

〔研究ノート〕

## ドローンムービー世界大会応募作品における 映像シーンと音響シーンの整合性調査の試み

梨子木 快晴\*, 沖田 和久\*<sup>2</sup>, 大里 一矢\*<sup>3</sup>, 河納 隼一\*<sup>4</sup>, 近藤 善隆\*<sup>5</sup>, 福島 学\*

\*日本文理大学工学部情報メディア学科

\*<sup>2</sup>日本文理大学大学院工学研究科環境情報学専攻

\*<sup>3</sup>日本文理大学大学院工学研究科環境情報学専攻 (2021年度卒業)

\*<sup>4</sup>株式会社アーネット

\*<sup>5</sup>株式会社ジェイテック

### An Investigation on the Consistency of Video and Audio Scenes in Drone Movie World Championship Entrée

Kaisei NASHIKI\*, Kazuhisa OKITA\*<sup>2</sup>, Kazuya OSATO\*<sup>3</sup>, Syun'ichi KAWANO\*<sup>4</sup>,  
Yoshitaka KONDO\*<sup>5</sup> and Manabu FUKUSHIMA\*

\*Department of Media Technologies, School of Engineering, Nippon Bunri University

\*<sup>2</sup>Department of Environmental Engineering and Applied Information Science,  
Graduate School of Engineering, Nippon Bunri University

\*<sup>3</sup>Department of Environmental Engineering and Applied Information Science,  
Graduate School of Engineering, Nippon Bunri University (Graduate, AY2021)

\*<sup>4</sup>Artnet Co., Ltd.

\*<sup>5</sup>J-TEC Co., Ltd.

### 1. はじめに

映像刺激により想起される印象と、音刺激により想起される印象の違和感が少ないことが、伝えたい印象を表現するのに重要である。しかし、表現したい印象を映像刺激または音刺激として記述したとしても相手が想起する印象と一致するか<sup>1)2)</sup>、さらに映像と音で同じ印象が記述できるかについて研究が十分行われているとは言い難い。特に作品という側面から考えると数値評価<sup>3)4)5)6)</sup>が適さないものと言われることもある。

一方、動画コンテンツの工学的側面で考えると、コンテンツの品質に対する質の保証は必要であり、このた

めの数値化は重要である。特にアクセシビリティ JIS<sup>7)</sup>のように Web でアクセス可能なコンテンツに対して、利用者に対する配慮を考えると、今後動画コンテンツにもガイドラインが出来ることが予想される。

本研究では動画コンテンツの1つであり、世界大会 (Drone Movie World Championship<sup>8)</sup>) で競う対象となっているドローンで撮影された映像とそれに付加する楽音に着目する。ドローンによる撮影映像は、世界大会のレギュレーションとして山間部にある施設を中心としたものであり、空間的印象がその中心となっている。音刺激は両耳の相関により想起する空間的印象を知ることが出来る<sup>9)</sup>。すなわち、映像刺激から受ける印象と音刺激から受ける印象を合わせる必要がある。そこ

で、本研究では、音刺激の両耳間相関から音刺激により想起される印象の切り替わりを、入賞作品と入賞しなかった作品を使用して調べる。

## 2. 対象映像コンテンツと映像シーンチェンジ

使用した作品は、入賞しなかった作品（以下、動画1）、入賞した作品（以下、動画2、動画3）とする。はじめに映像の1フレーム画像から2次元スペクトルを求め、スペクトルパタンが一定値以上の変化となった箇所をシーンチェンジとする。動画1におけるシーンチェンジの検出結果を図1に示す。図には、シーンチェンジ前（Before）とシーンチェンジ後（After）の画像と2次元スペクトルを示す。ここではカラー画像をグレースケールに変換し2次元スペクトルを求めている。

検出されたシーンチェンジが、動画1で3シーン、動画2で2シーン、動画3で4シーンであった。動画1の画像と2次元スペクトルを図1から図3、動画2を図4と図5、動画3を図6から図9に示す。

