

Multiple Vertical Panning を用いた立体音響システムにおいて 奥行表現手法が映像と音の一致度に及ぼす影響

木村 敏幸[†]

[†] 東北学院大学情報学部データサイエンス学科 〒984-8588 仙台市若林区清水小路 3-1

E-mail: †t-kimura@m.ieice.org

あらまし Multiple Vertical Panning を用いた立体音響システムにおいて音量変化による奥行表現手法をこれまで検討してきたが、画面の上側に映像を再生した場合については検討していなかった。本研究では、映像位置を上側にした場合の条件も加えたうえで、奥行表現手法が臨場感に及ぼす影響を検討する。その際、臨場感を評価する尺度として映像と音の一致度を適用する。視聴覚実験を実施したところ、Multiple Vertical Panning を用いた立体音響システムは映像が中心にある場合にステレオホニックよりも臨場感が向上し、比例的に音量を変化させればより臨場感が向上するということが分かった。

キーワード 超臨場感コミュニケーション, 立体音響システム, 垂直パニング, 奥行表現, 視聴覚実験

Effect of Depth Expression on Coincidence between Sound and Vision in 3D Audio System Using Multiple Vertical Panning

Toshiyuki KIMURA[†]

[†] Department of Data Science, Faculty of Informatics, Tohoku Gakuin University

3-1, Shimizu-koji, Wakabayashi-ku, Sendai, Miyagi, 984-8588 Japan

E-mail: †t-kimura@m.ieice.org

Abstract Although the depth expression method via sound level variation have been studied in the 3D audio system using multiple vertical panning, the depth expression method is not studied in the case where the visual object is placed on the upper side of the visual display. In this study, the effect of the depth expression method on the sense of presence is evaluated with the addition of the condition when the position of the visual object is on the upper side. In the experiment, the degree of coincidence between sound and vision is applied as the evaluation scale of the sense of presence. From the result of audio-visual experiments, it have shown that a 3D sound system using Multiple Vertical Panning improves the sense of presence compared to stereophonic when the visual object is placed in the center, and that the sense of presence is further improved if the sound level is varied proportionally.

Key words ultra-realistic communication, 3D audio system, vertical panning, depth expression, audio-visual experiment

1. はじめに

超臨場感コミュニケーション技術はこれまでに 4K 解像度のデジタルシネマや 8K 解像度の画面テレビに実用化されている。これらに適合した音響として、大画面ディスプレイの前に何も置かずに、複数の人が何も装着せずにどこでも聞いても映像の位置から音が出ているように聞こえる立体音響システムの開発が期待されている。

大画面ディスプレイに適合する立体音響システムとして、私は今までとは異なる観点に基づいた新たな立体音響システム (Multiple Vertical Panning, 以降「MVP 方式」と呼ぶ) をこれま

で提案し [1], 実用化の可能性も示してきた [2]. さらに、近年研究されている裸眼立体映像に適合した奥行きを表現するために、音量変化による奥行表現手法についても検討してきた [3], [4].

しかしながら、これまでの奥行表現手法に関する研究では映像の位置が画面の上側にある場合について検討していなかった。そこで、本報告では、映像の位置が画面の上側にある場合も追加して再度視聴覚実験を実施する。さらに、これまでの研究では「臨場感」を直接評価させていたが、本報告では個人差が小さくなるのが期待される「映像と音の一致度」を評価指標として用いる。

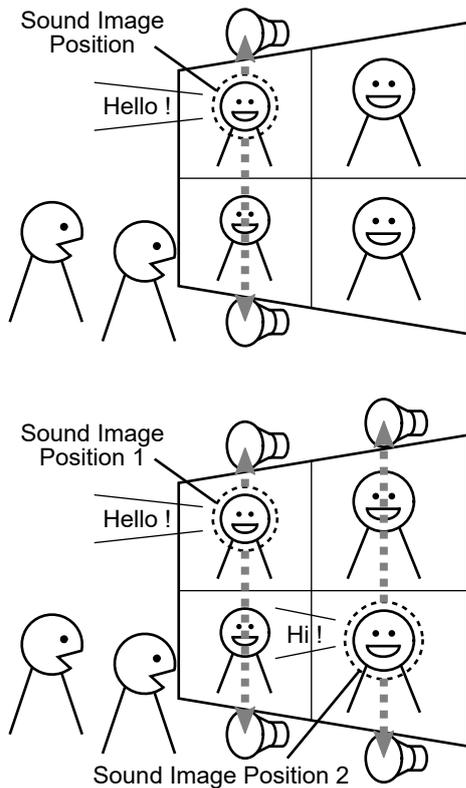


図1 Multiple Vertical Panning (MVP) 方式の基本構成 [4]

Fig. 1 Basic configuration of a multiple vertical panning (MVP) method [4].

