

Multiple Vertical Panning を用いた立体音響システムのための聴取実験

木村 敏幸[†] 安藤 広志[†]

[†] 独立行政法人情報通信研究機構 ユニバーサルコミュニケーション研究所
〒 619-0288 京都府相楽郡精華町光台 2-2-2
E-mail: †{t-kimura,h-ando}@nict.go.jp

あらまし 本報告では、大画面裸眼立体映像提示システムに適合するために、Multiple Vertical Panning (MVP) 方式に基づいた新たな立体音響システムを提案する。音像の垂直位置はスクリーンの上下に配置した2つのスピーカ間のパニングによって合成する。音像の水平位置は2つのスピーカ対の位置によって制御する。これにより、複数の聴取者が同時に映像提示システムによって描写された立体像の位置に音像を感じることができる。提案システムの聴覚性能を評価するために、垂直軸方向に27個のスピーカを配置したスピーカアレイを用いて聴取実験を行った。スピーカアレイの上部と下部に配置した2個のスピーカの間でパニングを行い、音像を合成した。12名の聴取者に音を聴かせて合成音像の位置を回答させたところ、聴取者たちはスピーカアレイの上部と下部に配置した2個のスピーカの間の位置に合成音像を感じていることが分かった。

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、裸眼立体映像システム、聴取実験

Listening test for 3D audio system using multiple vertical panning

Toshiyuki KIMURA[†] and Hiroshi ANDO[†]

[†] Universal Communication Research Institute,
National Institute of Information and Communications Technology
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan
E-mail: †{t-kimura,h-ando}@nict.go.jp

Abstract In this paper, the novel three-dimensional (3D) audio system is proposed. The proposed system is based on Multiple Vertical Panning (MVP) method and matches to the glasses-free 3D video display system in which the size of screen is very large. The vertical position of sound images is synthesized by the panning between two loudspeakers placed at the top and bottom of screen. The horizontal position of sound images is controlled by the position of two loudspeakers. By the proposed system, multiple listeners can simultaneously feel the sound images at the position of 3D objects depicted by the video display system. In order to evaluate the auditory performance of the proposed system, the listening test was designed by using the loudspeaker array in which twenty-seven loudspeakers were aligned on the vertical line. Sound images were synthesized by the panning between two loudspeakers placed at the top and bottom of the loudspeaker array. Twelve listeners listened to a sound and reported the position of synthesized sound images. As a result, it was indicated that listeners could feel the synthesized sound images at the position between two loudspeakers placed at the top and bottom of the loudspeaker array.

Key words Ultra-realistic communication, 3D audio system, vertical panning, glasses-free 3D video display system, listening test

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], [2]. 立体映像技術や立体音響技術によって三次元空間上に映像や音響をより

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

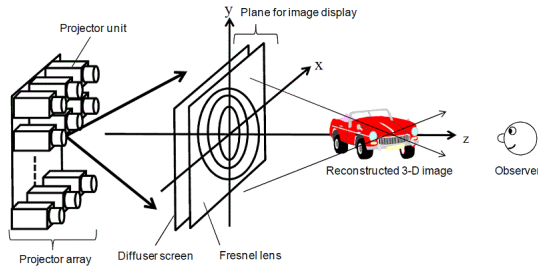


図 1 大画面裸眼立体映像提示システムの基本構成 [4]

Fig. 1 Basic configuration of a large glasses-free 3D video display system [4].

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、70 インチ [3], [4] 及び 200 インチ [5], [6] 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。図 1 にシステムの基本構成を示す。プロジェクタアレイを構成する各プロジェクタより視差付き映像をフレネルレンズに向かって再生する。その際、フレネルレンズの前に配置した拡散スクリーンの拡散特性（水平方向には狭い角度で拡散，垂直方向には広い角度で拡散）によって視差付き映像は水平方向にのみ提示される。これにより、観察者は水平位置に応じた視差付き映像を観察できるので、特殊な眼鏡なしに同時に複数の人が見る位置に応じた自然な立体像を観察できるようになる。

しかし、今までに開発してきたシステムは視覚のみを提示するもので、それ以外の感覚は提示できなかった。よりリアルな臨場感を表現するためには複数の感覚を同時に提示できるシステムを開発する必要がある。特に、聴覚を提示するためには大画面裸眼立体映像提示システムに適合した立体音響システムを開発する必要がある。そのためには、大画面裸眼立体映像提示システムの原理から下記の技術要件を満たす必要がある。