## 3. 相関係数と想起される空間印象

両耳に同一の音刺激が提示されると、両耳で発火した神経パルスが同時性検出器<sup>10)</sup>で同時であることを示し正中面に1つの音像が結像する。このような両耳の信号の相関係数は1となる。音像が右方向に移動するには、音像に近い日向耳である右耳の音圧が影耳である左耳より高くなり、左耳に到来する時刻は右耳よりも遅れる。これが音像定位の原理である。両耳の相関係数が0すなわち無相関となると同時性検出器が一意的な時間差を出すことが出来ないすなわち音像が定位しないこととなる。この時、空間的な広さを感じる。

作品の音刺激がヘッドフォン受聴のように直接耳に提示されることを想定し、3作品のチャンネル間相関係数を調べ、想起される空間印象から音刺激のシーンチェンジを検出する。

ここでは映像刺激と音刺激の対応を調べるのが目的であるため、音刺激のチャンネル間相関係数は映像刺激のフレーム間隔（1/30秒）を1ブロックとして導出する。

調べた結果を図10（動画1）、図11（動画2）、図12（動画3）、に示す。図は横軸時間を秒で示し、縦軸に相関係数を示す。図には2章で検出した映像のシーンチェンジのタイミングに緑の上向き矢印で示す。また相関係数が急激に変化している、すなわち音刺激におけるシーンチェンジのタイミングに赤の下向き矢印で示す。

## 4. 相関関数の導入

入賞できなかった動画1では、映像のシーンチェンジ（緑の上矢印）と音刺激のシーンチェンジ（赤の下矢印）で対応していないことがわかった。一方、入賞した動画3では対応していることが分かった。但し、入賞した動画2では音刺激のシーンチェンジが検出されていない。

相関係数は、チャンネル間の時間差が0での計算結果であるが、先の音像定位のメカニズムからチャンネル間に時間があることが考えられる。そこで、チャンネル間時間差を考慮するため、チャンネル間のラグタイムを考慮した相関関数を求める。

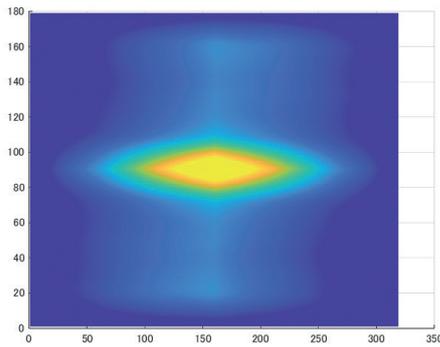
その結果を図13（動画1）、図14（動画2）、図15（動画3）、に示す。図の横軸は図10から図12と同じく時間を示し、縦軸にチャンネル間時間差、黄色が大きな値、青が小さな値を示す。

図11に示した相関係数で音響シーンチェンジが明確に判断できなかった動画2において、図14に示した相関関数ではラグタイム0の大きな振幅だけでなく、左右に定位すると思われるピークが見られた。

この結果より、入賞作品（動画2、動画3）は映像シーンチェンジと音刺激シーンチェンジのタイミングが対応する箇所が存在するが、入賞に至らなかった作品（動画1）では対応箇所が存在しないことがわかった。



