

音の方向知覚の手掛かりに基づいた個人化頭部伝達関数の生成*

- 高精度 3 次元音響再生の実用化に向けて -

○飯田一博（千葉工大・先進工）

1. はじめに

ヘッドホンを用いて 3 次元音響定位を実現するには聴取者に適合する HRTF を提供する必要がある。このようなプロセスを HRTF の個人化と呼んでいる。

音響定位精度の高い個人化 HRTF を生成できれば、現実と遜色ない音空間を再現できる。高精度な 3 次元音響再生はエンタテインメントに限らず、会議、教育、訓練、ナビゲーション、現地調査、医療、高齢者・障害者支援など生活の多岐に渡る局面において、対面と同等のコミュニケーションツールとなり得る。また、このようなツールが社会に普及すると人の移動が減少し、地球温暖化の緩和にも貢献できる。

HRTF 個人化の研究は国際的に活発に進められており、概ね以下の 3 種類に分類できる。

- (1) HRTF の振幅スペクトルの個人化[1]
- (2) hrir の時間波形の個人化[2]
- (3) 方向知覚の手掛かりの個人化[3]

従来の研究の多くは(1)もしくは(2)に着目して進められてきた。これらは物理的に本人と同等の HRTF/hrir の生成をめざしている。さまざまな方法が提案されているが、現時点では生成した個人化 HRTF/hrir の定位精度を音響定位実験により確認した結果は見あたらぬ。

一方、(3)では聴取者の方向知覚の手掛かり、すなわちスペクトラルキューと両耳間差キューを再現する。この場合、物理的に本人と全く同等の HRTF を生成する必要はなく、音響方向の知覚に必要な特性に重点をおいて HRTF を生成する。われわれは、これまで主に(3)の方針で個人化 HRTF の生成に取り組んできた。

本稿では、ヒトの方向知覚の手掛かりに着目した Parametric Notch-Peak HRTF モデル (PNP モデル)を紹介する。これは、音響の上昇角知覚の手掛かりであるスペクトラルノッチ・ピークと音響の側方角知覚の手掛かりである両耳間時間差・レベル差で HRTF を構成したモデルである。

さらに、PNP モデルを用いて個人化 HRTF を生成するアプリケーションソフトを紹介する。このアプリは以下の特徴を有する。

- (1) 1 方向の個人化 HRTF から全天空のあらゆる方向の個人化 HRTF が生成できる。
 - (2) 1 方向の HRTF の個人化においては、ただ 1 つのパラメータを最適化するだけでよい。
- 最後に、高精度 3 次元音響再生の実用化に向けて、音質面での課題について私見を述べる。

2. 上昇角知覚の手掛かりに基づいた HRTF のモデル化

3 次元音響再生のポイントは前後上下方向の音響定位である。したがって、音響の上昇角知覚の手掛かりを再現することが重要となる。

2.1 上昇角知覚の手掛かりと pHRTF

上昇角知覚の手掛かりは HRTF の振幅スペクトルに存在することが知られており、スペクトラルキューと呼ばれている。

われわれは実測 HRTF の振幅スペクトルを複数のノッチとピーク(以降 N/P と記述)に分解し、その一部または全部を用いて再合成したパラメトリック HRTF (pHRTF) を提案している[4]。さまざまな N/P の組み合わせで再合成した pHRTF を用いた音響定位実験により、上半球正中面のいずれの方向においても実測 HRTF と同等の音響定位精度を提供できる最小の pHRTF の構成要素は、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)であることを示した(Fig.1)[4,5]。

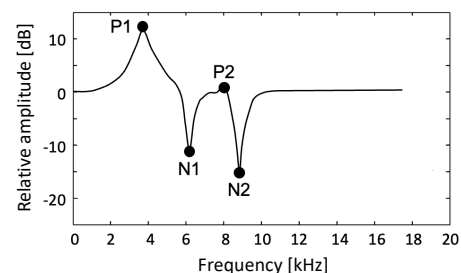


Fig. 1 Example of parametric HRTF recomposed of N1, N2, P1, and P2.

