

PNP HRTFモデルのノッチ・ピークレベルの最適化 - 2種類のヘッドホンを用いた音像定位精度の検証 -*

◎中村風香 (千葉工大・院), 加古達也, 渡邊悠希, 千葉大将, 野口賢一 (NTT),
飯田一博 (千葉工大・先進工)

1. はじめに

3次元音響や音のVRを実現するために、頭部伝達関数(HRTF: Head-Related Transfer Function)の個人化方法の研究が活発に進められている[1-5]. われわれは, Parametric Notch-Peak HRTF モデル(PNP モデル)を用いて全天空の個人化 HRTF を生成するアルゴリズムを提案し, それを搭載したアプリケーションソフトを開発している[6,7].

PNP モデルでは, 正中面において実測 HRTF と同等の音像定位精度を提供する最小の HRTF の構成要素が 4 kHz 以上で最も周波数の低い 2 つのノッチ(N1,N2)と 2 つのピーク(P1,P2)である知見[8,9]を利用し, これら 4 つのノッチとピークそれぞれの中心周波数, レベル, 尖鋭度を音源方向と聴取者に応じて設定することにより HRTF を生成している.

ノッチとピーク(以降 N/P)の周波数については, 音源方向や聴取者の聴覚特性に適した値を算出するアルゴリズムを確立している[6,7]. 一方, N/P のレベルについては音源方向との相関は得られているが, 聴取者の特性と強い相関が得られる物理量は見出せていない[10]. そのため, 従来の PNP モデルでは N/P レベルに音源方向依存性は反映されているが, 個人差には対応していない.

本稿では, PNP モデルで生成した個人化 HRTF の音像定位精度の向上を目的として, 被験者が N/P レベルを調整する実験を実施して, N/P レベルの設定値を見直した. また, その有効性を確認するために, 2 種類のオープン型ヘッドホンを用いて音像定位実験を行った.

2. N/P レベルの最適化

N/P レベルは少なくとも N/P の存在が検知される値を超えていることが必要である. 従来

の研究でノッチの検知閾は-9 から-20 dB 程度であることが示されている[11]. この検知閾を参考にしつつ, 従来の PNP モデルでは上半球正中面 7 方向の 118 耳の実測 HRTF の N/P レベルを分析し, N1/N2 レベルは分布の第一四分位数, P1/P2 レベルは分布の第三四分位数に設定している[6,7].

しかし, この設定値はそれぞれの聴取者において最適であるということではない. そこで, 従来の PNP モデルの N/P レベルに重み係数を掛け, 重み係数を聴取者に調整させて最適値を探ることを考えた.

2.1 実験方法

まず, 被験者は PNP アプリを用いて, 適合する N2 周波数カテゴリと頭部寸法カテゴリを選定した[6].

次に, 被験者は上半球正中面 7 方向(上昇角: 0-180°, 30°間隔)の各方向について, アプリの GUI にあるスライダを用いて, N/P レベルに乗じる重み係数を 0.0 から 1.5 のレンジで調整した(Fig. 1 青い囲み部分). ここで, 重み係数が 1 より大きい場合は, N/P レベルの絶対値が従来モデルよりも増大し, 方向知覚の手がかりが強調される. 一方, 1 より小さい場合は N/P レベルの絶対値が低減し, 方向知覚の手がかりは薄れるが, HRTF による音質の変化が小さくなる.

調整においては被験者に以下のように教示した: “最も音像を目標方向に知覚しやすくなる(音像の幅が狭くなる, 距離感が大きくなる, 頭外感が増える, 目標方向に近づく)ようにスライダを調整してください”.

広帯域白色雑音(上限周波数 20kHz)に個人化 HRIR を畳み込んだ刺激(提示時間 1s)は被験者自身が play ボタンを押すことにより何度でも提示できる. 使用した機器は PC(Apple, MacBook Air), ヘッドホンアンプ(TEAC, HA-

* Optimization of notch and peak levels of PNP HRTF model - Verification of sound image localization accuracy using two kinds of headphones -, by NAKAMURA, Fuka (Chiba Institute of Technology), KAKO, Tatsuya, WATANABE, Yuki, CHIBA, Hironobu, NOGUUCHI, Kenichi (NTT), and IIDA, Kazuhiro (Chiba Institute of Technology).

P50SE), ヘッドホン(beyerdynamic, DT990 PRO)である. ヘッドホン伝達関数は補正していない. 被験者は 13 名(男性 6 名, 女性 7 名)である.



