接しており視野内の位置にある。このため、スピーカ間で同じような音が発生していても、リアスピーカよりフロントスピーカの方が視線移動を促しやすい配置だったと考えられる。

2章の関連研究では、サラウンド音響環境として3画面で29個のスピーカを球面状に5段階の高さに設置したもの^[13]、3画面で7.1.4chのフロントスピーカ3つ、リアスピーカ2つ、バックスピーカ2つ、サブウーファ1つ、天井スピーカ4つの計12個のスピーカを設置したもの^[14]で、実験参加者へ人の数を数え

るタスクを課していた。タスクの影響により注視分布が1点に集まりやすく、周囲環境音の影響を受けやすかったことや、視野外に聴覚刺激となるスピーカが多く配置されたことで、視野外から視覚的注意を引きやすい状況であったことが考えられる。一方で、本実験では、一般的なテレビ視聴を想定したタスクを課さない自由視聴で、視野内に収まる1画面分の画角、視野外のスピーカはリアスピーカのみであった。このことから、関連研究とは異なり、視野外からは注意を引き寄せることが少なかったことが考えられる。

5.1節と同様に、垂直方向においても多重比較の結果(表5)から、シーン条件により有意差が生じた音響条件が異なることが確認された。ステレオが有意に大きかったシーン(Valley A)、サラウンドが有意に大きかったシーン(Forest A)、ステレオ・サラウンドが有意に大きかったシーン(Valley B, Valley C)があり、シーン条件の影響があったことが考えられる。特に、ステレオシーンではシーン条件間の単純主効果が認められており、シーンに依存する影響が強かったことが考えられる。

5.1節でも述べたように、その要因の1つとして、シーンの画像特徴の影響が音よりも強く出ていた可能性が示唆されるが、映像そのものの特徴の分析は今後の課題である。

5.3 注視領域面積

注視領域面積では、音響条件間で主効果が認められ、多重比較の結果から、モノラルとステレオ、モノラルとサラウンドの音響条件間で有意差がみられた(表1:注視領域の面積,表3)。

注視領域面積は、水平方向と垂直方向の標準偏差から算出されるため、垂直方向の違いが主となり、面積において音響条件間で有意差が得られたと考えられる。また、5.1節で示唆されたように、後方に強く注意を払う必要がなかったことが、ステレオとサラウンドの音響条件間に有意差が見られなかった要因と考えられる。以上のことから、モノラルと比較して、ステレオおよびサラウンドにおいて注視点分布が広がることを示し、音響条件によって視線が変化し、注視領域面積が大きくなることを明らかにした。

本実験で使用した自然映像のように、画像全体の傾向が似ている場合には、特定の注視対象が決まりにくく、ドラマやスポーツなどで生じやすい特定の視対象への注視が発生しにくい。そのため、スピーカから発生する音に対する視線移動を促せたと考えられる。このことから、自然映像ジャンルでなくとも、様々な視対象を配置した画面を満遍なく見せたい映像に対して適用できる可能性がある。

また、スピーカの配置はITU-R BS.775-1^[15]に従っており、一般的な映画やドラマなどのコンテンツでの

音響制作現場ではこの基準位置での制作が多く行なわれている。このため、物理的に配置できない場合を除き、この位置から大きく変更することはほとんどないと考えられる。

以上のことから、注意を強く引く注視対象物が少ない映像や様々な視対象を配置した映像かつITU-R BS.775-1^[15]に従ったスピーカ配置の状況下で、画面を満遍なく見せたい映像制作者への指標の1つとなる結果を示すことができたと考えている。