* Generation of the individualized HRTFs based on the cues for perception of direction of a sound image - Toward practical realization of accurate 3D sound reproduction -, by IIDA, Kazuhiro (Chiba Institute of Technology).

2.2 Parametric Notch-Peak HRTF モデル

4つのN/PはIIRフィルタを用いることにより、それぞれ中心周波数、レベル、尖鋭度(Q)の3つのパラメータで決定される。したがって、最小構成のpHRTFは12個のN/Pパラメータ(3パラメータ×(2ノッチ+2ピーク))で表され、これらを聴取者に適合させれば個人化HRTFを生成できる。しかし、N/Pパラメータは音源方向に依存するだけでなく、聴取者にも依存するため、その最適化は容易ではない。

ここで、ある方向で個人に適合したHRTFを生成するために、12種類のN/Pパラメータのうち1つのパラメータ(独立変数)を決定すれば、他の11種類のパラメータが従属的に決定されるHRTFモデルを考える。これをParametric Notch-Peak HRTFモデル(PNPモデル)と呼ぶ。妥当なPNPモデルを構築できれば、ただ1つのN/Pパラメータを適合させるだけで、その方向の個人化HRTFが生成できる。

2.3 正中面7方向のPNPモデルの構築

2.3.1 N2周波数によるN1/P1/P2周波数の推定

無響室で測定した成人男女59名(118耳)の上半球正中面7方向(上昇角:0°-180°, 30°間隔)のHRTFからN/Pパラメータを抽出した。N/P周波数の分布をFig.2に示す。N1/N2周波数(a)(b)は方向依存性および個人依存性が顕著であることが観察されるが、P1周波数(c)は方向依存性、個人依存性ともに低い。P2周波数(d)は方向依存性は低いが、個人依存性は高い。

これらのN/P周波数パラメータを用いて、正中面7方向のそれぞれにおいてN1/N2/P1/P2のいずれかの周波数を独立変数、残りの3種類の周波数を従属変数とする単回帰分析を行った。その結果、N2周波数を独立変数とした場合が最も強い相関が得られた。

N2周波数を独立変数としてN1/P1/P2周波数を回帰式により推定した際の平均残差を求めた。7方向を通じて大きな差異はなく、0.05-0.11 oct.の範囲にあった。方向知覚に関するN/P周波数の弁別閾は、N1/N2では0.1-0.2 oct., P1では低域側が0.35 oct., 高域側が0.47 oct.であることが報告されている[6]。118耳のうち残差が弁別閾内に収まる耳のパーセンテージを求めたところ、N1周波数では79-98%, P1では100%, P2では96-100%であった。

これらより、上半球正中面7方向での個人のN1/P1/P2周波数は、ほぼ弁別閾内の精度でN2周波数より推定できると言える。

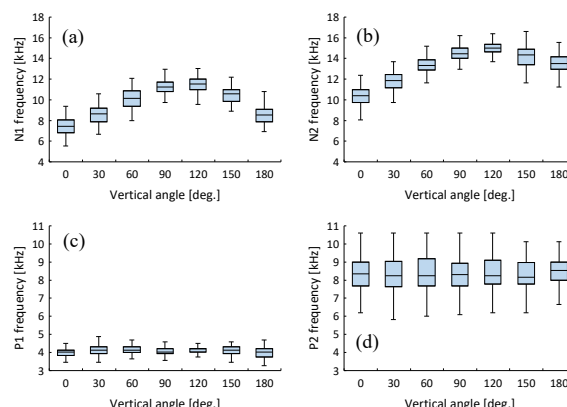


Fig. 2 Distribution of frequency of (a) N1, (b) N2, (c) P1, and (d) P2 for seven directions in the upper median plane for 118 ears.

