Vol.2022-MUS-134 No.10
Vol.2022-SLP-142 No.10
2022/6/17

3D 音響信号の効率的なバイノーラルレンダリングに関する検討

水谷勇貴[1]　西口正之[2]　渡邉貫治[2]　安倍幸治[2]　石川智一[3]　榎本成悟[3]

概要：AR・VR において頭部伝達関数を直接用いてリアルタイムで音環境をバイノーラルレンダリングする場合，受聴者と音源の位置関係の変化や，再現する環境の特性により演算量が膨大となってしまう場合がある．提案法ではパンニング処理をレンダリングに応用することを検討し，定位感や音声の品質の劣化を最小限にしつつ演算量の低減を図った．音声の品質と定位について主観評価実験によって，他の手法との比較を行った．

A study on efficient binaural rendering of 3D audio signals

YUKI MIZUTANI[†1]　MASAYUKI NISHIGUCHI[†2]　KANJI WATANABE[†2]
KOJI ABE[†2]　TOMOKAZU ISHIKAWA[†3]　SEIGO ENOMOTO[†3]

1. はじめに

近年，AR や VR 機器の普及が進み，3D の音環境をレンダリングする機会が増加している．頭部伝達関数を用いてリアルタイムで音空間をバイノーラルレンダリングする際，受聴者と音源の位置関係の変化や再現する環境の特性などによっては演算量が膨大となってしまう．そのような場合において，定位感や音質の劣化を抑えつつ，演算量を低減することを目的として本検討を行った．

本検討ではパンニング処理をレンダリングに応用することで演算量の低減を図り，SNR での客観的評価及び定位実験と MUSHRA 法での主観評価実験を行った．

2. 提案手法での処理の概要

パンニングでは，任意の目的信号をいくつかの代表点の方向（例えば 2 方向）に振り分けて，それらの合成音像で目的信号を再現する．パンニングの考え方をバイノーラルレンダリングに応用すると，目的信号にその方向の頭部インパルス応答（HRIR）を直接畳み込んで当該信号の音像を作るのではなく，代表点に振り分けられた信号にその代表点方向の HRIR を畳み込み，それらの合成音像で目的信号の音像を生成するという方法が考えられる．それにより，目的信号の数が増えても，代表点方向の，例えば 2 方向の HRIR の畳み込みだけで全ての目的信号の音像を生成することが出来，演算量の削減が可能になる．

スピーカー再生で従来より一般的に用いられてきた sin 則[1]によるパンニング方法では，合成に用いる 2 つの信号に，音源の方向に基づいて計算されたゲインをかけ，目的となる信号を合成していた[2][3]．この手法を本論文では従来法と呼ぶことにする．

本検討の提案手法では，最適なパンニング処理は音源方向の HRIR を事前に決めた複数の代表方向の HRIR でいかに良く近似するかという問題と等価であるという点に着目し，2 つの代表方向の HRIR で目的音方向の HRIR（オリジナル HRIR と呼ぶ）を合成する方法について検討した．まず合成に用いる 2 つの代表方向の HRIR をそれぞれ目的方向のオリジナル HRIR との相互相関が最大となるように時間シフトを行い，次に，それらに適当なゲインをかけて足し合わせることで合成 HRIR を生成する．その際，オリジナル HRIR と合成 HRIR の誤差ベクトルが最小となるようなゲインを用いる．上記の処理を行うことにより，合成 HRIR でオリジナル HRIR を模擬するのである．

オリジナル HRIR の時間波形をベクトルと見立てて $\vec{x}$ と記す．同様に，再現のために用いる二つの代表方向の HRIR をそれぞれ$\overrightarrow{x_1}$，$\overrightarrow{x_2}$とし，$\overrightarrow{x_1}$にかけるゲインをA，$\overrightarrow{x_2}$にかけるゲインをBとする．なお，合成に用いる二つの HRIR $\overrightarrow{x_1}, \overrightarrow{x_2}$はそれぞれオリジナル HRIR $\vec{x}$との相互相関が最大となるような時間シフトを行った後のものである．オリジナル HRIR と合成された HRIR との誤差をエラーベクトル$\vec{e}$とすると，

$$\vec{e} = \{\vec{x} - (A\overrightarrow{x_1} + B\overrightarrow{x_2})\}$$

となる．エラーベクトル$\vec{e}$とベクトル$\overrightarrow{x_1}$，$\overrightarrow{x_2}$によって張られる面が直交するときにエラーベクトルの大きさが最小になる．従って，エラーベクトルと代表方向の各ベクトルが直交するため，それらの内積は 0 となる．

$$\{\vec{x} - (A\overrightarrow{x_1} + B\overrightarrow{x_2})\}\cdot\overrightarrow{x_1} = 0$$
$$\{\vec{x} - (A\overrightarrow{x_1} + B\overrightarrow{x_2})\}\cdot\overrightarrow{x_2} = 0$$