2. MVP方式の原理 [1], [2]

Fig. 1 に MVP 方式の基本構成を示す。まず、Fig. 1 の上側に示すように、映像の位置の上下に 2 個のスピーカを配置する。そして、音源に音量差をつけて 2 個のスピーカから音を再生（すなわち「垂直パニング」）すると、視聴者は 2 個のスピーカの間で音が鳴っているように感じるようになる。その際、適切な音量差を設定することによって、視聴者は映像の位置で音が鳴っているように感じるようになる。音が鳴っているのはディスプレイの上下に配置した 2 個のスピーカだけなので、複数の視聴者はどこにいても常に映像の位置で音が鳴っているように感じることができる。

さらに、Fig. 1 の下側に示すように、ディスプレイの上下に複数のスピーカ対を配置することによって、音像を表現できる位置がディスプレイの上下方向のみならず左右方向にも拡大される。これにより、複数の視聴者はどこにいても常にディスプレイ上の映像の位置で音が鳴っているように感じることができる。

3. 制作システム

Fig. 2 に制作システムの実行風景を示す。55 インチの映像ディスプレイ（TOSHIBA：55J20X）の上下に計 18 個のスピーカで構成された MVP 方式用のスピーカアレイ、左右にはステレオホニック用のスピーカを配置した。スピーカには市販のスピーカユニット（ダイトーボイス：AR-10N）を市販の密閉型エンクロージャ（ダイトーボイス：EX-10 BK）に取り付けたもの



図2 制作システムの実行風景

Fig. 2 Image of developed audio-visual display system.

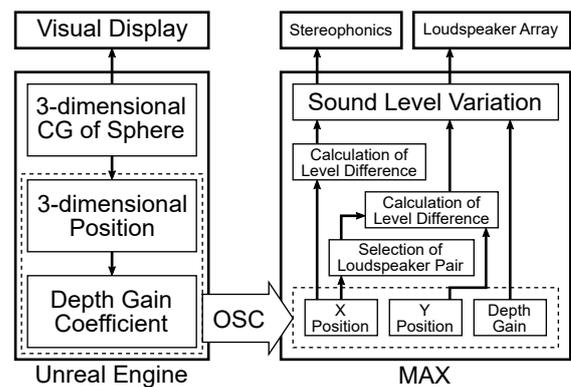


図3 制作システムの構成図

Fig. 3 Configuration of developed audio-visual display system.

を使用した。

Fig. 3 に制作システムの構成図を示す。ソフトウェアにはゲームエンジンの Unreal Engine [5] (ver. 4.16) とビジュアルプログラミング言語の Max [6] (ver. 7.3.5) を使用した。Unreal Engine では、白色球体の 3 次元 CG が奥行き方向に移動する映像を映像ディスプレイ上に表示し、球体の水平及び垂直位置、奥行き距離に基づくゲイン係数および再生する音の方式を OSC (Open Sound Control) 信号で送信した。一方、Max では、受信した OSC 信号から再生する音の方式を決定し、受信した水平及び垂直位置を基にスピーカから音を再生し、受信したゲイン係数に従って音量を変化させた。

Max における MVP 方式による音の再生は以下の手順により実施した。まずは受信した水平位置を基に、下記の式によってディスプレイ上下に配置した 2 個の再生用スピーカを選択する。

$$P'_H = \Delta d \text{round} \left(\frac{P_H + 0.54}{\Delta d} \right) - 0.54 \quad (1)$$

但し, $P_H(=-0.605\sim 0.605)$ は受信した水平位置, $P'_H(=-0.54, -0.405, -0.27, -0.135, 0, 0.135, 0.27, 0.405, 0.54)$ は選択した 2 個の再生用スピーカの水平位置を表し, 0 がディスプレイの横中心位置に相当する. また, $\Delta d(=0.135)$ は左右のスピーカ間隔を表す.