- (1) 同時に複数の人が観察しているので、複数の人がどこで聞いても立体像の位置から音が出ているように聞こえること。
- (2) 特殊な眼鏡なしに観察しているので、何も装着せずに音が聞けること。
- (3) 映像提示システムの妨げになるので、プロジェクタアレイとスクリーンの間には何も置かないこと。

(1) の条件から、一点で受聴することを前提としたシステム（ステレオホニック [7], 5.1ch システム [8]）は適用することが困難である。(2) の条件から、ヘッドホンの使用を前提としたシステム（バイノーラル [9]）は適用することが困難である。(3) の条件から、プロジェクタアレイとスクリーンの間にスピーカアレイを配置することを前提としたシステム（22.2ch システム [10], 高次アンビソニック [11], 波面合成法 [12]）は適用することが困難である。また、通信会議システムとしての応用を考えた場合、收音用マイクロホンもスクリーンと観察位置の間に配置できないので、スクリーンと聴取位置の間に收音用マイクロホンを配置することを前提としたシステム（トランスオーラル [13], 境界音場制御 [14]）も適用することが困難である。つまり、上記の技術要件を全て満たすためには今までに提案されてきた立体音響システムでは対応することができない。従って、この問題

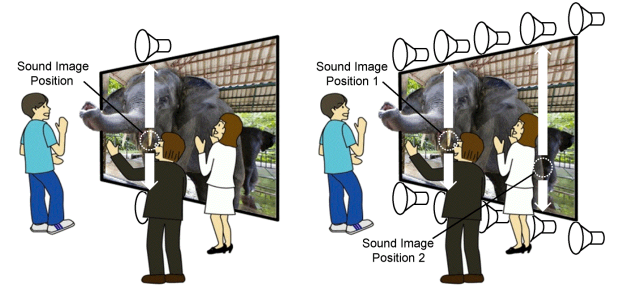


図 2 提案する立体音響システムの基本構成

Fig. 2 Basic configuration of the proposed 3D audio system.

を解決するためには今まで提案されてきた立体音響システムとは異なる観点で新たにシステムを開発する必要がある。

そこで、本報告ではこれまでの立体音響システムとは異なる観点（Multiple Vertical Panning (MVP) 方式）に基づいた新たな立体音響システムを提案する。2 章では提案システムの原理について述べ、提案システムが大画面裸眼立体映像提示システムに適合していることを示す。3 章では提案システムの聴覚性能を評価するために実施した聴取実験について述べ、提案システムが十分な聴覚性能を持っていることを示す。

2. 提案システムの原理

図 2 に提案システムの基本構成を示す。まず、図 2 の左側に示すように、立体像の位置の上下に 2 個のスピーカを配置する。その際、大画面裸眼立体映像提示システムによる立体像の奥行きは図 1 におけるフレネルレンズの位置にほぼ近いので、2 個のスピーカはスクリーンの上下に配置する。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、聴取者は 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、聴取者は立体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スクリーンの上下に配置した 2 個のスピーカだけなので、複数の聴取者はどこにいても常に立体像の位置で音が鳴っているように感じるができる。

さらに、図 2 の右側に示すように、スクリーン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スクリーン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、複数の聴取者はどこにいても常に大画面裸眼立体映像提示システムがスクリーン上に描写する立体像の位置で音が鳴っているように感じるができる。提案したシステムでは聴取者はヘッドホンを装着する必要がない。また、スピーカはスクリーンの上下に配置しているので、プロジェクタアレイとスクリーンの間には何も配置されていない。また、通信会議システムとしての応用を考えた場合、話者の声を直接收音するだけで良く、收音用マイクロホンの配置場所は限定されないもので、スクリーンと聴取位置の間に配置する必要がない。従って、提案したシステムは 1 章で述べた 3 つの技術要件を全て満たしていると言える。