2.3.2 N1/N2/P1/P2のレベルとQの設定

次に上半球正中面7方向の118耳のN/PレベルとQの分布を求めた。紙面の都合で詳細は割愛するが、N1/N2レベルは120°, 150°で小さく(深く)、個人依存性も観察された。P1レベルは前方で大きい(高い)傾向があるが、個人依存性は少ない。P2レベルは上方で大きい傾向があることが観察された。

N/Pレベルは他のパラメータとの間に強い相関関係は見られなかった。しかし、N/P周波数では絶対値が重要であることと異なり、N/PレベルはN/Pの存在が検知される値を超えていることが重要であると考えられる。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[7]。

これらの知見に基づいて、本研究ではN/PのレベルとQについては、方向依存性は与えるが聴取者間では共通の値とした。聴取者がN/Pを検知できることを意図して、N1/N2レベルは上半球正中面7方向のそれぞれにおいて118耳の第一四分位数、P1/P2レベルは第三四分位数に設定した。Qは第二四分位数(中央値)に設定した。

2.3.3 P0の周波数・レベル・Q

最近の研究により、後方のHRTFにおいては1kHz付近にピーク(P0)が存在し、これが後方の音像定位に寄与することが示された[8]。そこで正中面上昇角120°, 150°, 180°のPNPモデルにはP0を付加した。ただし、P0は個人差

が非常に小さいため、聴取者によらず周波数は 1031.25 Hz, レベルは 2.0 dB(120°), 3.0 dB(150°), 5.0 dB(180°), Q は 1.0 とした。

3. 全天空の個人化 HRTF の生成

以上より PNP モデルを構築した方向においては, N2 周波数から pHRTF を生成することが可能となった。しかし, 全天空の個人化 HRTF を生成するには, あらゆる方向の PNP モデルが必要となる。

そこで, ある 1 つの方向において聴取者が N2 周波数を設定すれば, 全天空の個人化 HRTF が生成される仕組みを構築した。設定する方向は天頂方向とした。その理由は, 天頂方向では N2 周波数の違いによる音像方向の違いがわかりやすく, 一般の聴取者が自身で適した N2 周波数を設定しやすいと考えられるからである。

天頂の N2 周波数の設定から全天空の個人化 HRTF を生成する手順を以下に示す(Fig.3)。

- (1) 天頂方向に音像を知覚する N2 周波数カテゴリを聴取者が選択(Fig.3 の赤い円)。
 - (2) 天頂方向の N2 周波数から上半球正中面の他の 6 方向(30°間隔)の N2 周波数を回帰式で推定(黒の曲線の矢印)。
 - (3) 上半球正中面 7 方向の PNP モデルを用いて, N2 周波数から他の N/P パラメータを推定(黒の点線の矢印)。
 - (4) 直線回帰で上半球正中面の任意の方向の N/P パラメータを推定(青とオレンジの破線)。
 - (5) 両耳間差情報を付加することにより, 全天空の任意の方向の個人化 HRTF を生成[9]。
- 手順(1)–(5) の詳細については[10]を参照されたい。

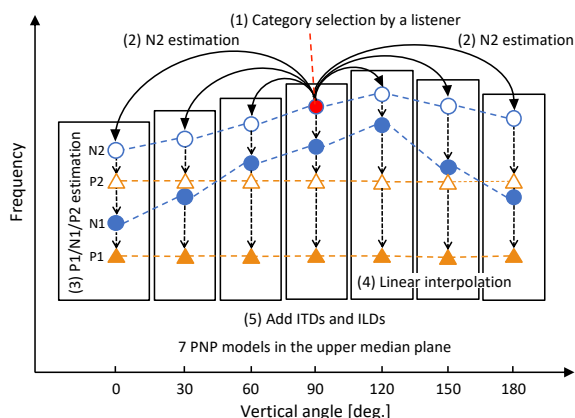


Fig. 3 Schematic diagram for generating individualized HRTFs in the upper hemisphere using seven PNP HRTF models in the upper median plane.