Fig. 1 GUI of the application software of PNP model for HRTF individualization. Subjects adjusted the weighting coefficient for N/P level using a slider in the regions outlined in blue [12].

2.2 実験結果

被験者が調整した重み係数の平均値と標準偏差を Table 1 に示す. 重み係数の平均値は全方向で 1 より大きく, N/P をより強調する方が目標方向に知覚しやすくなった. 方向別に見ると, 上半球正中面の前半分は後半分と比較して値が大きい. この傾向は従来の音像定位実験結果で示されている”前方は後方と比較して定位が難しい”という知見と一致する. 標準偏差は最大でも 0.21 であり, 個人差は比較的小さい.

重み係数の調整値の個人差が小さいことから, 本研究で提案する PNP モデルでは N/P レベルは個人化するのではなく, 聴取者に共通の値として取り扱うこととした. ただし, 従来のモデルと比較すると N/P レベルの絶対値は増大しており, これが音像定位精度に及ぼす影響を第 3 章で検証する.

Table 1 Weighting coefficient for N/P level for seven vertical angles in the upper median plane.

	Vertical angle [deg.]						
	0	30	60	90	120	150	180
Ave.	1.37	1.28	1.31	1.20	1.22	1.19	1.22
S.D.	0.17	0.21	0.13	0.21	0.17	0.16	0.20

3. 個人化 HRTF の音像定位精度の検証

2.2 節で得られた N/P レベルの重み係数を反映させた全天空個人化 HRTF 生成アプリを開発し, アプリで生成した個人化 HRTF の音像定位精度を検証した.

3.1 実験方法

Table 2 に示す 4 つの実験を実施した. 目標方向は水平面内の 12 方向(方位角: 0-330°, 30°間隔)または上半球正中面内の 7 方向(上昇角: 0-180°, 30°間隔)である. 音源信号は広帯域白色雑音(上限周波数 20kHz)で, 提示時間は 1s である. 音源信号に個人化 hrir を畳み込んだ刺激をヘッドホンにより被験者に提示した. ヘッドホンは DT990 PRO (beyerdynamic) と nwm ONE (NTT ソノリティ)の 2 種類を用いた (Fig.2). 両者とも PDR (Pressure Division Ratio)が 1 に近い値をとるため[13], FEC (free air equivalent coupling to the ear)ヘッドホン[14]とみなせる.

各刺激をランダムな順に 10 回ずつ提示した. 被験者は成人 5 名(男性 2 名, 女性 3 名)である. 被験者はマッピング法により実験 1,2 では音像の方位角と仰角を回答し, 実験 3,4 では音像の上昇角を回答した. 頭内定位した場合は頭内定位ボックスにチェックを入れた. 実験は防音室で実施した.

Table 2 Target directions and headphones for four kinds of sound image localization tests.

	Target directions	Headphones
Test 1	12 directions in the horizontal plane	DT990 PRO
Test 2	7 directions in the median plane	nwm ONE
Test 3	7 directions in the horizontal plane	DT990 PRO
Test 4	7 directions in the median plane	nwm ONE



Fig. 2 Headphones used in the localization tests. (a): DT990 PRO, (b):nwm ONE.

3.2 実験結果

3.2.1 N/P レベルの最適化が音像定位精度に及ぼす影響

まず, N/P レベル最適化が音像定位精度に及ぼす影響を検証した. 5 人の被験者のうち 3 人は N/P レベルの最適化をしていない従来の PNP モデルで生成した音像定位実験も実施している[6]. Figs.3,4 に 3 人の被験者による従来モデルと提案モデル(実験 1,3 の 3 名分)の音像定位実験の結果を示す. ただし, 従来モデルの音像定位実験では水平面, 正中面ともに目標方向は 45°間隔であったことに注意されたい.

Fig.3 に目標方向が水平面の場合の方位角の回答を示す。水平面では、従来モデルでも概ね目標方向付近に回答したが、提案モデルでも同様の結果であり、両者の布置に顕著な差はみられなかった。

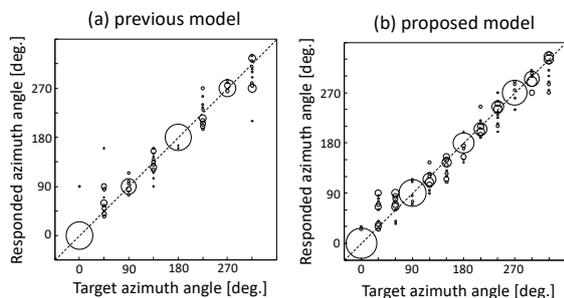


Fig. 3 Responded azimuth angle for individualized HRTFs in the horizontal plane. (a) previous model, (b) proposed model.