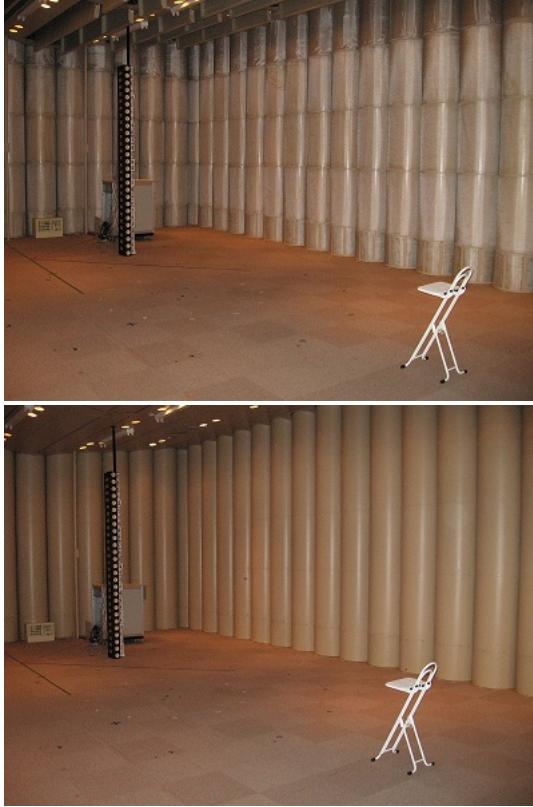


図 3 実験環境の概観図 (上: 残響時間 140 ms, 下: 残響時間 1030 ms)

Fig. 3 Image of experimental environment (Top: Reverberation time 140 ms, Bottom: Reverberation time 1030 ms).

3. 聴取実験

提案した立体音響システムの最小の構成要素はスクリーンの上下に配置した 2 個のスピーカである。提案システムの全体の聴覚性能は最小の構成要素の重ね合わせで表現できるので、スクリーンの上下に配置した 2 個のスピーカによる合成音像の高さ知覚を評価すれば、提案システムの聴覚性能を評価するには十分である。

また、過去に実施された垂直パニングの実験条件 [15]~[18] は今回検討している大画面裸眼立体映像提示システムの条件には当てはまらないため、2 章で述べた 2 個のスピーカ間の適切な音量差は不明のままである。そこで、本報告では 200 インチ画面を持った大画面裸眼立体映像提示システムのスクリーンの上下に配置することを想定した 2 個のスピーカによる合成音像の高さ知覚を評価する聴取実験を実施し、2 個のスピーカ間の適切な音量差を設定する垂直パニング曲線を検討する。

3.1 実験環境・条件

聴取実験は ATR 可変残響室 [19] にて行った。この部屋は壁面を構成する円筒及び天井回転板をモータで回転させることによって、部屋の残響時間を図 3 に示すように 140 ms(全吸音) から 1030 ms(全反射) にまで変化させることができる。部屋の暗騒音レベルは部屋の残響時間が 140 ms のときは 14 dBA、部屋の残響時間が 1030 ms のときは 22 dBA だった。

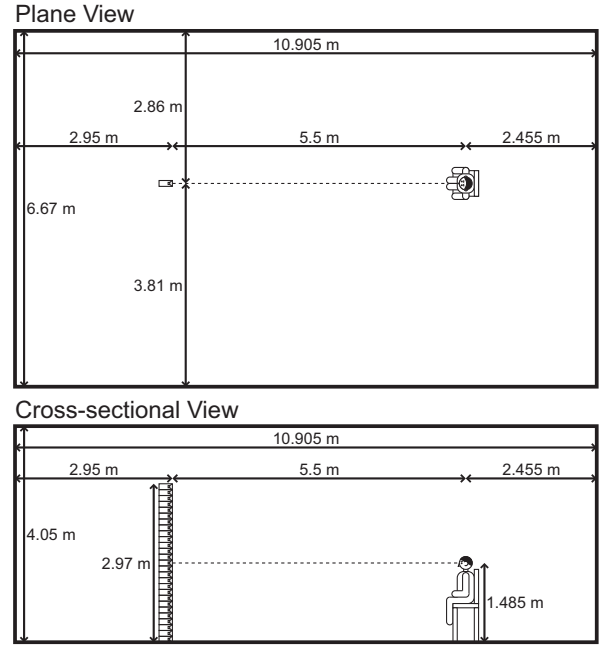


図 4 聴取実験における聴取者及びスピーカアレイの配置

Fig. 4 Position of the listener and the loudspeaker array in the listening test.