6. まとめ

本論文では、大画面・サラウンド音響において、映像視聴時の人の生体反応の1つである視線に着目し、モノラル、ステレオ、サラウンドの音響条件で映像を視聴した際の視聴者の視線を計測した。その分析および検定結果から、ステレオおよびサラウンドの音響条件において、モノラルと比較して注視領域面積が有意に広がり、注視点分布が広がることを確認した。本実験のような自然の風景映像ではステレオとサラウンドの音響条件間では注視点分布の広がりにより有意差が認められないことを実験的に確認した。このことから、サラウンド音響により、モノラルと比較して注視点の分布が広がるという仮説を支持する結果が示された。一方で、ステレオとの比較では仮説を支持しない結果が示された。

モノラルとの比較で仮説を支持した要因としては、スクリーン真下から生じるフロントスピーカが垂直方向の視線移動を生じさせたことが考えられる。ステレオとサラウンドの比較で仮説を支持しなかった要因としてはリアスピーカから強く注意を引くような音が発生せず、注意を引きつけられることによる水平方向の眼球運動があまり生じず、視線移動する必要が生じなかったことが考えられる。

本実験の結果より、本論文の目的である、「人の視覚に対する聴覚の作用のメカニズムを解明するための基礎知識となる情報を得る」、「音と視覚の関係性を明らかにする」について、ステレオ、サラウンドのような適切な音響環境状況の映像視聴時において、垂直方向注視点の標準偏差について音響条件間で違いが見られたことから、視覚に対する聴覚の作用が存在すること、音と視覚に関係性があることを確認した。

この結果から視聴者が映像を見る際に、視線移動を促すスピーカ配置による映像視聴環境を構築することにより、映像内の情報を満遍なく取り込むために、画面全体に広く動かすような意味のある視線移動を誘導させる。このことから、視聴者のより広範囲からの視覚情報の取得に寄与する効果があると考えており、映像コンテンツ作成への指針の提供と視聴者の映像視聴体験の向上に寄与するものであると考えている。

今後の課題は、音響条件に加え、映像視聴時に関連が深い要因との相互作用を検証することであり、例としては映像中の物体移動や顕著性などの映像効果要因や主観評価との関係を視線の動きから明らかにすることである。また、本実験で違いが見られなかったステレオとサラウンドの関係について、サラウンドを効果的に活用する方法を検討することも課題である。これらの結果が、映像制作やより良い映像視聴環境構築の礎となると考えている。