次に, 下記の式によって音源信号 $s(n)$ に上下の音量差をつける.

$$s_U(n) = a_U s(n) \quad (2)$$

$$s_D(n) = a_D s(n) \quad (3)$$

但し, $s_U(n)$, $s_D(n)$ は上側及び下側のスピーカへの音信号, a_U , a_D はそれぞれのゲイン係数を表し, 音量差 $\Delta A[\text{dB}]$ から下記のように定義される.

$$a_U = \frac{10^{\frac{\Delta A}{20}}}{\sqrt{10^{\frac{\Delta A}{10}} + 1}} \quad (4)$$

$$a_D = \frac{1}{\sqrt{10^{\frac{\Delta A}{10}} + 1}} \quad (5)$$

本システムでは音量差 ΔA は先行研究 [1] に従い, 下記のように設定した.

$$\Delta A = \frac{P'_V + 0.1437}{0.1065} \quad (6)$$

但し, $P'_V \left(= \frac{2.5}{0.86} P_V + 0.0654 \right)$ は先行研究 [1] とは上下のスピーカ間隔が異なることにより設けた補正した垂直位置であり, P_V は受信した垂直位置を表す. P'_V の導出手順は付録にて示す.

最後に, 下記の式によって音量を変化させ, 選択した 2 個の再生用スピーカから再生する.

$$x_U(n) = a_Z s_U(n) \quad (7)$$

$$x_D(n) = a_Z s_D(n) \quad (8)$$

但し, $x_U(n)$, $x_D(n)$ は上側及び下側のスピーカから再生する音信号, a_Z は受信した奥行距離に基づくゲイン係数を表す. 今回は下記の 2 種類を設定した.

$$\text{「比例増大」 } a_Z = 1 - \frac{P_Z}{10} \quad (9)$$

$$\text{「指数増大」 } a_Z = \frac{1.05}{1.05 + P_Z} \quad (10)$$

但し, $P_Z(=0\sim 10)$ は球体の奥行距離を表す. 「比例増大」では先行研究 [3], [4] のように距離によって音量を比例的に変化させているのに対し, 「指数増大」では点音源の物理的な音量変化を模擬するため, 音量を逆数的に変化させている.

一方, 従来のステレオホニック方式の場合は, 下記の式によって音源信号 $s(n)$ に左右の音量差をつけた音信号をディスプレイの左右に配置した 2 個のスピーカから再生した.

$$x_L(n) = a_Z a_L s(n) \quad (11)$$

$$x_R(n) = a_Z a_R s(n) \quad (12)$$

但し, $x_L(n)$, $x_R(n)$ は左側及び右側のスピーカから再生する音

表 1 視聴覚実験における実験条件

Table 1 Experimental conditions in the audio-visual experiment.

| Index | Sphere | Sound level variation | Sound playing |
|---------|--------|-----------------------|---------------|
| (i) | Center | None | Streophonic |
| (ii) | | | MVP |
| (iii) | | Propotion | Streophonic |
| (iv) | | | MVP |
| (v) | | Exponent | Streophonic |
| (vi) | | | MVP |
| (vii) | Left | None | Streophonic |
| (viii) | | | MVP |
| (ix) | | Propotion | Streophonic |
| (x) | | | MVP |
| (xi) | | Exponent | Streophonic |
| (xii) | | | MVP |
| (xiii) | Upper | None | Streophonic |
| (xiv) | | | MVP |
| (xv) | | Propotion | Streophonic |
| (xvi) | | | MVP |
| (xvii) | | Exponent | Streophonic |
| (xviii) | | | MVP |

信号を表す. a_L , a_R は左右の 2 個のスピーカのゲイン係数を表し, 本実験ではタンゼント則 [7] に従い下記の式のように設定した.

$$a_L = \frac{1 - \frac{\tan \theta}{\tan \theta_0}}{\sqrt{2 \left\{ 1 + \left(\frac{\tan \theta}{\tan \theta_0} \right)^2 \right\}}} \quad (13)$$

$$a_R = \frac{1 + \frac{\tan \theta}{\tan \theta_0}}{\sqrt{2 \left\{ 1 + \left(\frac{\tan \theta}{\tan \theta_0} \right)^2 \right\}}} \quad (14)$$

但し, θ 及び θ_0 はディスプレイ正面の視聴位置における音像及び左右の 2 個のスピーカの開き角を表す. 本システムでは, 下記のように設定した.

$$\frac{\tan \theta}{\tan \theta_0} = \frac{P_H}{0.69} \quad (15)$$

但し, P_H は受信した水平位置を表す.

