

受聴者の耳介形状による上半球正中面の 初期頭部伝達関数のノッチ周波数の推定

○島崎光, 飯田一博 (千葉工大・工)

1 はじめに

精度の高い音像定位には, 受聴者本人もしくは受聴者に適合する頭部伝達関数(HRTF: head-related transfer function)を用いる必要があるが, 一般の受聴者の HRTF を測定することは困難であり, 耳介形状などから HRTF を推定することが試みられている[1].

著者の 1 人は, 上半球正中面の音像定位における必要最小限の HRTF の構成要素は 2 つのノッチと 2 つのピーク(N1,N2,P1,P2)であることを報告した[2,3]. さらに, この知見を利用して, 各受聴者の N1,N2,P1,P2 を耳介形状から推定し, これら 4 つのファクタで構成した個人化 HRTF をパラメトリックに生成することを目指している. この個人化 HRTF は 4 つのピーキングフィルタで構成できるため, HRTF データベースは不要となる.

しかし, ノッチやピークの推定は部分的な成功に留まっている. 正面方向に限れば, 耳介形状から N1,N2 周波数を弁別閾内の誤差で推定可能であるが[4], 正面以外の方向, 特に上方においては, 耳介形状と N1,N2 周波数の間に強い相関関係は見出せておらず, その推定方法の確立は解決すべき課題として残されている.

ここで, ノッチやピークの成因を考えを巡らせると, これらは耳介で生成されることが報告されている[5]. また, 耳介の影響は HRTF の観測点(外耳道入口)との位置関係から, 主に頭部インパルス応答(HRIR: head-related impulse response)の初期部分に含まれていると考えられる. したがって, HRIR の初期部分からノッチを抽出することにより, 耳介形状と強い相関が得られる可能性がある.

そこで, HRIR の継続長をパラメータとして音像定位実験を行ったところ, 上半球正中

面において上昇角に関わらず従来の HRIR と同等の音像定位精度が得られる必要最小限の継続長は 1ms 程度であった[6]. この結果より, 初期 1ms の HRIR は音像定位精度を保ちつつ, 耳介形状との相関の高さも兼ね備えることが期待される.

本研究では, 継続長 1ms の初期頭部伝達関数に着目し, 上半球正中面の N1 周波数を受聴者の耳介形状から推定することを試みた.

2 初期 HRTF の N1 の抽出

日本人成人 27 名の 54 耳の上半球正中面(0°-180°, 30°間隔)の初期 1ms の HRTF を求め, N1 周波数を抽出した. 抽出アルゴリズムは以下の通りである.

- 1) HRIR の振幅の絶対値が最大となるサンプル(Smax)を検出し, Smax を中心として 4 項, 96(時間軸上で前に 48, 後ろに 48)サンプルの Blackman-Harris 窓で初期 HRIR を切り出す. サンプリング周波数は 48 kHz である.
- 2) 512 サンプルの FFT により初期 HRTF の振幅スペクトルを算出する. 周波数分解能は 93.75 Hz である.
- 3) 差分法により極小値を求め, 4 kHz 以上の帯域で最も周波数の低い極小値を N1 とする.

初期 1ms の HRTF のスペクトルの例を Fig. 1 に示す. 初期 HRTF のスペクトル(破線)の概形は全区間 HRTF(実線)と同様であるが, 微細な山谷は消滅している.

Fig. 2 に初期 HRTF の N1 周波数とレベルの分布を方向ごとに示す. 全区間 HRTF と同様, N1 周波数は音源方向が正面から上昇するにしたがって高くなり, 120°付近で最も高くなる. また, N1 周波数の分布範囲は隣接する方向間で重なるだけでなく, 0°と 180°, 30°と

Estimation of spectral notch frequency of early head-related transfer function in the upper median plane based on the anthropometry of the listener's pinnae, by SHIMAZAKI, Hikaru and IIDA, Kazuhiro (Chiba Institute of Technology)