Table 3 に従来モデルと提案モデルで目標方向が共通する 0, 90, 180, 270°での平均方位角誤差を示す。4 方向の平均値は従来モデルでは 3.4°, 提案モデルでは 4.6°であり、顕著な差はみられなかった。

Table 3 Mean azimuth angle error [deg.]

Target azimuth angle	Previous model	Proposed model
0	3.7	3.0
90	5.1	3.7
180	1.5	7.3
270	3.1	4.5
Ave.	3.4	4.6

Fig.4 に目標方向が正中面の場合の上昇角の回答を示す。0°(正面)では、従来モデルでは音像が上昇する傾向がみられたが、提案モデルでは概ね目標方向に回答した。90, 180°では従来モデル、提案モデルともに概ね目標方向に回答したが、その間の方向では両者ともに回答にばらつきがみられた。

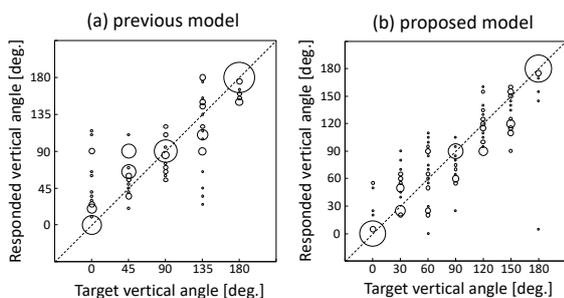


Fig. 4 Responded vertical angle for individualized HRTFs in the horizontal plane. (a) previous model, (b) proposed model.

Table 4 に上昇角 0, 90, 180°の平均上昇角誤差を示す。3 方向の平均値は従来モデルでは 16.3°, 提案モデルでは 11.0°であり、N/P レベルの最適化により定位精度に改善がみられた。特に 0°で顕著な改善がみられたが、一方で 90, 180°では従来モデルよりも若干大きな値となった。

Table 4 Mean vertical angle error [deg.]

Target vertical angle	Previous model	Proposed model
0	30.5	8.2
90	11.1	15.2
180	8.5	9.5
Ave.	16.7	11.0

3.2.2 ヘッドホンの機種が音像定位精度に及ぼす影響

次に、提案モデルと 2 種類のオープン型ヘッドホンを用いてヘッドホンの機種の違いが音像定位精度に及ぼす影響を検証した。

Fig.5 に目標方向が水平面の場合(実験 1,2)の 5 人の被験者の方位角の回答を示す。

DT990 PRO, nwm ONE とともに 0, 90, 180, 270°では目標方向付近に回答した。30, 150, 330°ではやや大きなばらつきが両者ともにみられた。総じて両者の布置に顕著な差はみられなかった。

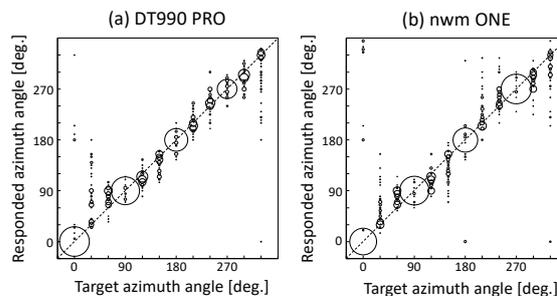


Fig. 5 Responded azimuth angle for individualized HRTFs in the horizontal plane. (a) DT990 PRO and (b) nwm ONE.

平均方位角誤差を Table 5 に示す。12 方向の平均値は DT990 PRO で 18.5°, nwm ONE で 18.4°でありほぼ同等であった。ただし、0, 30, 330°の前方では nwm ONE の定位誤差が小さく、150, 180, 210°の後方においては DT990 PRO の定位誤差が小さい傾向がみられた。

Table 5 Mean azimuth angle error [deg.].