上記の式より，ゲイン A，B は以下のように求まる．

1 秋田県立大学大学院
Graduate School of　Akita Prefectural University
2 秋田県立大学
Akita Prefectural University
3 パナソニック ホールディングス株式会社
Panasonic Holdings Corporation

Vol.2022-MUS-134 No.10
Vol.2022-SLP-142 No.10
2022/6/17

$$A = \frac{\overrightarrow{x_1}\cdot\vec{x}|\overrightarrow{x_2}|^2 - \overrightarrow{x_2}\cdot\vec{x}(\overrightarrow{x_1}\cdot\overrightarrow{x_2})}{|\overrightarrow{x_1}|^2|\overrightarrow{x_2}|^2 - |\overrightarrow{x_1}\cdot\overrightarrow{x_2}|^2}$$

$$B = \frac{\overrightarrow{x_2}\cdot\vec{x}|\overrightarrow{x_1}|^2 - \overrightarrow{x_1}\cdot\vec{x}(\overrightarrow{x_1}\cdot\overrightarrow{x_2})}{|\overrightarrow{x_1}|^2|\overrightarrow{x_2}|^2 - |\overrightarrow{x_1}\cdot\overrightarrow{x_2}|^2}$$

提案手法では，時間シフトと上記のように算出されたゲインを用いてパンニング処理を行った．

3. 客観評価

オリジナル HRIR と上記手法により合成された HRIR との SNR の平均値を算出した．客観評価では，水平面の全周 360° の範囲において，2° 間隔の HRIR(FABIAN[4])を用いた．時間シフト無し，sin 則によるゲインを用いたパンニングによる HRIR の合成を従来法とし，提案手法との比較の結果を以下の表 1, 2 に示す．パンニングにおける代表点の位置のパターンは 4 方向_斜め（45°, 135°, 225°, 315°），4 方向_縦横（0°，90°，180°，270°），6 方向（30°，90°，150°，210°，270°，330°）の 3 パターンである．これらのパターンを図 1 に示す．左から順に 4 方向_斜め，4 方向_縦横，6 方向である．

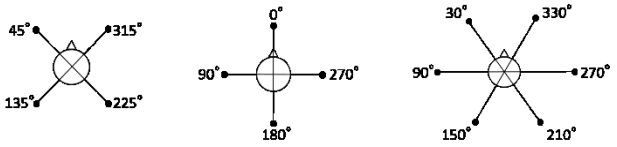

図 1　各代表点パターン

Figure 1　Each pattern of representative points.

ある方位の HRIR は，その近傍の 2 つの代表点の HRIR を用いて合成される．例として，代表点の位置が 4 方向_斜めの場合であれば，45° から 135° までの 90° の範囲角に含まれる点の HRIR は，45° の HRIR と 135° の HRIR の 2 つを用い，それらに各々時間シフトとゲインをかけ，足し合わせることで合成される．表 1，表 2 に示す SNR の平均値は, 360° を 2° 間隔で分割した 180 個のオリジナル HRIR と上記手法により合成された同間隔の HRIR との SNR の平均である. 尚, この SNR はオリジナルの HRIR を信号 S, 合成した HRIR とオリジナル HRIR との差分をノイズ N として算出したものである．代表点と角度が一致する点に関しては，SNR が無限大となってしまうため，SNR の平均値を求める際には除外した．

表 1　各条件での SNR の平均値（右耳）

Table 1　Average SNR under each condition(right ear).

	従来法	提案法
4 方向_斜め	-2.86[dB]	6.67[dB]
4 方向_縦横	-1.46[dB]	7.22[dB]
6 方向	-0.80[dB]	8.30[dB]

表 2　各条件での SNR の平均値（左耳）

Table 2　Average SNR under each condition(left ear).