図 4 に示すように計 27 個のスピーカを垂直軸方向に配置した。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)) に取り付けしたものを用いた。従って、スピーカアレイ全体の高さは 2.97 m (=11 cm×27) となった。また、大画面裸眼立体映像提示システムの最適視聴距離を考慮してスピーカアレイから 5.5 m 離れた地点に聴取位置を設定した。聴取位置の高さは聴取者の耳の位置において 1.485 m とした。音圧レベルは聴取位置において約 70 dBA に設定した。

本実験で設定した実験条件を図 5 に示す。条件 (a), (b) における灰色のスピーカは音が再生されないスピーカを表す。(a) に示すパニング条件では上下に配置した 2 個のスピーカから音源信号 $s(n)$ に下記の式によって音量差をつけた音信号を再生した。

$$x_U(n) = a_U s(n) \quad (1)$$

$$x_D(n) = a_D s(n) \quad (2)$$

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_U , a_D はそれぞれのゲイン係数を表す (但し、 $a_U^2 + a_D^2 = 1$)。音量差 ΔA [dB] を下記のように定義すると、

$$\Delta A = 20 \log_{10} \frac{x_U(n)}{x_D(n)} = 20 \log_{10} \frac{a_U}{a_D} \quad (3)$$

a_U , a_D はそれぞれ下記の式のように表される。

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

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

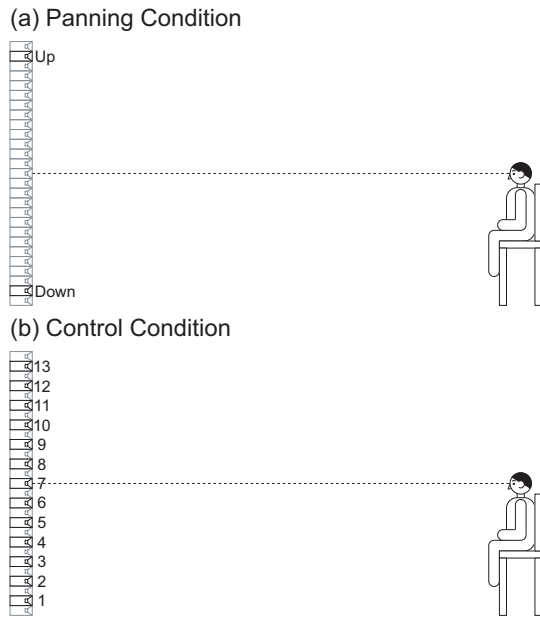


図 5 聴取実験で用いた実験条件

Fig. 5 Experimental conditions used in the listening test.

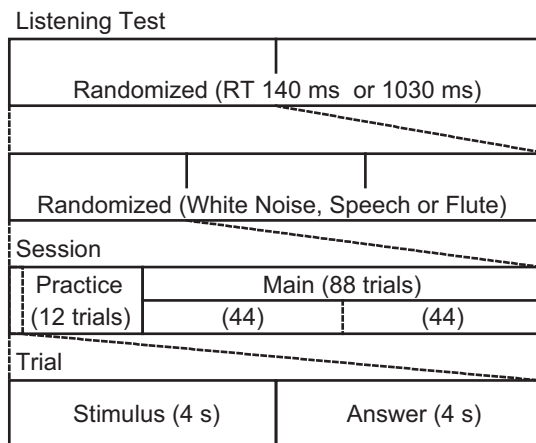


図 6 聴取実験の流れ図

Fig. 6 Flowchart of the listening test.

本実験では音量差 ΔA は -15 dB から 15 dB まで 1 dB 間隔で設定した。(b) に示す統制条件は聴取者の音像の高さ知覚能力を評価するために設定した。この条件では 13 個の中から選ばれた 1 個のスピーカから音源信号 $s(n)$ を直接再生した。

3.2 実験計画・手順

聴取者は日常生活において聴力に異常の見られない 21-32 歳の計 12 名 (男性 6 名, 女性 6 名) である。音源には白色雑音, 音声, フルートの 3 種類を用いた。聴取実験の流れ図を図 6 に示す。まず, 実験全体を残響時間ごとに 2 つに分割した。さらに, 音源ごとに分割を行い, 計 6 つのセッションを設定した。各セッションにおいて 12 回の練習試行の後, 88 回の本試行を行った。その際, 44 試行ごとに休憩時間を設けた。残響時間, 音源及び試行の提示順序は聴取者ごとにランダム化した。練習試行及び本試行の詳細を表 1 に示す。

聴取者には 27 個のスピーカの中から音が再生されたと感じた番号を回答用紙に記入するように教示した。その際, 図 7 に

表 1 聴取実験における練習試行及び本試行

Table 1 Practice and main trials in the listening test.