利益相反の開示

本論文に関して、報告すべき利益相反関連事項はない。

謝辞

本研究の一部は、JSPS 科研費 JP17H00745 の助成を受けたものである。

参考文献

- [1] 小峯一晃：生体情報を用いた映像評価技術, NHK 技研 R&D, No.159, pp.13–22 (2016). (Komine, K.: Video evaluation technology using biological information, NHK STRL R&D, No.159, pp.13–22 (2016) (in Japanese)).
- [2] Takahira, H., Kikuchi, K. and Yamada, M.: A system for measuring gaze movement and hand movement simultaneously for hand-held devices, IEICE Transactions on Communications, Vol.E98-B, No.1, pp.51–61 (2015).
- [3] Hatada, T., Sakata, H. and Kusaka, H.: Psychophysical analysis of the ‘sensation of reality’ induced by a visual wide-field display, SMPTE Journal, Vol.89, No.8, pp.560–569 (1980).
- [4] Mochiduki, S., Takahira, H. and Yamada, M.: Analysis of gazing points while viewing super-high-definition images at various viewing positions, Proceedings of the 2015 International Conference on Computer Application Technologies, pp.54–57 (2015).
- [5] Mochiduki, S., Watanabe, R., Takahira, H. and Yamada, M.: Analysis of head movement during gaze movement with varied viewing distances and positions, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E101-A, No.6, pp.892–899 (2018).
- [6] Fang, Y., Emoto, M., Nakashima, R., Matsumiya, K., Kuriki, I. and Shioiri, S.: Eye-position distribution depending on head orientation when observing movies on ultrahigh-definition television, ITE Transactions on Media Technology and Applications, Vol.3, No.2, pp.149–154 (2015).
- [7] Corneil, D. B. and Munoz, P. D.: The influence of auditory and visual distractors on human orienting gaze shifts, Journal of Neuroscience, Vol.16, No.24, pp.8193–8207 (1996).
- [8] Rorden, C. and Driver, J.: Does auditory attention shift in the direction of an upcoming saccade?, Neuropsychologia, Vol.37, No.3, pp.357–377 (1999).
- [9] Shomstein, S. and Yantis, S.: Control of attention shifts between vision and audition in human cortex, Journal of Neuroscience, Vol.24, No.47, pp.10702–10706 (2004).
- [10] Jack, E. C. and Thurlow, R. W.: Effects of degree of visual association and angle of displacement on the ‘ventriloquism’ effect, Perceptual and Motor Skills, Vol.37, No.3, pp.967–979 (1973).
- [11] Sato, J., Fukue, K., Kinoshita, Y. and Ozawa, K.: Relationship between gaze direction and sound localization in ventriloquism effect, Acoustical Science and Technology, Vol.32, No.1, pp.40–42 (2011).
- [12] Komiyama, S.: Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems, Journal of the Audio Engineering Society, Vol.37, No.4, pp.210–214 (1989).
- [13] Mendonça, C., Rummukainen, O. and Pulkki, V.: 3D sound can have a negative impact on the perception of visual content in audiovisual reproductions, Proceedings of the 21st International Conference on Auditory Display, pp.138–143 (2015).
- [14] Korshunova, V., Remijn, B. G., Elischka, S. and Mendonça, C.: The impact of sound systems on the perception of cinematic content in immersive audiovisual productions, Proceedings of the 12th IEEE Asia Pacific Workshop on Mixed and Augmented Reality, pp.1–5 (2019).
- [15] ITU-R BS.775-1: Multichannel stereophonic sound system with and without accompanying picture (1994).
- [16] ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level (2015).
- [17] 畑田豊彦：生理光学 (16) 自然視画像と視覚特性, O plus E, Vol.74, pp.121–130 (1986). (Hatada, T.: Physiological optics. 16. Images in usual vision and vision characteristics., O plus E, Vol.74, pp.121–130 (1986) (in Japanese)).
- [18] Yamada, M. and Fukuda, T.: Quantitative evaluation of eye movements as judged by sight-line displacements, SMPTE Journal, Vol.95, No.12 pp.1230–1241 (1986).
- [19] 山田光穂, 福田忠彦:画像における注視点の定義と画像分析への応用, 電子情報通信学会論文誌 (D), Vol.J69-D, No.9, pp.1335–1342 (1986). (Yamada, M. and Fukuda, T.: Definition of gazing point for picture analysis and its applications, IEICE Transactions on Information and Systems (Japanese Edition), Vol.J69-D, No.9, pp.1335–1342 (1986) (in Japanese)).
- [20] 福田亮子, 佐久間美能留, 中村悦夫, 福田忠彦:注視点の定義に関する実験的検討, 人間工学, Vol.32, No.4, pp.197–204 (1996). (Fukuda, R., Sakuma, M., Nakamura, E. and Fukuda, T.: An experimental consideration on the definition of a fixation point, Japanese Journal of Ergonomics, Vol.32, No.4, pp.197–204 (1996) (in Japanese)).

(2022年2月19日受付, 5月30日再受付)

著者紹介

高比良 英朗



2014年東海大学情報通信学部卒業。2016年同大学大学院情報通信学研究科修士課程修了。2020年名古屋大学大学院情報科学研究科博士後期課程満期退学, 現在, 同大学大学院情報学研究科協力研究員。映像メディア, 視線計測・分析の研究に従事。電子情報通信学会, 電気学会各会員。

篠原 未歩



2020年東海大学情報通信学部卒業。2022年同大学大学院情報通信学研究科修士課程修了。現在、パナソニックコネクタ株式会社に勤務。在学中、高精細映像の癒し効果に着目した人間の生理的特性の研究に従事。