	従来法	提案法
4 方向_斜め	-2.25[dB]	6.93[dB]
4 方向_縦横	-1.18[dB]	7.83[dB]
6 方向	-0.47[dB]	8.82[dB]

4. 主観評価

定位実験と MUSHRA 法により，主観評価を行った．各実験での共通している条件は以下の表 3 に示す．主観実験では FABIAN の HRIR ではなく被験者の頭の HRIR を測定し，使用した．この HRIR は，水平面の全周 360° の範囲において，15° 間隔のものである．

表 3　主観評価実験での各条件

Table 3　Each condition of subjective evaluation experiment.

提示角度	0° ～345°，15° 間隔
使用 HRIR	オリジナル HRIR，従来法による合成 HRIR，提案法による合成 HRIR
代表点パターン	4 方向_斜め，4 方向_縦横，6 方向
提示音圧レベル	70dB（オリジナル，0°）
使用ヘッドホン	SENNHEISER HD 580 precision
被験者数	1 名（男性，20 代）

4.1 定位実験

使用音源はオリジナル HRIR，従来法のパンニングで合成した HRIR，および提案法によるパンニングによって合成した HRIR を各々ホワイトノイズに畳み込んだものである．また，各音源の繰り返し回数は 2 回として定位実験を行った．定位実験の結果を以下に示す．横軸が被験者に提示したホワイトノイズに畳み込んだ HRIR の方位である．また，縦軸は回答方位である．

Vol.2022-MUS-134 No.10
Vol.2022-SLP-142 No.10
2022/6/17

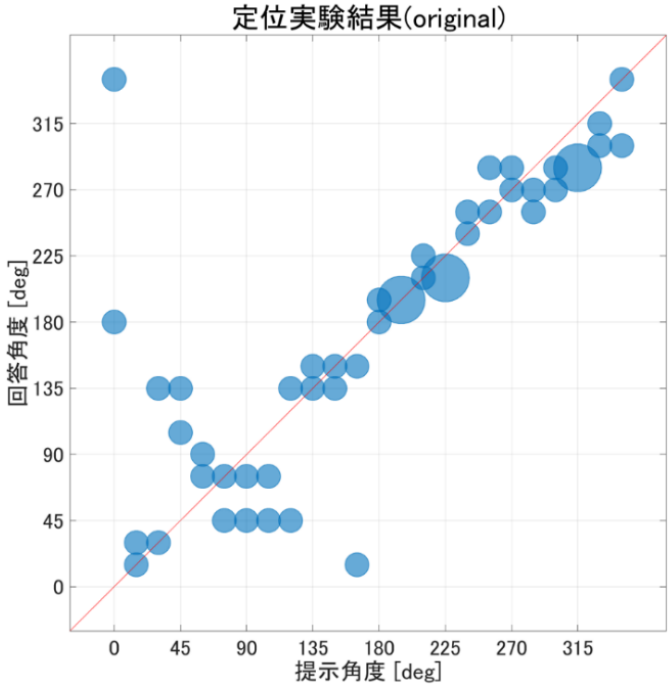

図 2　定位実験の結果（オリジナル）

Figure 2　Result of localization experiment(Original).