4. 視聴覚実験

4.1 実験環境・条件

視聴覚実験では Table 1 に示すような計 18 種類の実験条件を設定した. 音には白色雑音を用いた. 音量の変化がない条件では, a_Z は常に 1 とした.

実験は暗騒音レベルが 40.2 dBA である研究室の実験室の一角にて行った. Fig. 4 に示すように, 映像ディスプレイから 1.05 m 離れた地点に 2 カ所の視聴位置 (中央, 側方) を設定した. 視聴位置の高さは視聴者の耳の位置において 1.44 m とした.

実験で使用した白色球体の映像は, Fig. 5 に示すように, 画

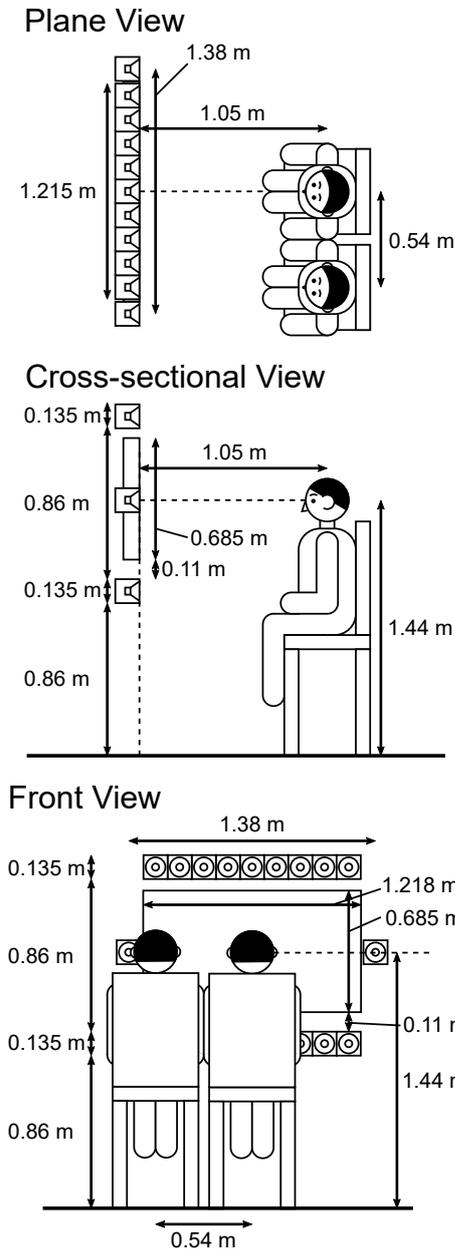


図4 視聴覚実験における視聴者、映像ディスプレイ、スピーカの配置
Fig. 4 Position of viewers, visual display and the loudspeakers in the audio-visual experiment.

面上の三カ所の位置（中央、左、上）の奥行 10 m から 5 秒間で手前方向に一定速度で移動させた。

4.2 実験計画・手順

視聴者は 10 名である。実験の流れ図を Fig. 6 に示す。セッションでは、3 試行の練習試行の後に、Table 1 で示した計 18 種類の実験条件を 2 回繰り返した計 36 回の本試行を行った。試行の提示順序は視聴者ごとにランダム化した。その際、セッションの間には休憩時間を設けた。

視聴者には提示した刺激の映像と音の一致度を評定させた。その際、評定は Table 2 に示すような 5 段階とした。また、評定の際には視聴者は頭部や上半身を自由に動かすことができた。

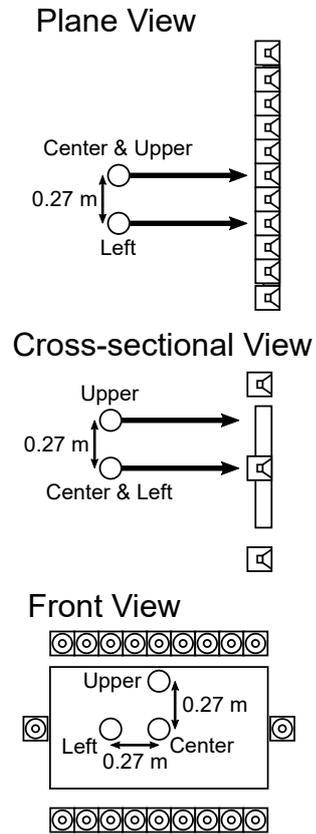


図5 視聴覚実験における球体の位置
Fig. 5 Position of the sphere object in the audio-visual experiment.