4. PNP モデルを用いた個人化 HRTF 生成アプリの開発

PNP モデルを用いて個人化 HRTF を生成するデスクトップアプリを MATLAB® の App Designer 用いて開発した(Fig.4)[10,11]。

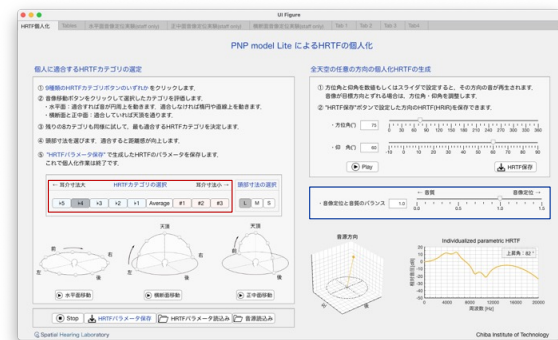


Fig. 4 GUI of the application software for HRTF individualization using Parametric Notch-Peak HRTF model.

4.1 天頂方向の N2 周波数の設定方法

聴取者とアプリの唯一の接点となる”天頂方向に音像定位する N2 周波数”の設定方法として以下の 3 種類が考えられる。

- (1) 聴取者による N2 周波数の調整
- (2) 聴取者による N2 周波数カテゴリの選択
- (3) 耳介形状画像による N2 周波数の推定

これまで, 簡便性と定位精度の両立を求めて様々な方法を考案した。学外も含めて多くの方にアプリを試用していただいた結果, 天頂方向の N2 周波数設定の簡便性と音像定位精度はトレードオフの関係にあった。例えば, スライダーなどでフィットする N2 周波数を連続的に調整すれば音像定位精度は向上する。しかし, その操作は一般の聴取者にとっては困難であった。

われわれは” (2) 聴取者自身による N2 周波数カテゴリの選択”が操作の簡便性と定位精度の両面で妥当であると考えている。本研究では, 天頂方向の 118 耳の N2 周波数の分布範囲を, 弁別閾以下の 1/12 oct.(0083 oct.)間隔で 9 種類のカテゴリに分類した(Fig.4 の赤枠部分)。

4.2 生成した個人化 HRTF の音像定位精度

8 人の被験者がアプリで生成した水平面(0-315°, 45°間隔)および正中面(0-180°, 45°間隔)の個人化 HRTF に白色雑音を畳み込み, ヘッドホンから提示した。全被験者の回答を Fig.5 (a), (b)に示す。比較のため, 8 名中 7 名の被験者の水平面および正中面(それぞれ 0-180°, 30°間隔)の実音源に対する回答を(c), (d)に示す。

水平面の個人化 HRTF の回答分布は 0°, 90°,

180°では実音源と同様であったが、45°, 135°, 225°, 315°ではばらつきがやや大きい。正中面の回答分布は上方および後方では実音源とほぼ同様であったが、正面では音像が上昇する場合があった。音像定位精度とアプリの操作性評価の詳細は別稿[11]で報告する。

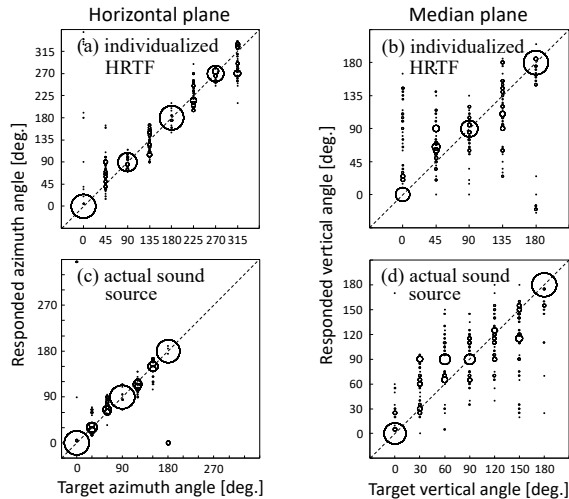


Fig. 5 Results of localization tests in the horizontal and the median planes. (a) and (b): individualized HRTFs. (c) and (d): actual sound sources.