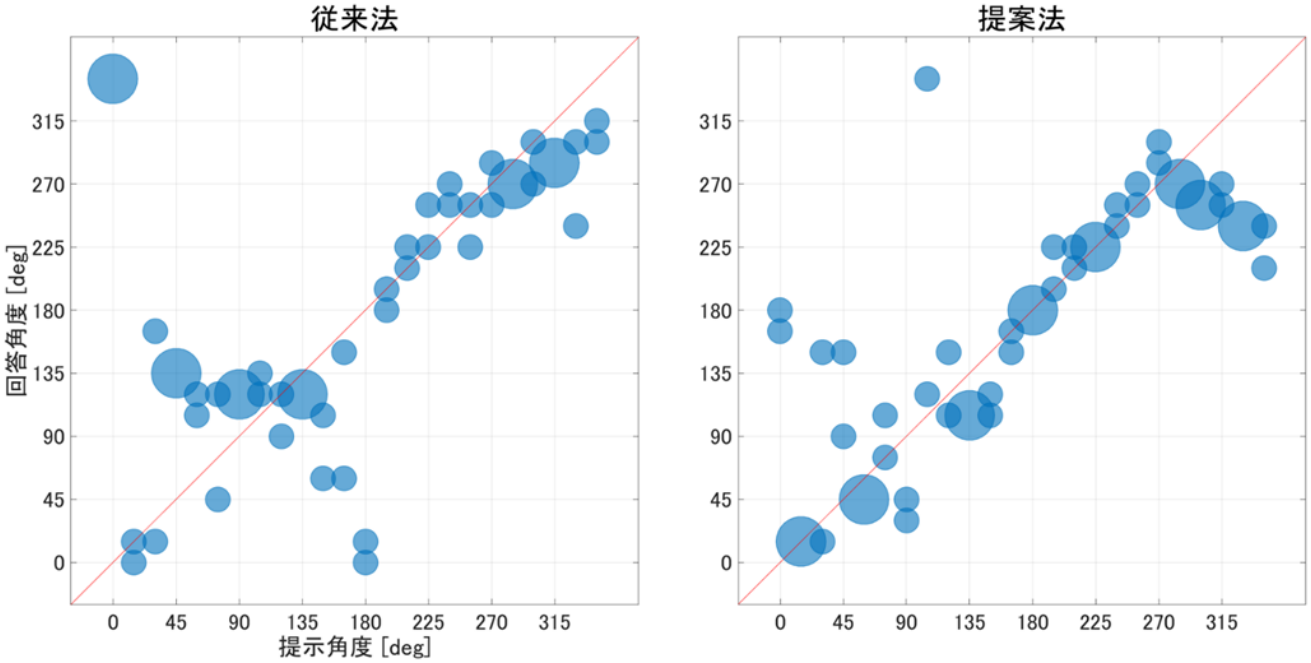

図 3　定位実験の結果（4 方向_斜め）

Figure 3　Result of localization experiment(four directions_diagonal).

Vol.2022-MUS-134 No.10
Vol.2022-SLP-142 No.10
2022/6/17

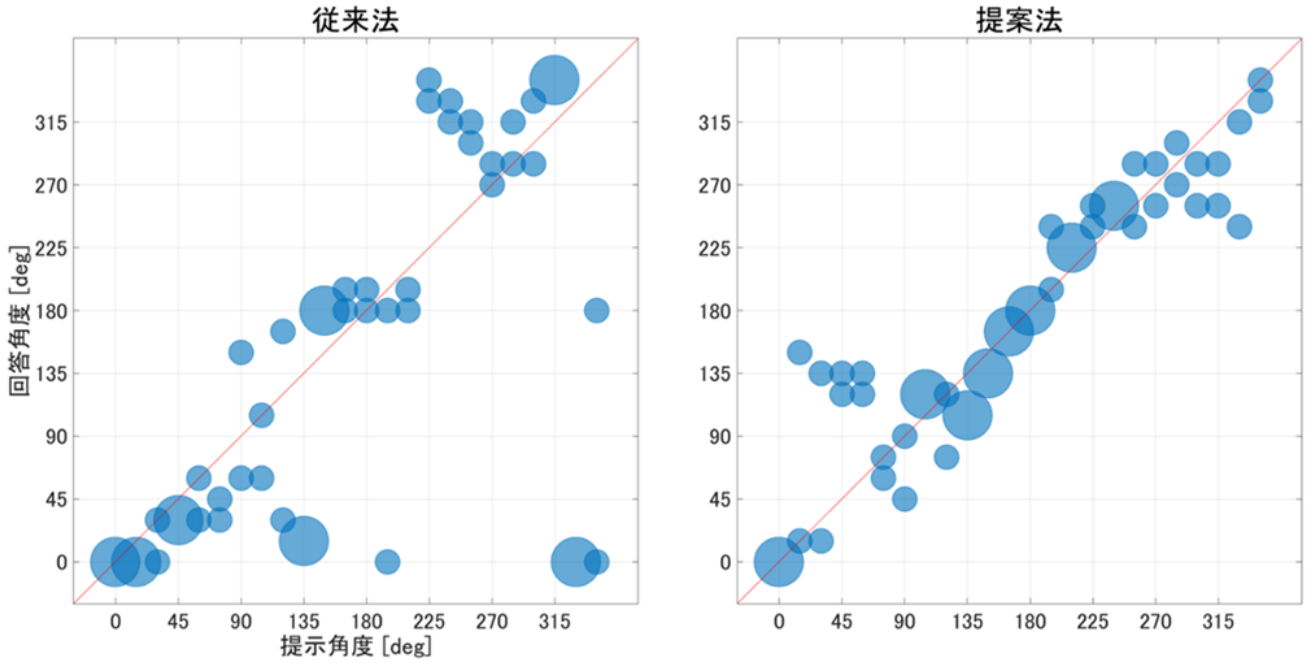

図 4　定位実験の結果（4 方向_縦横）

Figure 4　Result of localization experiment(four directions_vertical and horizonal).