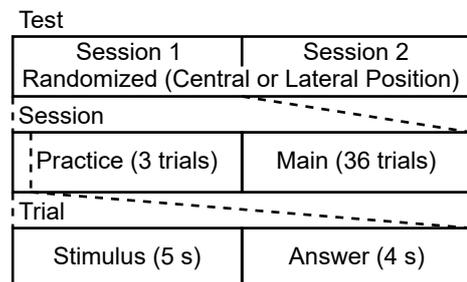


図6 視聴覚実験の流れ図
Fig. 6 Flowchart of the audio-visual experiment.

表2 視聴覚実験における評価尺度

Table 2 Evaluation scale in the audio-visual experiment.

| Grade | Judgment |
|-------|-----------|
| 5 | Excellent |
| 4 | Good |
| 3 | Fair |
| 2 | Poor |
| 1 | Bad |

4.3 実験結果及び考察

実験条件ごとの平均評定値を Fig. 7 に示す。エラーバーは 95% 信頼区間を表す。

球体位置 (3 水準)、視聴位置 (2 水準)、音量変化手法 (3 水準) 及び音の再生方式 (2 水準) を要因とする四要因分散分析

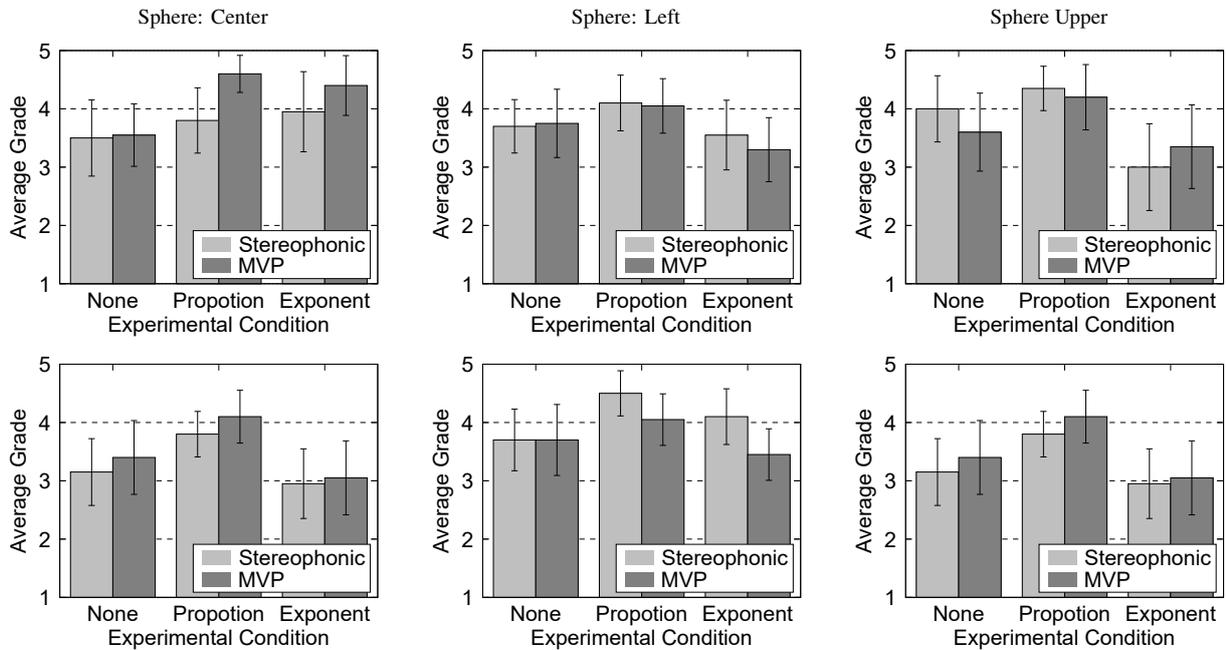


図7 視聴覚実験の結果（上：中央視聴位置，下：側方視聴位置）

Fig. 7 Results of the audio-visual experiment. (Upper: Central, Lower: Lateral viewing position)