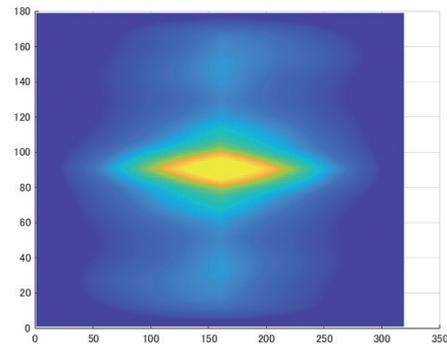
Before) 画像



Before) 2次元スペクトル



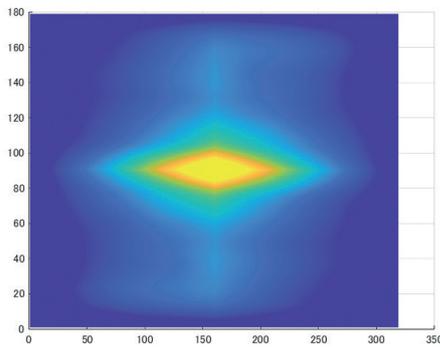
Before) 画像



Before) 2次元スペクトル



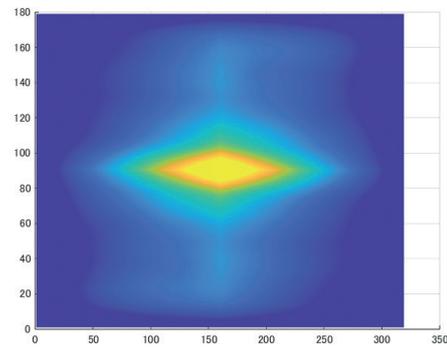
After) 画像



After) 2次元スペクトル



After) 画像



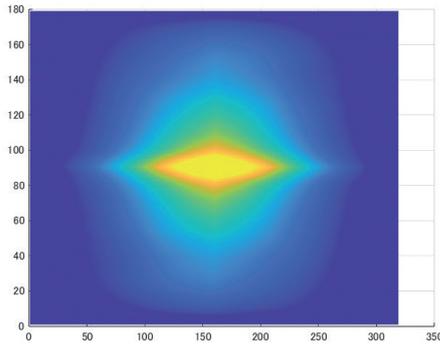
After) 2次元スペクトル

図1 動画1のシーンチェンジ(1)

図2 動画1のシーンチェンジ(2)



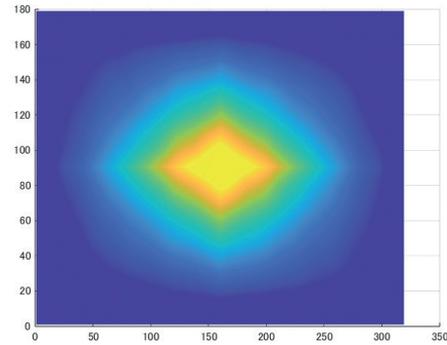
Before) 画像



Before) 2次元スペクトル



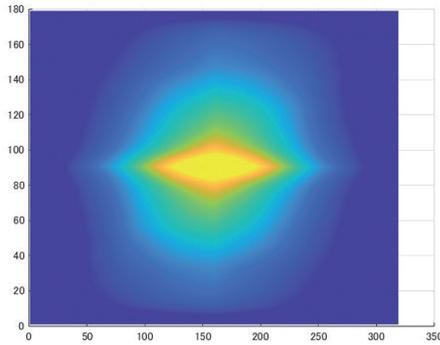
Before) 画像



Before) 2次元スペクトル



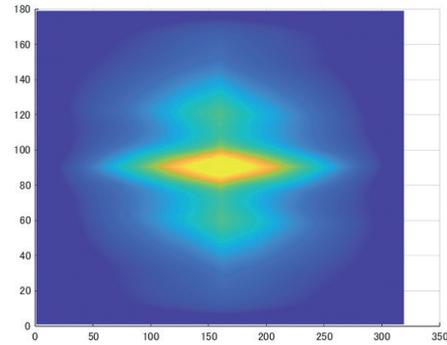
After) 画像



After) 2次元スペクトル



After) 画像



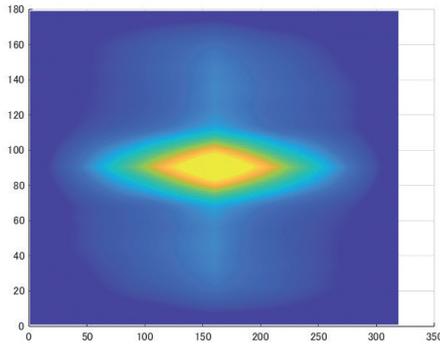
After) 2次元スペクトル

図3 動画1のシーンチェンジ(3)