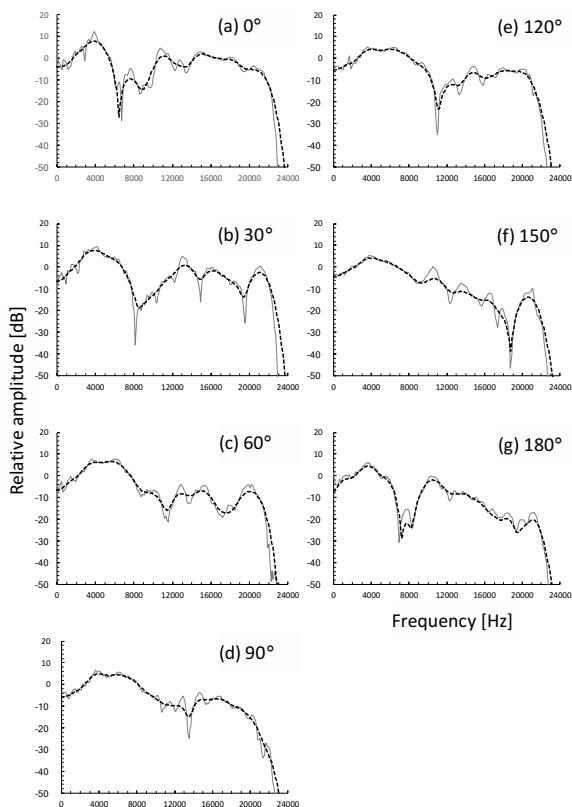


Fig. 1 Examples of the amplitude spectra of the full-length HRTFs (solid lines) and the early HRTFs (broken lines) for seven vertical angles.

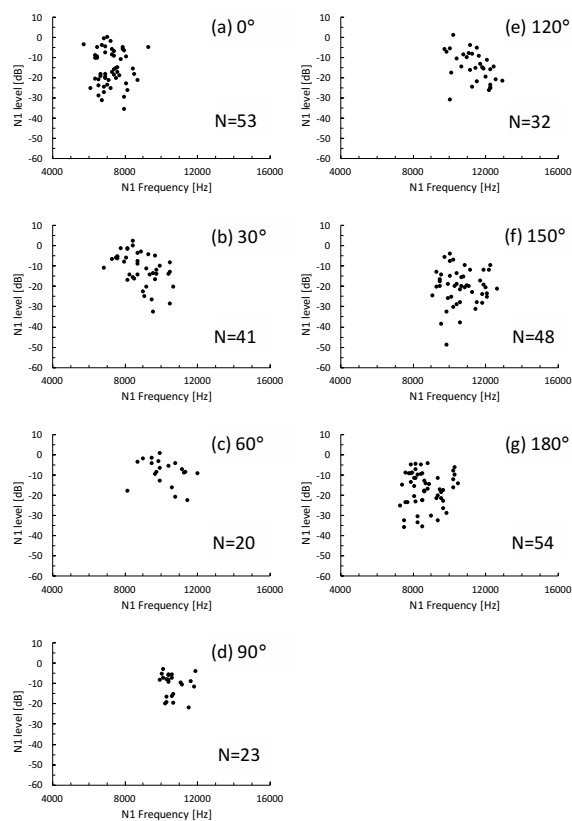


Fig. 2 Distribution of the frequency and level of N1 for seven vertical angles.

150°, 60°と 120°でも重なっており，他人の HRTF を用いた場合の前後誤判定と符合している．ノッチのレベルについては，前方と後方では深く，上方では浅い．

Table 1 に示すように，N1 周波数の個人差は約 2,000 Hz から 4,000 Hz に及び，前方と後方では大きく，上方では小さい傾向がある．オクターブ数で表すと 0.26 から 0.70 [oct.]であり，音像の上昇角に関する弁別閾(0.1-0.2 [oct.][4])と比較して十分大きい．