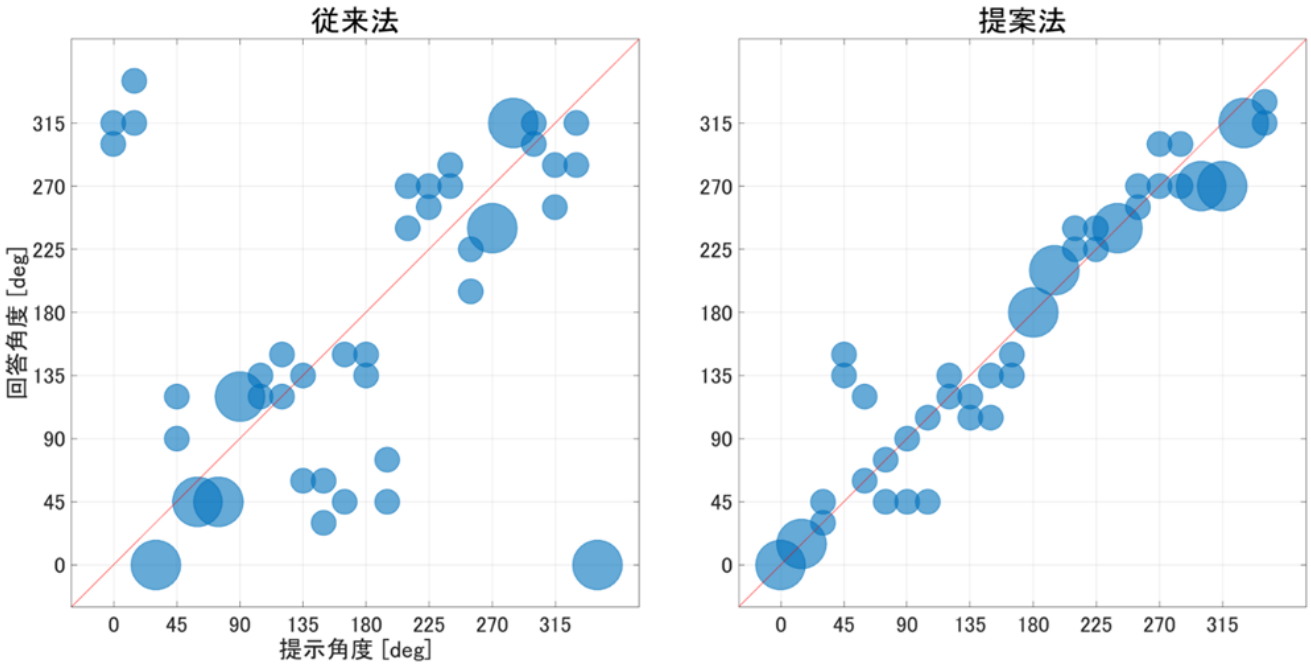

図 5 定位実験の結果（6 方向）

Figure 5　Result of localization experiment(six directions).

提案法では，回答角度が提示角度近傍に位置することが多くなっており，提案法の定位感は従来法に比べ全体的に向上していたと言える．特に 6 方向の場合では前後誤りも少なく，かなり正確な定位となっていた．この結果は表 1，表 2 に示す SNR の結果ともほぼ整合したものとなっている．上記の定位実験の結果より，提案法の処理が有効であったと考えられる．

4.2 MUSHRA 法での評価

MUSHRA 法での音声の評価実験を行った．使用音源は 4.1 節同様，オリジナル HRIR，従来法によるパンニング，提案法によるパンニングによって合成した HRIR を各々jvs コーパス[5]（男声 1 種，女声 1 種）に畳み込んだものであ

Vol.2022-MUS-134 No.10
Vol.2022-SLP-142 No.10
2022/6/17

る．MUSHRA 法での実験結果を以下に示す．図 6，図 7 の各バーは HRIR の生成方法毎の全周 24 方向の評価点の平均値と 95%信頼区間を示す．

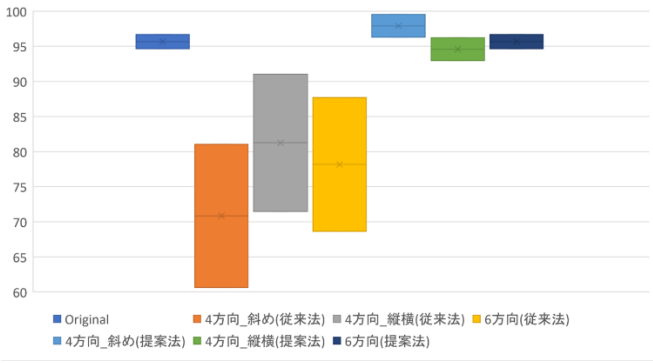

図 6　MUSHRA 法での実験結果（男声）
Figure 6　Result of experiment by MUSHRA(male voice).