図4 動画2のシーンチェンジ(1)



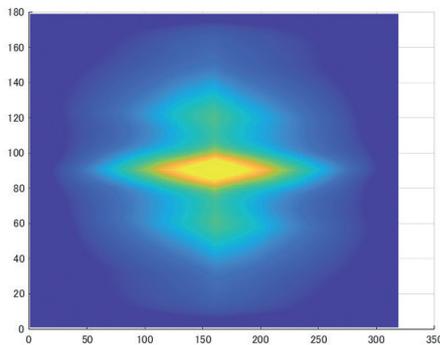
Before) 画像



Before) 2次元スペクトル



After) 画像

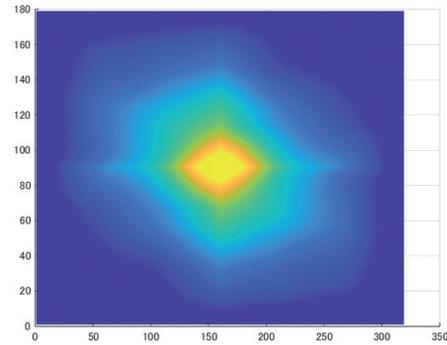


After) 2次元スペクトル

図5 動画2のシーンチェンジ(2)



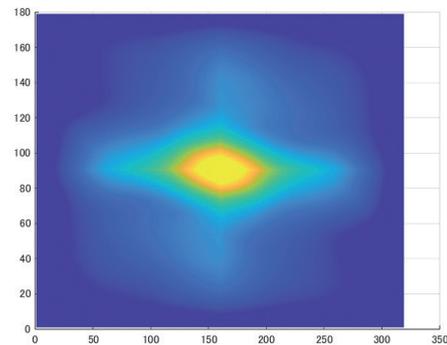
Before) 画像



Before) 2次元スペクトル



After) 画像

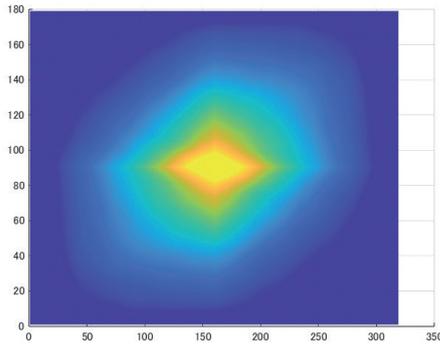


After) 2次元スペクトル

図6 動画3のシーンチェンジ(1)



