

Multiple Vertical Panning を用いた立体音響システムにおける 音源の垂直位置の離散化の影響*

○木村敏幸, △安藤広志 (NICT)

1 はじめに

独立行政法人情報通信研究機構 (NICT) では超臨場感コミュニケーション技術に関する研究を進めている [1]. 立体映像技術や立体音響技術によって三次元の映像や音響をよりリアルに表現することができるようになれば, 今までの映像・音響メディアでは実現できなかったより臨場感のあるコミュニケーション (立体テレビや立体遠隔通信会議など) が可能になると期待される.

NICT では次世代の立体映像技術として裸眼立体映像技術に着目し, これまでにプロジェクタアレイを用いた方式を提案し, 200 インチの画面を持った大画面裸眼立体映像提示システムを開発してきている [2]. 我々はさらに大画面裸眼立体映像提示システムに適合する立体音響システムとして, 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) を提案し, MVP 方式が十分な性能を有することを実証してきた [3]. また, スピーカの数減らしていった場合にどのくらい臨場感に影響があるのかを視聴覚実験によって検討し, スピーカを 10 個 (スクリーン上部及び下部にそれぞれ 5 個) にまで減らしても視聴者は臨場感の違いを識別することができないということも示してきた [4].

MVP 方式は立体像の垂直位置に従って垂直パニング用のスピーカ対の相対的な音量を調節する方式である. そのため, MVP 方式を放送や通信に応用する場合には, 立体像の位置情報を伝送する必要がある. しかし, 音源が移動している場合には立体像の位置情報をその都度伝送する必要があるため, それに伴い伝送量が増大する恐れがある. このことが MVP 方式の実用化の妨げになる可能性がある.

もし, 音源の垂直位置を離散化し, 段階数が少ない場合でも視聴者が臨場感の違いを識別できないのであれば, 立体像の位置情報を伝送するのに必要なビットレートを削減することができるので, 実用的なシステムを構築できるようになると期待される. そこで, 本発表では MVP 方式において実用的なシステムを構築するために, 音源の垂直位置を離散化させた場合にどのくらい臨場感に影響があるのかを視聴覚実

験によって評価する.

2 視聴覚実験

2.1 実験環境・条件

実験は 200 インチ背面投影型映像スクリーンが設置されている会議室にて行った. スクリーンの後方には 2 台のプロジェクタが設置されており, それぞれ左目用と右目用の映像が投影される. プロジェクタとスクリーンの間には偏光フィルタが設置されているので, 視聴者は偏光方式の立体メガネをかけることによって立体映像を観ることが出来る. 部屋の残響時間は 402 ms, 部屋の暗騒音レベルは 38 dBA だった.

Fig. 1 に示すようにスクリーンの上下に計 42 個のスピーカを配置した. スクリーンは会議室の壁面に設置されており, スピーカをスクリーンの真上と真下に配置することができないため, スピーカはスクリーンより 0.275 m 前方に配置した. スピーカは市販のスピーカユニット (Fostex: FE103En) を自作の密閉型エンクロージャ (11 cm (幅) × 25 cm (奥行) × 11 cm (高さ)) に取り付けたものを用いた. 大画面裸眼立体映像提示システムの視聴範囲 (最適視聴距離は 5.5 m) を考慮して, スクリーンから 3.5, 5.5, 7.5 m 離れた地点に 3 か所の視聴位置 (以降「前方位置」「中心位置」「後方位置」と呼ぶ) を設定した. 大画面裸眼立体映像提示システムの最適視聴幅はスクリーンから 5.5 m 離れた場合, スクリーン正面を中心として左右に 2 m である. そこで, スクリーン正面から 5.5 m 離れ, さらに 2 m 左側に移動した視聴位置 (以降「側方位置」と呼ぶ) をもう 1 か所設定した. 計 4 か所の視聴位置の高さは視聴者の耳の位置において 1.5 m とした. 音圧レベルは中心位置において約 70 dBA に設定した.