Element	Note
Practice (12)	= 7 conditions + 5 positions
Main (88)	= [31 conditions + 13 positions] × 2 repetitions

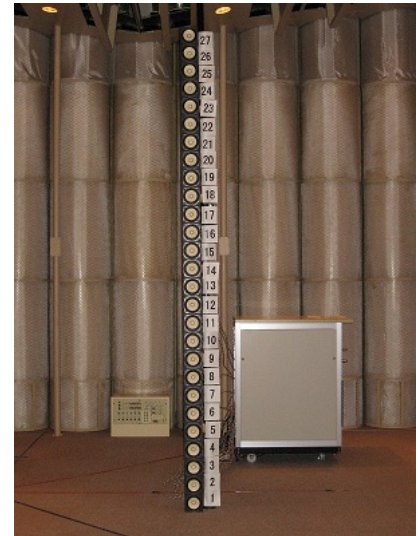


図 7 知覚した音像の高さと回答番号の関係図

Fig. 7 Relation between perceived heights of sound images and answer indexes.

示すようにスピーカの右側に配置した指標に従って回答させた。この番号は 1 から 27 までの 27 通りで, 14 番のスピーカが被験者の耳及び目の高さと同じ。2 つ以上の音が鳴っていると思った場合は, 2 つ以上の番号で回答させた。また, 回答の際には聴取者は頭部及び上半身を自由に動かすことができた。

3.3 実験結果及び考察

実験結果を分析するために, 聴取者の回答からスピーカ番号が複数記入されているものを全て除去した上で, 下記の式に従って聴取者が回答したスピーカ番号から知覚した音像の高さを算出した。

$$H_{\text{per}}[\text{m}] = (I_{\text{ans}} - 14) \times 0.11 \quad (6)$$

ただし, I_{ans} は回答したスピーカ番号, H_{per} は音像の高さを表す。

パニング条件における知覚した音像の高さの平均をプロットしたものを図 8 に示す。エラーバーは 95%信頼区間を表す。全ての条件において, 音量差がおよそ 0 dB の場合に知覚した音像の高さが 0 m (つまり, スクリーンの上下に配置した 2 個のスピーカの間点) 付近であり, 聴取者は 2 個のスピーカの間音像があると感じていることが分かる。特に, 音量差が -3 dB から 9 dB の範囲では知覚した音像の高さが音量差に対して直線的に推移しているように見える。そこで, -3 dB から 9 dB

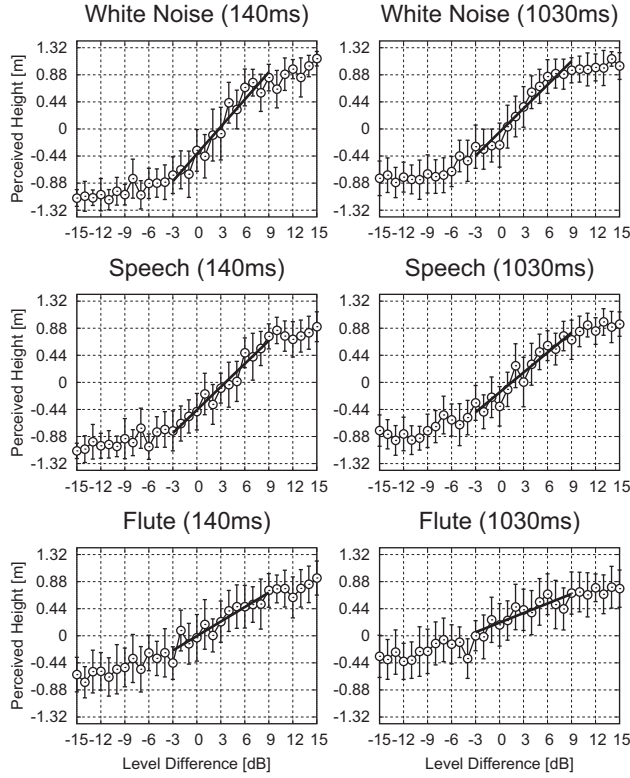


図 8 パニング条件における実験結果及び回帰直線

Fig. 8 Results and regression lines in the panning condition.