表3 視聴覚実験における分散分析表

Table 3 Analysis of variance in the audio-visual experiment.

| source | SS | df | MS | F |
|--------------------------|---------|-----|-------|-----------------------|
| A: Sphere position | 7.54 | 2 | 3.77 | 2.72 ⁺ |
| B: Viewing position | 10.51 | 1 | 10.51 | 7.58 ^{**} |
| C: Sound level variation | 66.05 | 2 | 33.03 | 23.82 ^{****} |
| D: Sound playing | 0.61 | 1 | 0.61 | 0.44 |
| AB | 17.03 | 2 | 8.52 | 6.14 ^{***} |
| AC | 10.46 | 4 | 2.62 | 1.89 |
| AD | 9.10 | 2 | 4.55 | 3.28 [*] |
| BC | 1.36 | 2 | 0.68 | 0.49 |
| BD | 0.23 | 1 | 0.23 | 0.17 |
| CD | 0.41 | 2 | 0.20 | 0.15 |
| ABC | 12.71 | 4 | 3.18 | 2.30 ⁺ |
| ABD | 2.88 | 2 | 1.44 | 1.04 |
| ACD | 4.44 | 4 | 1.11 | 0.80 |
| BCD | 2.84 | 2 | 1.42 | 1.02 |
| ABCD | 1.16 | 4 | 0.29 | 0.21 |
| error[WC] | 948.25 | 684 | 1.39 | |
| Total | 1095.60 | 719 | | |

（全て被験者間要因）を ANOVA4 on the web [8] にて実施した。実施した結果の分散分析表を Table 3 に示す。

まず、音の再生方式による影響を検討するため、Table 3 において要因 D が関係する結果に着目したところ、球体位置との 1 次交互作用 (Table 3 中の「AD」) に有意差が見られる一方、主効果 (Table 3 中の「D」)、その他の 1 次交互作用 (Table 3 中の「BD」「CD」)、2 次交互作用 (Table 3 中の「ABD」「ACD」「BCD」) 及び 3 次交互作用 (Table 3 中の「ABCD」) には有意差が見られなかった。球体位置との 1 次交互作用に関する単純主効果を実施したところ、球体位置が中心の場合に、音の再生方式による

主効果に有意差が見られた ($F(1, 684) = 4.57$. $p < 0.05$) 一方で、球体位置が左や上の場合には有意差が見られなかった (左: $F(1, 684) = 2.19$. $p > 0.05$, 上: $F(1, 684) = 0.24$. $p > 0.05$)。従って、球体が中心にある場合、どの視聴位置においても MVP 方式はステレオホニックよりも映像と音の一致度が高くなるということが言える。

次に、音量変化手法による影響を検討するため、Table 3 において要因 C が関係する結果に着目したところ、主効果 (Table 3 中の「C」) に有意差が見られる一方、1 次交互作用 (Table 3 中の「AC」「BC」「CD」)、2 次交互作用 (Table 3 中の「ABC」「ACD」「BCD」) 及び 3 次交互作用 (Table 3 中の「ABCD」) には有意差が見られなかった。さらに主効果に関して Ryan の手法による多重比較を実施したところ、変化なしと比例増大及び比例増大と指数増大の組み合わせに対して 5% 水準で有意差がみられ、変化なしと指数増大の組み合わせに関しては有意差が見られなかった。従って、奥行表現のための音量変化手法としては、比例的に音量を変化させれば常に映像と音の一致度がより向上するということ言える。

5. まとめ

本報告では、これまでに提案してきた MVP 方式による奥行表現の可能性を検討するために、映像位置を上側にした場合の条件も加えたうえで、映像と音の一致度に基づいた臨場感の評価を実施した。その結果、映像が中心にある場合、MVP 方式はステレオホニックよりも臨場感が向上することが分かった。また、比例的に音量を変化させればさらに臨場感が向上することも分かった。

一方で、物理的な現象に即していないにも関わらず、音量変化がない条件においても高い一致度を示していた。そこで、今

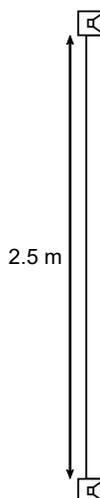


図8 先行研究[1]における映像ディスプレイ及びスピーカの配置
Fig. 8 Position of visual display and the loudspeakers in the previous study [1].