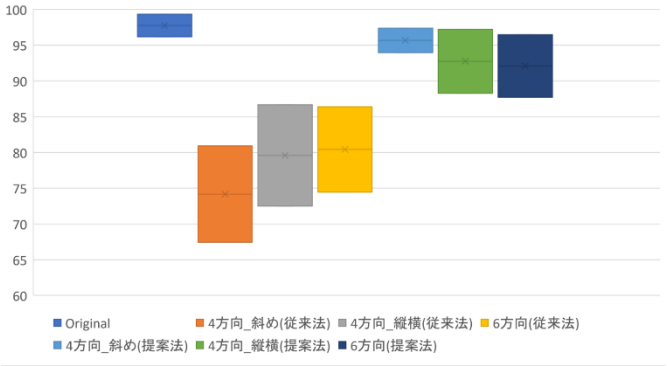

図 7　MUSHRA 法での実験結果（女声）
Figure 7　Result of experiment by MUSHRA(female voice).

MUSHRA 法での実験結果より，すべての代表点の位置のパターンにおいて従来法に比べ提案法での評価点は高くなっている．また提案法では全ての評価においてオリジナルと 95%信頼区間がオーバーラップしている事が分かる．

この評価点の差異の最も大きな要因は，リファレンス音であるオリジナルの HRIR を畳み込んだ音声との方位感の違いによるものである．また，遠近感の違いも影響している．

5.　演算量について

提案手法を用い，予め時間シフト量とゲインを計算し，テーブルに格納しておくものとして，レンダリング時の演算量を見積もった．複数音源を，既に算出した時間シフト・ゲインを反映させて代表点に振り分けてから代表点の HRIR を畳み込むことで，各音源についてその方向の HRIR を各々直接畳み込むよりも畳み込みの回数が減るため，演算量の低減が見込める．受聴者から見て 2 つの代表点の間にある音源オブジェクト数をM，HRIR のタップ数をLとする．パンニングを行わず，それぞれの音源に対して直接畳み込みを行った場合，1 サンプル時間当たりの処理に必要な積和数は，

$$M \times L = ML$$

となる．提案法によるパンニングを行った場合は，時間シフトを反映した信号へのゲインの掛け算，代表点への足し込み，および 2 つの代表点での HRIR の畳み込みを行うこととなる．従って必要な積和数は，

$$(2 \times M) + 2 \times (M - 1) + 2 \times L = 4M + 2L - 2$$

となる．例として音源オブジェクト数が 3，HRIR のタップ数が 256 の場合を考える．パンニングを行わない場合の積和数は 768，パンニングを行った場合の積和は 522 であり，概ね 30%程度削減できている．音源オブジェクト数や HRIR のタップ数が増加するほど，パンニングを行うことでの演算量削減の効果が大きくなる．尚，演算量見積もり値は，処理に用いる CPU や DSP の構造によって多少増減する．

6.　考察

提案手法を用いることで，音質の劣化を抑えながら，バイノーラルレンダリングの演算量を抑えることができることを確認できた．実使用においては，予め代表方向を決めておき，音源方位毎のシフト量やゲイン等のパラメータを事前に算出して，テーブルに格納しておくことで，リアルタイムでの音環境レンダリングに活かすことが出来ると考えている．

謝辞

本研究は，JSP 科研費 JP19K12021，JP19K12066 の助成を受けた．

参考文献

[1] Benjamin Bernfeld. “Attempts for better understanding of the directional stereophonic listening mechanism”. 44th AES convention, 1973
[2] 小野一穂. マルチチャネルオーディオ. 映像情報メディア学会誌, 2014, vol. 68, No. 8, pp. 604-607.
[3] 安藤彰男. 音響の高臨場感技術. 映像情報メディア学会誌, 2012, vol. 66, No. 8, pp. 671-677.
[4] “The FABIAN head-related transfer function data base”. https://depositonce.tu-berlin.de/handle/11303/6153, (参照 2021-12-10).
[5] “JVS (Japanese versatile speech) corpus”. https://sites.google.com/site/shinnosuketakamichi/research-topics/jvs corpus, (参照 2021-12-16).