の範囲において直線回帰を行ったところ、それぞれの条件に対して下記のような結果が得られた。

(White noise, Reverberation time 140 ms)

$$H_{\text{per}} = 0.1475\Delta A - 0.4066 \quad (7)$$

(Speech, Reverberation time 140 ms)

$$H_{\text{per}} = 0.1253\Delta A - 0.4499 \quad (8)$$

(Flute, Reverberation time 140 ms)

$$H_{\text{per}} = 0.0784\Delta A - 0.0045 \quad (9)$$

(White noise, Reverberation time 1030 ms)

$$H_{\text{per}} = 0.1279\Delta A - 0.0510 \quad (10)$$

(Speech, Reverberation time 1030 ms)

$$H_{\text{per}} = 0.1079\Delta A - 0.1635 \quad (11)$$

(Flute, Reverberation time 1030 ms)

$$H_{\text{per}} = 0.0518\Delta A + 0.2130 \quad (12)$$

得られた回帰直線を図 8 中の太線にて示す。回帰直線が正しく推定されていることが分かる。

6 条件において得られた回帰直線の平均から、下記の式に示すようなパニング曲線 H_{pan} を算出した。

$$H_{\text{pan}} = \begin{cases} -1.32 & (\Delta A < -11.05) \\ 0.1065\Delta A - 0.1437 & (-11.05 \leq \Delta A \leq 13.74) \\ 1.32 & (\Delta A > 13.74) \end{cases} \quad (13)$$

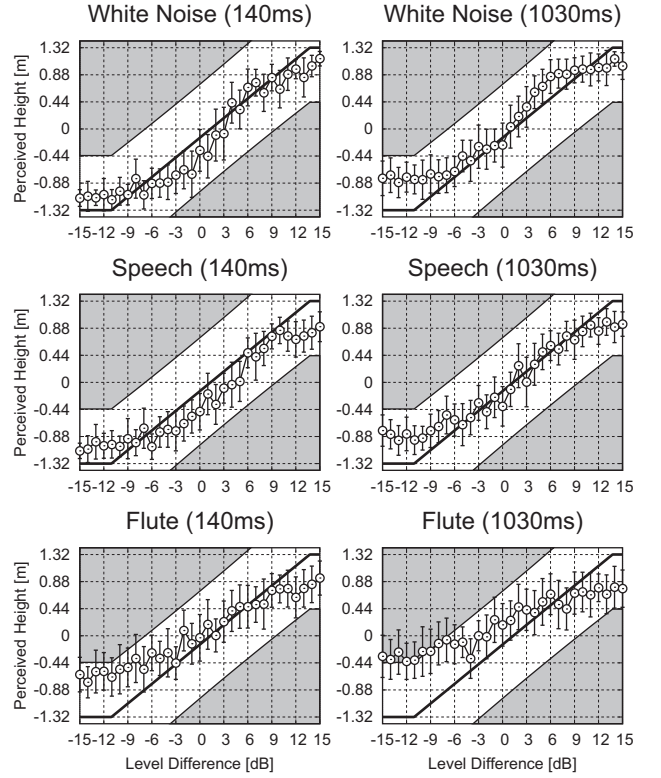


図 9 パニング条件におけるパニング曲線及び弁別閾

Fig. 9 Panning curves and differential limens in the panning condition.