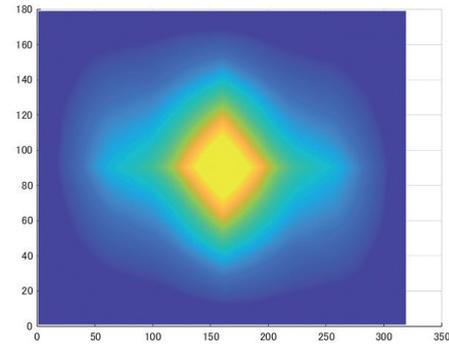
Before) 画像



Before) 2次元スペクトル



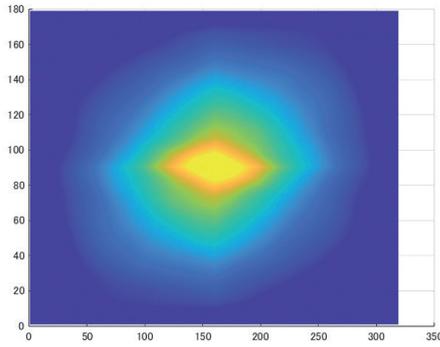
Before) 画像



Before) 2次元スペクトル



After) 画像

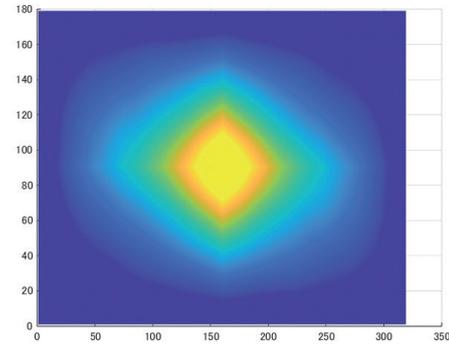


After) 2次元スペクトル

図7 動画3のシーンチェンジ(2)



After) 画像

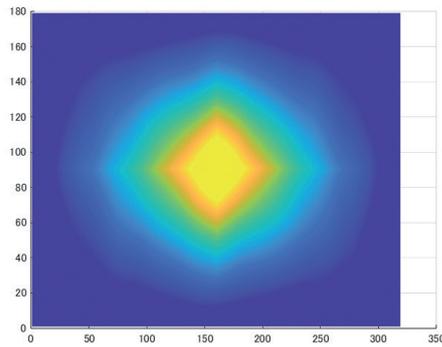


After) 2次元スペクトル

図8 動画3のシーンチェンジ(3)



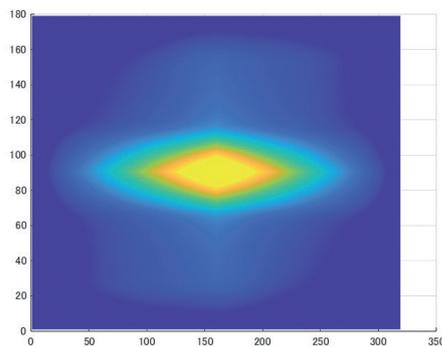
Before) 画像



Before) 2次元スペクトル



After) 画像



After) 2次元スペクトル

図9 動画3のシーンチェンジ (4)

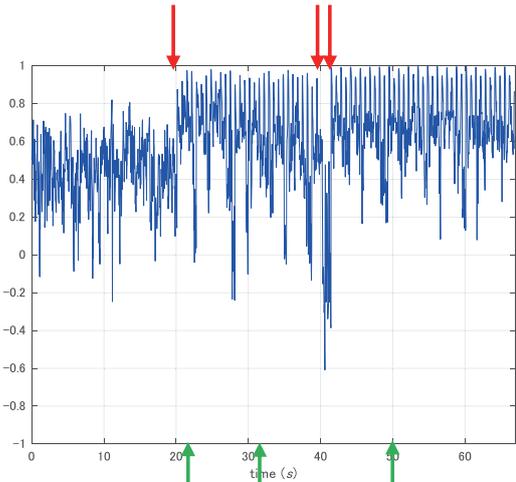


図10 音刺激のチャンネル間相関係数 (動画1)

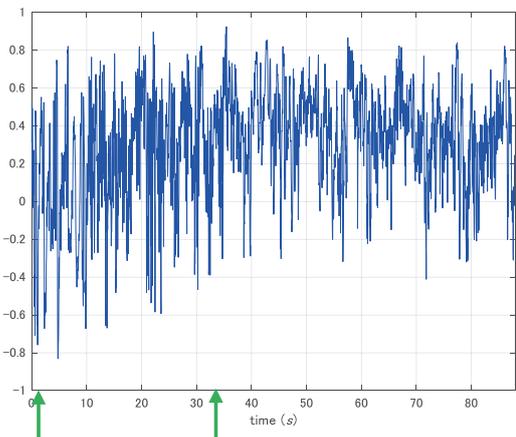


図11 音刺激のチャンネル間相関係数 (動画2)