実験に用いた立体映像を Fig. 2 に示す. この立体映像 (5 秒間) は UFO (Fig. 2 内の黄色の枠線) が音を出しながら画面内を動き回り, UFO が星やボール (Fig. 2 内の赤色の枠線) に当たると星やボールの場所から音が再生されるという内容である. 立体映像の最適視聴距離は 5.5 m, 視差は 0.0625 m とした. また, 大画面裸眼立体映像提示システムは視聴位置が左右方向に移動するとそれに合わせて立体像の見え

*Effect of the discretization of the vertical position of sound sources in 3D audio system using multiple vertical panning. by KIMURA, Toshiyuki, ANDO, Hiroshi (National Institute of Information and Communications Technology)

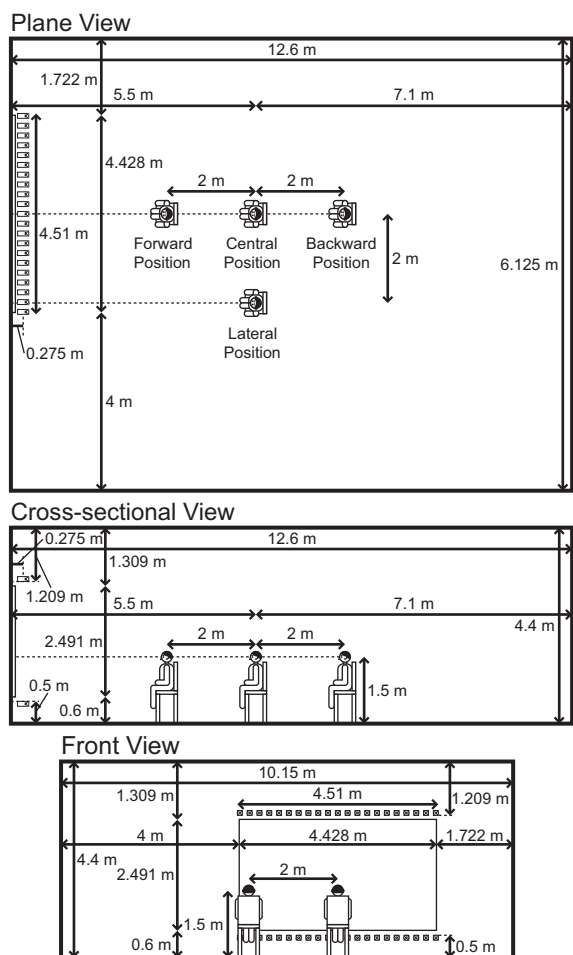


Fig. 1 Position of viewers, screen and the loud-speaker array in the audio-visual experiment.

方が変わるので、本実験においても視聴位置が左右方向に移動するたびにその位置に合わせた立体映像を提示した。

本実験で設定した音の再生条件を Fig. 3 に示す。条件 (a)-(f) における灰色のスピーカは音が再生されないスピーカ、スクリーン上の黒色の太線は垂直方向に離散化されていない音源位置、黒色の丸は垂直方向に離散化された音源位置を表す。条件 (f) は前回の実験 [4] においてスピーカの数を超えても臨場感の違いが識別できないと判断された条件である。条件 (a) は前回の実験 [4] において臨場感に違いがあると判断された条件であり、今回の実験における視聴者の臨場感の識別能力を検査するために設定した。

時刻 $T = \frac{m-1}{F_v}$ において位置 (P_H, P_V) の立体像から音が再生される場合の各条件における再生音の合成は下記の手順により実施した。但し、 $F_v (=30 \text{ fps})$, $m (=1, \dots)$ は映像のフレームレート及びフレームインデックスを表す。また、 $P_H (= -2.214 \sim 2.214)$ は立体像の横位置を表し、0 がスクリーンの横中心位置に相当する。一方、 $P_V (= -1.2455 \sim 1.2455)$ は立体像の縦

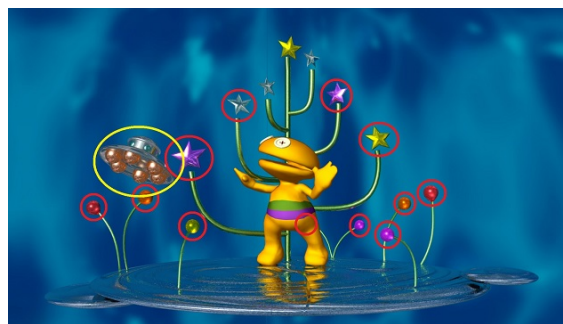


Fig. 2 3D video used in the audio-visual experiment.