さらに、知覚する音の高さの弁別閾 DL_{pan}^+ , DL_{pan}^- を下記の式に従って算出した。

$$DL_{\text{pan}}^+ = \tan(\tan^{-1}(H_{\text{pan}}/5.5) + \phi) \times 5.5 \quad (14)$$

$$DL_{\text{pan}}^- = \tan(\tan^{-1}(H_{\text{pan}}/5.5) - \phi) \times 5.5 \quad (15)$$

但し、 ϕ は音の高さの弁別角で、過去の研究結果 [20] を参考に 9 度とした。パニング条件の実験結果にパニング曲線及び弁別閾を追加した図を図 9 に示す。灰色の領域は弁別閾の外側を表しており、この領域に知覚した音の高さの平均がある場合、パニング曲線に従って提示した音像の高さと (映像などで提示される) 実際の音源の高さの違いを聴取者が判別できることを示している。残響時間 1030 ms で音源がフルートの場合以外の条件においては、灰色の領域に知覚した音の高さの平均が入っていないため、聴取者が音像の高さの違いを判別できないくらいにパニング曲線の聴覚性能が高いと言える。しかし、残響時間 1030 ms で音源がフルートの条件においては、スピーカの下側に音像を提示したときに灰色の領域に知覚した音の高さの平均が入っているため、このときには聴取者が音像の高さの違いを判別できる可能性がある。

一方、統制条件の実験結果及び弁別閾を図 10 に示す。エラーバーは 95%信頼区間を表す。統制条件における弁別閾 DL_{ctrl}^+ , DL_{ctrl}^- は下記の式に従って算出した。

$$DL_{\text{ctrl}}^+ = \tan(\tan^{-1}(H_{\text{pre}}/5.5) + \phi) \times 5.5 \quad (16)$$

$$DL_{\text{ctrl}}^- = \tan(\tan^{-1}(H_{\text{pre}}/5.5) - \phi) \times 5.5 \quad (17)$$

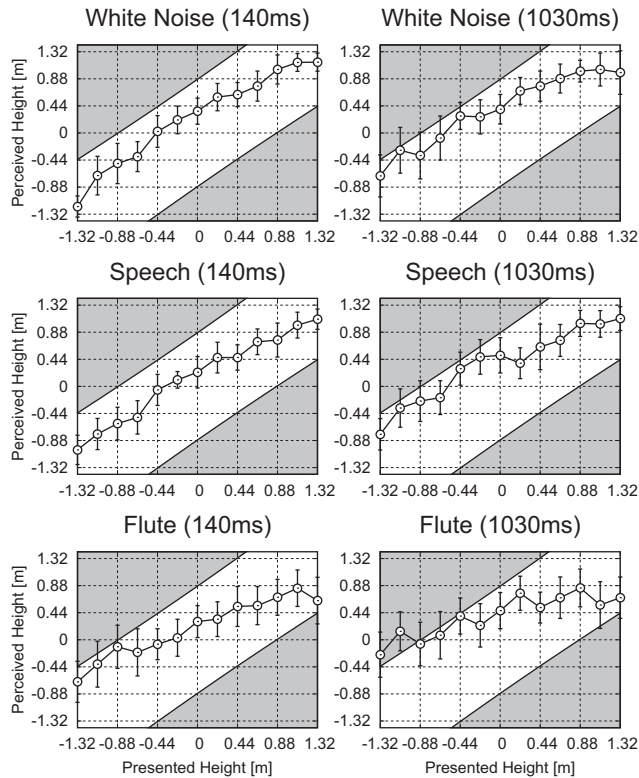


図 10 統制条件における実験結果及び弁別閾

Fig. 10 Results and differential limens in the control condition.

但し, $H_{pre} (= (I_{pre} - 14) \times 0.11)$ は提示した音の高さを表し, I_{pre} は提示したスピーカ番号を表す. 残響時間 1030 ms で音源がフルートの場合以外の条件においては, 灰色の領域に知覚した音の高さの平均が入っていないが, 残響時間 1030 ms で音源がフルートの条件においては, 下側のスピーカで音像を提示したときに灰色の領域に知覚した音の高さの平均が入っている. 従って, 音源がフルートのときは, 残響時間によって聴取者が音像の高さ知覚そのものができていない可能性がある. そこで, 今後残響時間を詳細に変更した上で更なる追加実験を行う必要があると考えられる.

4. まとめ

本報告では, 従来提案されている大画面裸眼立体映像提示システムに適合するために, Multiple Vertical Panning (MVP) 方式に基づいた新たな立体音響システムを提案した. 提案システムの聴覚性能を評価するために, 垂直軸方向に 27 個のスピーカを配置したスピーカアレイを用いて聴取実験を行ったところ, 残響時間が 1030 ms で音源がフルートの場合以外は聴取者たちが提示した音像の高さの違いを判別できないほど提案システムの聴覚性能が高いということが分かった.