5. おわりに：3次元音響再生に求められる音質について

個人化 HRTF を実用に供するための条件として、音像定位精度と HRTF 生成操作の簡便性に加えて音質(音色)があげられる。HRTF を用いると音質が変化するとよく耳にする。音質の基準はどこに置くべきだろうか？

ヒトは日常生活において(ヘッドホン受聴時を除き)自分自身の HRTF を含んだ音を聴いている。私は3次元音像定位の実現が最優先事項であるので、ひとまず音源信号に本人 hrir が畳み込まれた音質を基準としている。

一方、通常のヘッドホン受聴の音質を基準とする考えもある。これを支持する方はコンテンツクリエイターには多いのではないだろうか。彼らの責務を鑑みるとこの考えも頷ける。この場合は、本人 HRTF も音質の阻害要因となる。

求められる音像定位精度と音質はコンテンツやサービスに依存する。Fig.6に示すように、音像定位よりも音質変化が少ないことが重要なコンテンツもあれば、音質変化の許容範囲は広いが音像定位は必須のサービスもある。

本稿で紹介した個人化 HRTF 生成アプリは、その存在理由ゆえ、音像定位と音質のバランス

を音像定位寄りに設定しているが(2.3.2 項参照)、このバランスを GUI 上で簡単に変更・調整できるパラメータを設けている(Fig.4 の青枠部分)。これにより、音像定位を控えめにして音質変化が少ない個人化 HRTF も生成できる。3次元音響再生を様々な用途に展開するうえで活用できる機能ではないかと考えている。

今後われわれは音像定位精度と音質の両立や個人化プロセスのさらなる簡易化を進めるが、高精度3次元音響再生は、研究者とサービスプロバイダやコンテンツクリエイターのみなさんが協働することにより、より早く実用化に到達できるのではないだろうか。お力添えいただければ幸いである。

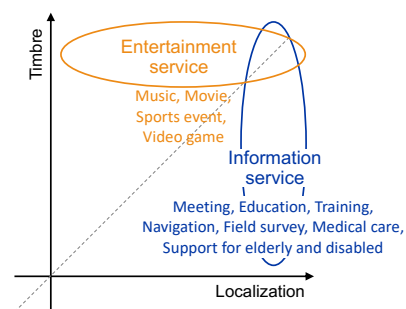


Fig. 6 Required sound image localization accuracy and timbre for various 3D audio services.

謝辞

本研究の一部は JSPS 科研費 19K12068 の助成を受けて実施した。学内外の多数の方に被験者およびアプリの評価者としてご協力いただいた。修了生の相崎翼氏、菊地勇成氏、現大学院生の中村風香氏には共同研究者として多大な貢献をしていただいた。記して感謝します。

参考文献

- [1] 例えば Bomhardt *et al.*, Proc. of meetings on acoustics, 2016.
- [2] 例えば Chun *et al.*, AES 143th Convention, New York, USA. 2017 October 18-21.
- [3] 例えば Iida *et al.*, J. Acoust. Soc. Am., 136, pp.317-333, 2014.
- [4] Iida *et al.*, Appl. Acoust. 68, 835-850, 2007.
- [5] Iida *et al.*, Appl. Acoust. 129, 239-247, 2018.
- [6] 飯田, 頭部伝達関数の基礎と3次元音響再生への応用, 66-67, コロナ社, 東京, 2017.
- [7] Moore *et al.*, J. Acoust. Soc. Am., 85, pp.820-836, 1989.
- [8] 中村と飯田, 聴覚研究会資料, H-2023-46, 2023.
- [9] Morimoto *et al.*, Acoust. Sci. & Tech. 24, 267-275, 2003.
- [10] 中村と飯田, 聴覚研究会資料, 2024.2
- [11] 中村と飯田, 音講論(春), 3-6-4, 2024.