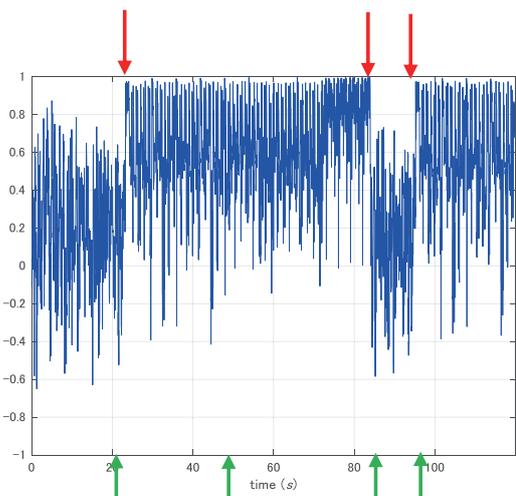


図12 音刺激のチャンネル間相関係数 (動画3)

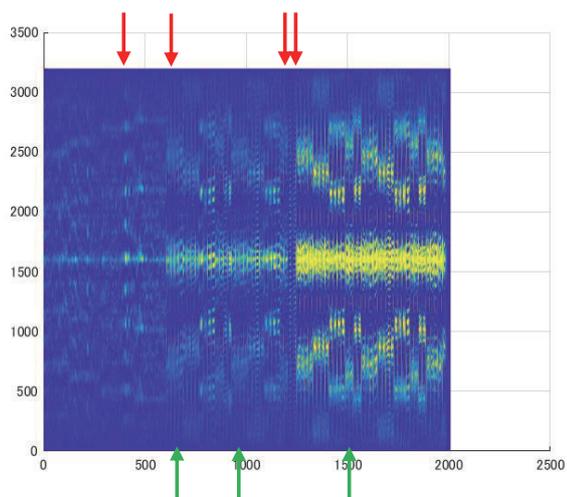


図13 音刺激のチャンネル間相関関数（動画1）

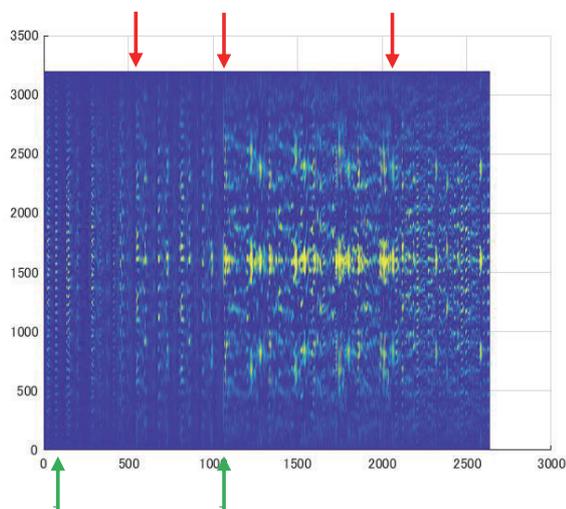


図14 音刺激のチャンネル間相関関数（動画2）

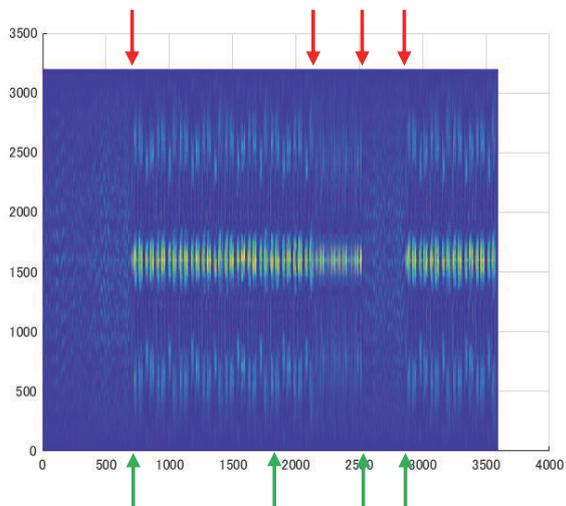


図15 音刺激のチャンネル間相関関数（動画3）

## 5. おわりに

インターネットの普及と利用端末の普及により、個人でも手軽に動画配信が可能となった。これに伴い誰もが動画コンテンツを作ることが出来るようになった。しかし、例えば映像と音の融和性等において違和感のある作品となることが少なくない。このため、動画コンテンツ制作をアシストするシステムが今後必要となると考えている。