Target azimuth angle	DT990 PRO	nwm ONE
0	26.2	21.0
30	52.6	26.4
60	22.0	19.6
90	3.8	4.3
120	11.9	15.7
150	14.2	21.3
180	6.4	19.2
210	10.9	20.9
240	15.3	20.2
270	7.4	5.0
300	16.2	17.1
330	34.4	29.7
Ave.	18.5	18.4

Fig.6 に目標方向が正中面の場合(実験 3,4)の 5 人の被験者の上昇角の回答を示す. DT990 PRO, nwm ONE とともに 90, 180°では概ね目標方向付近に回答し, 60, 120, 150°では回答のばらつきが大きい. しかし, 両者の布置には顕著な差はみられなかった.

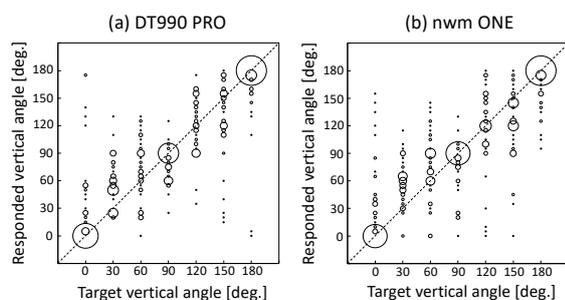


Fig. 6 Responded vertical angle for individualized HRTFs in the median plane. (a) DT990 PRO and (b) nwm ONE.

平均上昇角誤差を Table 6 に示す. 7 方向の平均値は DT990 PRO で 26.4°, nwm ONE で 27.3°であり, 水平面の方位角誤差と同様, 両者に顕著な差はみられなかった.

Table 6 Mean vertical angle error [deg.].

Target vertical angle	DT990 PRO	nwm ONE
0	34.3	32.2
30	32.0	31.3
60	28.3	30.8
90	14.3	16.9
120	25.4	32.1
150	32.3	33.5
180	18.1	14.3
Ave.	26.4	27.3

最後に水平面および正中面での頭内定位率を Table7 に示す. いずれの場合も低い値であったが, 水平面では nwm ONE がより低く, 正中面では DT990 PRO がより低い結果であった.

Table 7 Ratio of inside-of-head localization.

Horizontal plane		Median plane	
DT990 PRO	nwm ONE	DT990 PRO	nwm ONE
0.04	0.02	0.03	0.07

5. おわりに

本稿では, PNP モデルで生成した個人化 HRTF の音像定位精度の向上を目的として, 被験者が N/P レベルを調整する実験を実施し, N/P レベルの設定値を見直した. また, その有効性を確認するために, FEC ヘッドホンとみなせる 2 種類のオープン型ヘッドホンを用いて音像定位実験を行い, 以下のことを示した.

- (1) レベルの最適化をしない従来モデルと提案モデルの定位精度を比較すると, 水平面では両者ともに概ね目標方向付近に定位した. 一方, 正中面では正面方向において, 従来モデルでは音像が上昇する傾向がみられたが, 提案モデルでは概ね目標方向に定位し, 改善がみられた.
- (2) 2 種類の FEC ヘッドホン(DT990 PRO, nwm ONE)の音像定位精度に顕著な差はみられなかった. 平均定位誤差は水平面では 18°程度, 正中面では 27°程度であった.

謝辞

本研究の一部は JSPS 科研費 24K15052 の助成を受けて実施した.

参考文献

- [1] Kistler and Wightman, J. Acoust. Soc. Am. 91,1637-47, 1992.
- [2] Middlebrooks and Green, J. Acoust. Soc. Am. 92, 597-599, 1992.
- [3] Iida *et al.*, J. Acoust. Soc. Am. 136, 317-333, 2014.
- [4] Bomhardt *et al.*, Proc. of meetings on acoustics, 2016.
- [5] Miccini and Spagnol, IEEE 5th VR Workshop on Sonic Interactions in Virtual Environments, Mar. 2020.
- [6] 中村と飯田, 聴覚研究会資料, 54, No.1, H-2024-20 もしくは信学技報, HIP2024-97.
- [7] 飯田, 音講論(春), 3-6-3, 2024.
- [8] Iida *et al.*, Appl Acoust, 68, 835-850, 2007.
- [9] Iida and Ishii, Appl Acoust, 129, 239-247, 2018.
- [10] 設楽と飯田, 音講論(秋), 3-P-32, 2024.
- [11] Moore *et al.*, J. Acoust. Soc. Am., 85, 820-836, 1989.
- [12] <https://www.iida-lab.it-chiba.ac.jp/webapps/home/>
- [13] 千葉他, 音講論(秋), 1-R-47, 2024.
- [14] Møller *et al.*, J. Audio Eng. Soc. 43, 203-217, 1995.