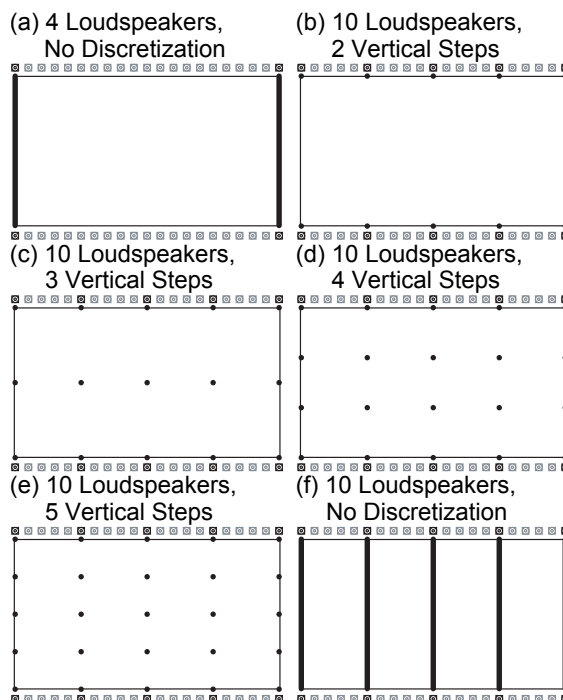


Fig. 3 Sound conditions used in the audio-visual experiment.

位置を表し、 -0.3455 が視聴者の耳の高さに相当する。

まず、提示する立体像の横位置 P_H を元に下記の式によってスクリーン上下に配置した 2 個の再生用スピーカを選択する。

$$P'_H = \Delta d_H \text{ round} \left(\frac{P_H + 2.2}{\Delta d_H} \right) - 2.2 \quad (1)$$

但し、 $P'_H (= -2.2, \dots, 2.2)$ は選択した 2 個の再生用スピーカの横位置を表す。また、 Δd_H は左右のスピーカ間隔を表し、Fig. 3 の条件 (a) では 4.4 m, その他の条件では 1.1 m となる。

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た音信号を選択した 2 個の再生用スピーカ

カから再生する。

$$x_U(n) = a_U w(n) s(n) \quad (2)$$

$$x_D(n) = a_D w(n) s(n) \quad (3)$$

$$(n = \frac{F_s}{F_v}(m-1), \dots, \frac{F_s}{F_v}m + LF_s)$$

但し, $F_s(=48 \text{ kHz})$, $n(=0, \dots)$ は音の標準化周波数及び標本時間, $x_U(n)$, $x_D(n)$ は上側及び下側のスピーカから再生する音信号を表す。また, $w(n)$, $L(=1 \text{ ms})$ は下記のように定義される音の窓関数及びそのクロスフェード時間を表す。

$$w(n) = \begin{cases} \frac{1}{LF_s} \{n - \frac{F_s}{F_v}(m-1)\} & (n = \frac{F_s}{F_v}(m-1), \dots, \frac{F_s}{F_v}(m-1) + LF_s) \\ 1 & (n = \frac{F_s}{F_v}(m-1) + LF_s, \dots, \frac{F_s}{F_v}m) \\ -\frac{1}{LF_s} (n - \frac{F_s}{F_v}m) + 1 & (n = \frac{F_s}{F_v}m, \dots, \frac{F_s}{F_v}m + LF_s) \end{cases} \quad (4)$$

一方, 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。

$$a_U = \frac{10^{\frac{\Delta A}{20}}}{\sqrt{10^{\frac{\Delta A}{10}} + 1}} \quad (5)$$

$$a_D = \frac{1}{\sqrt{10^{\frac{\Delta A}{10}} + 1}} \quad (6)$$

本実験では音量差 ΔA は先行研究 [3] に従い, 下記のように設定した。

$$\Delta A = \frac{\alpha P'_V + 0.1437}{0.1065} \quad (7)$$

但し, $\alpha(= \frac{2.7}{2.5})$ は先行研究 [3] では上下のスピーカ間隔が 2.5 m であったのに対し, 今回の実験における上下のスピーカ間隔が 2.7 m であることにより設けた補正係数である。また, $P'_V(=-1.25, \dots, 1.25)$ は立体像の離散化した縦位置を表し, 下記のように定義される。

$$P'_V = \begin{cases} \Delta d_V \text{ round} \left(\frac{P_V + 1.25}{\Delta d_V} \right) - 1.25 & (\Delta d_V \neq 0) \\ P_V & (\Delta d_V = 0) \end{cases} \quad (8)$$