後の課題としては、視聴者に物理的な現象を理解させようとして一致度の評価をさせるために、球体をより遠くから移動させた場合において視聴覚実験を実施する必要があると考えられる。

謝辞

研究の遂行にあたり、システム制作および評価実験を担当して下さった細谷蓮氏に感謝の意を申し上げます。本報告における視聴覚実験は東北学院大学人間対象研究審査委員会の承認を得たうえで実施した。

文献

- [1] T. Kimura and H. Ando, "3D audio system using multiple vertical panning for large-screen multiview 3D video display," ITE Transactions on Media Technology and Applications, vol.2, no.1, pp.33-45, Jan. 2014.
- [2] T. Kimura and H. Ando, "Effect of discretization of sound source position on sense of presence in 3D audio systems using multiple vertical panning," Transactions of Virtual Reality Society of Japan, vol.20, no.3, pp.179-188, Sept. 2015.
- [3] 木村敏幸, "Multiple vertical panning を用いた立体音響システムにおける奥行き表現の臨場感への影響," 日本音響学会秋季研究発表会講演論文集, no.2-1-16, pp.1265-1268, Sept. 2019.
- [4] 木村敏幸, "Multiple vertical panning を用いた立体音響システムにおける奥行き表現の比較検討," 日本音響学会春季研究発表会講演論文集, no.2-11-8, pp.1309-1312, March 2021.
- [5] Unreal Engine. <http://www.unrealengine.com/>.
- [6] Max. <http://cycling74.com/products/max/>.
- [7] B. Bernfeld, "Attempts for better understanding of the directional stereophonic listening mechanism," Proc. 44th Audio Eng. Soc. Convention, no.C-4, pp.1-24, Feb. 1973.
- [8] ANOVA4 on the Web. <https://www.hju.ac.jp/kiriki/anova4/>.

付 録

先行研究[1]における映像ディスプレイ及びスピーカの配置を Fig. 8 に示す。映像ディスプレイの大きさが 200 インチであったため、映像ディスプレイの高さは 2.5 m であり、スクリーンの上部と下部にスピーカが直に配置されている。その際に設定した垂直パニング曲線に基づく垂直位置 P'_V は以下の式で示される。

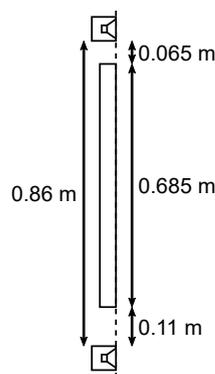


図9 本研究における映像ディスプレイ及びスピーカの配置
Fig. 9 Position of visual display and the loudspeakers in this study.

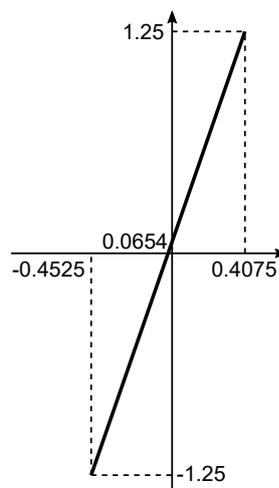


図10 本研究における垂直位置の補正直線
Fig. 10 Adjustment line of the vertical position in this study.

$$P'_V = \begin{cases} -1.32 & (\Delta A < -11.05) \\ 0.1065\Delta A - 0.1437 & (-11.05 \leq \Delta A \leq 13.74) \\ 1.32 & (\Delta A > 13.74) \end{cases} \quad (16)$$

但し、 $\Delta A[\text{dB}]$ は音量差を表す。

一方、本研究における映像ディスプレイ及びスピーカの配置を Fig. 9 に示す。映像ディスプレイの大きさが 55 インチになったため、映像ディスプレイの高さが 0.685 m に変更されたのに加え、スクリーンの上部と下部に配置されているスピーカと映像ディスプレイとの間に間隔が生じている。

従って、本研究における垂直位置 P_V を先行研究における垂直位置 P'_V に補正する直線は Fig. 10 に示すようになり、下記の式で示される。

$$P'_V = \frac{2.5}{0.86} P_V + 0.0654 \quad (17)$$