このための第一歩として、ドローン撮影映像を用いた競技会に応募した作品において、映像刺激により想起される印象を画像2次元スペクトルから予測し、音刺激により想起される印象をチャンネル間相関から予測した。予測の正当性を得るために、競技会の審査方式が、1) 一般Web投票, 2) 審査員投票, であることから、落選した動画コンテンツと入選した動画コンテンツを用いた。

調べた結果、画像の2次元スペクトルから予測したシーンチェンジ時刻と、音刺激のチャンネル間相関関数を用いて予測したシーンチェンジが、落選した作品では対応しておらず、入選した作品では対応していることがわかった。今後も競技会に参加し、この仮説が正しいかについて検証するとともに、映像と音のシーンチェンジを調える方法について検討を進める。

## 参考文献

- (1) [https://ipsj.ixsq.nii.ac.jp/ej/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=36944&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=36944&item_no=1&page_id=13&block_id=8) (2023年6月14日アクセス)
- (2) [https://ipsj.ixsq.nii.ac.jp/ej/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=48931&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=48931&item_no=1&page_id=13&block_id=8) (2023年6月14日アクセス)
- (3) 高橋靖, “映画の定量的評価・解釈のための標準セマンティックスコア法: ヒューマン・コンテンツ・インタフェースデザインに向けて (5)”, デザイン学研究, 47巻, 4号, pp. 1-10, 2000, [https://www.jstage.jst.go.jp/article/jssdj/47/4/47\\_KJ00001647560/\\_article/-char/ja/](https://www.jstage.jst.go.jp/article/jssdj/47/4/47_KJ00001647560/_article/-char/ja/)
- (4) 小峯一晃, “生体情報を用いた映像評価技術”, NHK 技研 R&D, 2016年9月号, <https://www.nhk.or.jp/strl/publica/rd/159/3>.

- html (2023年6月14日アクセス)
- (5) デジタルコンテンツ協会, “動画映像の視覚評価に関する調査研究”,  
[https://www.dcaj.or.jp/project/report/pdf/2007/dc08\\_04.pdf](https://www.dcaj.or.jp/project/report/pdf/2007/dc08_04.pdf) (2023年6月14日アクセス)
- (6) 会津昌夫, 堀田裕弘, “映像, 画像メディアの評価－像メディア評価学の勧め－”, 電子情報通信学会誌, Vol. 96, No. 4, pp. 1-pp6, 2013,  
<https://www.ieice.org/jpn/books/kaishikiji/2013/201304.pdf>
- (7) <https://waic.jp/docs/jis2016/compliance-guidelines/202104/> (2023年6月14日アクセス)
- (8) Drone Movie World Championship,  
<https://dronemoviecs.com/>  
(2023年6月13日アクセス)
- (9) 柳川博文, “音場の拡がり感 (〈小特集〉音場の空間的印象)”, 日本音響学会誌, 48巻, 9号, pp. 1-pp. 5, 1992, [https://www.jstage.jst.go.jp/article/jasj/48/9/48\\_KJ00001456543/\\_article/-char/ja/](https://www.jstage.jst.go.jp/article/jasj/48/9/48_KJ00001456543/_article/-char/ja/)
- (10) Jeffress モデル, [https://bsd.neuroinf.jp/w/index.php?title=Jeffress%E3%83%A2%E3%83%87%E3%83%AB&mobileaction=toggle\\_view\\_desktop](https://bsd.neuroinf.jp/w/index.php?title=Jeffress%E3%83%A2%E3%83%87%E3%83%AB&mobileaction=toggle_view_desktop) (2023年6月13日アクセス)

---

(2023年6月14日受理)