但し, Δd_V は離散化した縦位置の間隔を表し, Fig. 3 の条件順に 0, 2.5, 1.25, 0.833, 0.625, 0 m となる。

2.2 実験計画・手順

視聴者はメガネあり立体映像が鑑賞可能で, かつ日常生活において聴力に異常の見られない 22–38 歳の計 9 名 (男性 5 名, 女性 4 名) である。評価手法には Scheffé の一対比較法 (浦の変法)[5] を用いた。聴取実験の流れ図を図 4 に示す。まず, 評価指標を 2 種

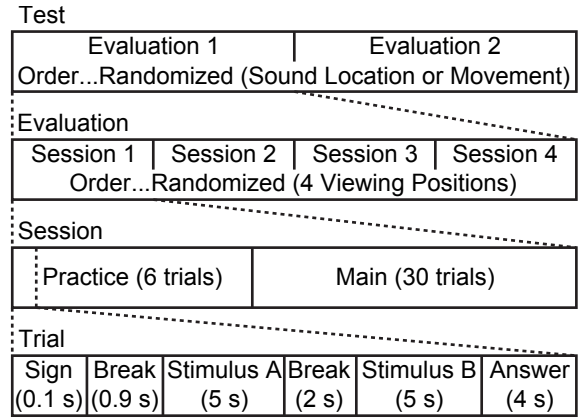


Fig. 4 Flowchart of the audio-visual experiment.

Table 1 Scale of Scheffé's paired comparison.

Grade	Judgment
3	Very good
2	Fairly good
1	Little good
0	The same
-1	Little bad
-2	Fairly bad
-3	Very bad

類 (「音の位置の一致度」「音の動きの一致度」) 設定し, 実験全体を評価指標ごとに 2 つに分割した。「音の位置の一致度」とは「星やボール (Fig. 2 内の赤色の枠線) の音が常に映像の位置で鳴っているように感じるか」ということを表し, 「音の動きの一致度」は「UFO (Fig. 2 内の黄色の枠線) の音が常に映像の位置に合わせて動いているように感じるか」ということを表す。さらに, 視聴位置ごとに分割を行い, 計 8 つのセッションを設定した。各セッションにおいて 6 回の練習試行の後, 30 回の本試行を行った。練習試行は図 3 における 3 つの実験条件 [条件 (a), (b), (f)] の順列の総数 ($3 \times 2 = 6$) であり, 本試行は図 3 における 6 つの実験条件 [条件 (a)-(f)] の順列の総数 ($6 \times 5 = 30$) である。評価指標, 視聴位置及び試行の提示順序は視聴者ごとにランダム化した。

視聴者には先に提示した刺激音 (A) を基準にして後に提示した刺激音 (B) の音の位置や動きの一致度を聴取者に評定させた。その際, 評定は表 1 に示すような 7 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。

2.3 実験結果及び考察

得られた実験結果を 8 種類 [= 評価指標 (2) \times 視聴位置 (4)] のセッションごとに Scheffé の一対比較法

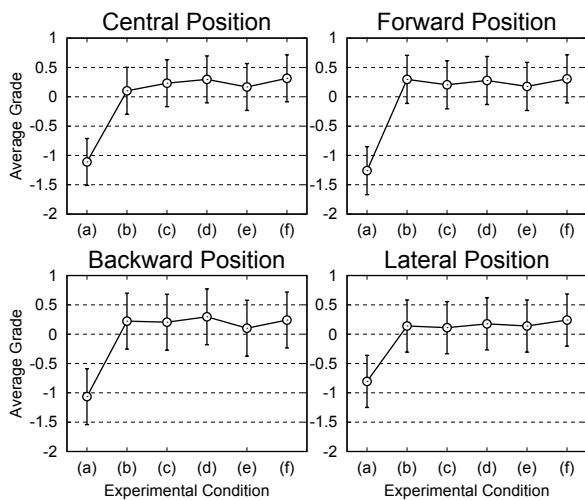


Fig. 5 Results of the audio-visual experiment (Sound location).