野坂 祐介



2020年東海大学情報通信学部卒業。2022年同大学大学院情報通信学研究科修士課程修了。在学中、視覚科学、特に絵画を見る眼球運動についての研究に従事。

横内 雅也



2021年東海大学情報通信学部情報メディア学科卒業。現在、株式会社あつまDXソリューション事業部プロデューサー。在学中、サラウンド音声映像視聴時の視線の動きに与える影響についての研究に従事。

山田 光穂 (正会員)



1978年名古屋大学工学部卒業。1980年同大学大学院工学研究科修士課程修了。同年NHK入局。同放送技術研究所を経て、2003年より東海大学に在籍。現在、同大学院情報通信学研究科教授。眼球運動の分析、視覚情報処理、次世代マルチメディアの研究に従事。電子情報学会、IEEE各シニア会員、映像情報メディア学会、ヒューマンインタフェース学会、視覚学会、パーソナルコンピュータ利用技術学会各会員。工学博士。

平山 高嗣



2005年大阪大学大学院基礎工学研究科博士後期課程修了。博士(工学)。同年京都大学大学院情報学研究科特任助教。2011年名古屋大学大学院情報科学研究科特任助教。2012年同助教。2014年同特任准教授。2017年同大学未来社会創造機構特任准教授。2021年人間環境大学人間環境学部環境科学科教授。現在に至る。顔画像認識、注視行動分析、視覚的注意の計算モデル、マルチモーダルインタラクションに関する研究に従事。電子情報通信学会、情報処理学会、IEEE、ACM各会員。

川西 康友



2006年京都大学工学部卒業。2012年同大学大学院情報学研究科博士後期課程修了。博士(情報学)。2012年同大学特定研究員。2014年名古屋大学特任助教。2015年同助教。2020年同講師。2021年理化学研究所感覚データ認識研究チームチームリーダー。現在に至る。ロボットによる周囲環境認識及び、人物追跡・属性認識・行動認識などの人物画像処理に関する研究に従事。IEEE、画像電子学会各会員、電子情報通信学会シニア会員。

出口 大輔



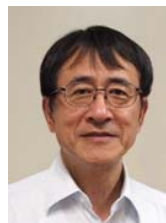
2001年名古屋大学工学部卒業。2006年同大学院情報科学研究科博士後期課程修了。博士(情報科学)。2004年(～2006年)まで日本学術振興会特別研究員。2006年名大大学院工学研究科研究員。2008年同大学院情報科学研究科助教。2012年同大情報連携統括本部情報戦略室准教授。2020年同大学院情報学研究科准教授。現在に至る。主に画像処理・パターン認識技術の開発とそのITS応用に関する研究に従事。電子情報通信学会、情報処理学会、IEEE各会員。

井手 一郎



1994年東京大学工学部卒業。2000年同大学工学系研究科電気工学専攻博士課程修了。博士(工学)。同年国立情報学研究所助手。2004年名古屋大学大学院情報科学研究科助教。2020年同大学数理・データ科学教育研究センター教授。2022年同大学大学院情報学研究科教授。現在に至る。この間、2010～2011年オランダアムステルダム大学情報学研究所上級訪問研究員。電子情報通信学会、情報処理学会、IEEE各シニア会員、映像情報メディア学会、人工知能学会、ACM各会員。

村瀬 洋



1978年名古屋大学工学部卒業。1980年同大学大学院修士課程修了。同年日本電信電話会社(現NTT)入社。1992年から1年間米国コロンビア大客員研究員。2003年名古屋大学大学院情報科学研究科教授。2021年同名誉教授。現在に至る。文字・図形・画像認識、コンピュータビジョンの研究に従事。工学博士。1994年IEEE CVPR最優秀論文賞、2004年IEEE Trans. MM論文賞、2012年紫綬褒章、他受賞。IEEE、電子情報通信学会、情報処理学会各フェロー。