今後は残響時間を詳細に変更した上で音像の高さ知覚について更に検討する必要がある. また, 大画面裸眼立体映像提示システムと統合した上で視聴覚実験を行い, 視聴覚システムにおける提案したシステムの有効性を検討する必要がある.

文 献

[1] 榎並和雅, “超臨場感システムの研究推進に向けて - (独) 情報通

信研究機構の研究と、産学官フォーラム URCF の活動 -,” 画像ラボ, vol.18, no.9, pp.41–45, Sept. 2007.

[2] K. Enami, “Research on ultra-realistic communications,” ECTI Trans. Electr. Eng. Electron. Commun., vol.6, no.1, pp.22–25, Feb. 2008.

[3] 河北真弘, 岩澤昭一郎, 矢野澄男, R. Lopez-Gulliver, 吉田俊介, S. Gurbuz, D. Moldovan, 安藤広志, 井ノ上直己, “眼鏡なし立体画像システム,” 情報通信研究機構季報, vol.56, no.1/2, pp.43–52, March/June 2010.

[4] M. Kawakita, S. Iwasawa, G. Sabri, and N. Inoue, “Development of glasses-free 3D video system,” Proc. International Universal Communication Symposium, pp.323–327, Oct. 2011. http://www.iucs2011.org/IUCS2011_USB_Mod_8th_1026/paper/Development%20of%20Glasses-free%203D%20Video%20System.pdf.

[5] 安藤広志, 河北真弘, 岩澤昭一郎, “世界初「200 インチの自然な裸眼立体表示技術」の開発に成功,” 情報通信研究機構報道発表, <http://www.nict.go.jp/press/2011/01/25-1.html>, Jan. 2011.

[6] H. Ando, M. Kawakita, and S. Iwasawa, “The world’s largest 200-inch glasses-free 3D display has been successfully developed,” NICT Press Release, http://www2.nict.go.jp/pub/whatsnew/press/h22/110125/110125_e.html, Jan. 2011.

[7] A.D. Blumlein, “Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems,” British Patent, no.394325, 1931.

[8] ITU-R Recommendation BS.775-1, “Multichannel stereophonic sound system with and without accompanying picture,” 1992-1994.

[9] J. Blauert, “Spatial hearing,” pp.372–392, MIT Press, Cambridge, Mass, revised edition, 1997.

[10] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando, “A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV),” SMPTE Mot. Imag. J., vol.117, no.3, pp.40–49, April 2008.

[11] M.A. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” J. Audio Eng. Soc., vol.53, no.11, pp.1004–1025, Nov. 2005.

[12] A.J. Berkhout, D. deVries, and P. Vogel, “Acoustic control by wave field synthesis,” J. Acoust. Soc. Am., vol.93, no.5, pp.2764–2778, May 1993.

[13] M.R. Schroeder, D. Gottlob, and K.F. Siebrasse, “Comparative study of european concert halls: Correlation of subjective preference with geometric and acoustic parameters,” J. Acoust. Soc. Am., vol.56, no.4, pp.1195–1201, Oct. 1974.

[14] S. Ise, “A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems,” ACUSTICA - Acta Acustica, vol.85, no.1, pp.78–87, January/February 1999.

[15] 黒住幸一, “垂直面内に配置した 2 つのスピーカによる合成音像の定位,” 音講論, no.2-2-9, pp.437–438, Oct. 1989.

[16] 黒住幸一, “垂直面内に配置したスピーカによる合成音像の移動感,” 音講論, no.1-7-21, pp.359–360, Sept. 1990.

[17] 播摩敏雄, 鈴木陽一, 曾根敏夫, “正中面におかれた 2 音源による合成音像定位,” 音講論, no.3-3-2, pp.461–462, Sept. 1995.

[18] V. Pulkki, “Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning,” J. Audio Eng. Soc., vol.49, no.9, pp.753–767, Sept. 2001.

[19] 平原達也, カールミュラー, 東倉洋一, “ATR 可変残響室の構造と音響特性,” 音響誌, vol.48, no.5, pp.301–308, May 1992.

[20] J. Blauert, “Spatial hearing,” pp.37–50, MIT Press, Cambridge, Mass, revised edition, 1997.