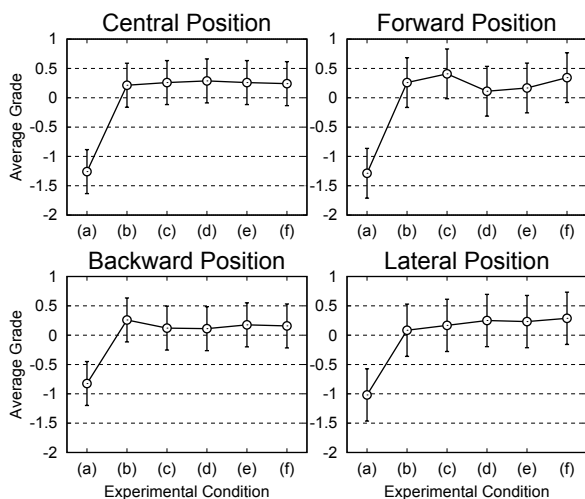


Fig. 6 Results of the audio-visual experiment (Sound movement).

(浦の変法)に基づいて分散分析を行ったところ、すべてのセッションにおいて実験条件に基づく主効果が0.1%有意であった。その結果、音の再生条件による臨場感への影響に有意差が見られると判断されたので、セッションごとに平均評定値を算出し、評定値の差によって音源の垂直位置の離散化による影響を検討することにした。

平均評定値をセッションごとに Figs. 5-6 に示す。エラーバーはヤードスティックによる95%信頼区間を表す。条件 (a) では音像が左端や右端に偏っており、映像位置と音像位置のずれを明らかに知覚できるので、他の条件よりも平均評定値が有意に低くなっている。ゆえに、この結果から見ると視聴者は臨場感の違いをおおむね正しく識別していると考えられる。

検討した音の再生条件の中で最も平均評定値が高い条件を基準として音源の垂直位置の離散化による

臨場感への影響を検討する。条件 (b)-(f) の場合は全てのセッションにおいて最も平均評定値が高い条件との有意差は見られない。つまり、音源の垂直位置を2段階に離散化したとしても、視聴者は離散化していない音との臨場感の違いを識別できないということが言える。従って、MVP方式を用いて視聴覚システムを構築する場合、音源の垂直位置は2段階に離散化することが可能であると考えられる。

3 おわりに

本発表では、我々が従来提案している立体音響システム (MVP方式) において、音源の垂直位置を離散化した場合にどのくらい臨場感に影響があるのかを視聴覚実験によって検討した。その結果、音源の垂直位置を2段階に離散化しても視聴者は臨場感の違いを識別することができないということが分かった。

これまでの研究では MVP方式を放送や通信に適用する場合において音源信号と立体像の位置情報を伝送することを前提としているが、多人数が参加する遠隔通信会議システムでは人数が多くなりすぎると伝送量が非常に多くなってしまふ。このような問題を解決するために、今後は立体像の位置情報を伝送する必要のない MVP方式のための收音用マイクロホンアレイを開発する必要がある。

謝辞 視聴覚実験の実施にあたり、実験環境の構築にご協力下さった岩澤昭一郎主任研究員、立体映像を提供して下さいた牧野真緒研究員に感謝の意を申し上げます。本発表における視聴覚実験は独立行政法人情報通信研究機構倫理規定に従い実施した。

参考文献

- [1] 榎並, 画像ラボ, 18 (9), 41-45, 2007.
- [2] 安藤ら, 情報通信研究機構報道発表, 2011. <http://www.nict.go.jp/press/2011/01/25-1.html>
- [3] Kimura *et al.*, ITE Trans. on Media Tech. and App., 2 (1), pp. 33-45, 2014.
- [4] 木村ら, 音講論 (春), pp. 953-956, 2014.
- [5] 日科技連官能検査委員会編, “官能検査ハンドブック,” 日科技連出版社, 1973, pp. 366-374.