ところで，前方と後方ではほぼ全ての耳で N1 を抽出できたが，上方になるにつれ N1 が現れない耳の割合が増加した．特に，60°および 90°で N1 が抽出できたのは 20 および 23 耳であった．全区間 HRTF においても，上方では微細な山谷しか現れない耳があるが，初期 HRTF ではスペクトルがより滑らかになるため極小点が生じなかったものと考えられる．一方，このような初期 HRTF でも全区間 HRTF と同等の音像定位精度を有することが示されており[6]，上方においては N1 が浅い，もしくは生じないことが重要な情報となっている可能性がある．今後，ノッチレベルの推定と併せて検討すべき課題である．

Table 1 Range of N1 frequencies.

	vertical angle [°]						
	0	30	60	90	120	150	180
Min. [Hz]	5719	6844	8156	9938	9750	9094	7313
Max. [Hz]	9281	10688	12000	11906	12938	12656	10500
Range [Hz]	3562	3844	3844	1968	3188	3562	3187
Range [oct.]	0.70	0.64	0.56	0.26	0.41	0.48	0.52

3 耳介形状パラメータ

N1 抽出に用いた 54 耳について，Fig. 3 に示す 10 箇所の耳介形状パラメータ[4]を計測した．このうち， x_1 - x_8 および x_d の 9 箇所は耳型よりノギスで計測し， x_a は被験者の横顔の写真から計測した．計測結果を Table 2 に示す． x_1 - x_8 および x_d の個人差は 6.6-25 mm， x_a の個人差は 36°であった．

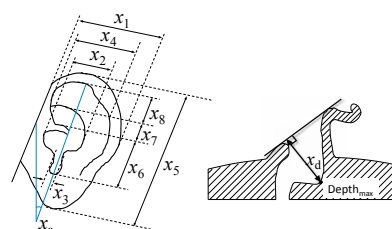


Fig. 3 Ten anthropometric parameters of the pinna.

Table 2 Range of pinnae dimensions for 54 ears [mm].

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_d	x_a
Min.	31.2	14.8	5.3	18.1	58.2	17.1	2.6	13.2	9.7	4
Max.	43.8	21.8	11.9	31.8	83.2	25.1	10.3	24.1	17.6	40
Range	12.6	7.0	6.6	13.7	25.0	8.0	7.7	10.9	7.9	36

4 耳介形状による N1 周波数の推定

54 耳の 10 種類の耳介形状パラメータを説明変数, 各上昇角の N1 周波数を目的変数として重回帰分析を行った(式(1)).

$$f(\beta)_{N1} = a_1(\beta)x_1 + a_2(\beta)x_2 + \dots + a_n(\beta)x_n + b(\beta) \text{ [Hz]} \quad (1)$$

ここで, β , a_i , b , x_i はそれぞれ上昇角(0° - 180° , 30° 間隔), 回帰係数, 定数, 耳介形状パラメータを表す.

それぞれの上昇角において, 変数減増法を用いて耳介形状パラメータを選択した.

$F_{in} = F_{out} \geq 2.0$ を満たす耳介形状パラメータの組合せのうち, 重相関係数が最大となるものを採用した. いずれの場合も, 多重共線性は見られなかった.

各上昇角の回帰係数 a_i を Table 3 に示す. 前方及び後方(0° , 30° , 180°)では, 回帰式に共通して x_3 (珠間切痕の幅), x_8 (舟状窩の長さ), x_d (耳甲介腔の深さ)が含まれていた. 一方, 上方の 90° , 120° では x_8 は含まれず, x_2 (耳甲介腔の幅)と x_3 が共通して含まれていた.

Table 3 Regression coefficients for seven vertical angles.

	Vertical angle, β [$^\circ$]						
	0	30	60	90	120	150	180
a_1							
a_2	186.95			-161.72	409.10		
a_3	-230.25			119.91	-219.35		-143.99
a_4		-91.33			-215.10	-126.56	-148.87
a_5					91.48	80.12	
a_6	-92.44	-199.83	-124.25	94.91		-272.68	
a_7						-160.23	-86.73
a_8	-117.41	-129.35	-271.68			-194.85	-85.65
a_d	-102.86	-123.18			-223.62		-131.65
a_a	-33.44	-45.50	-59.62				29.22
b	11969.78	20625.95	19371.93	10690.74	7875.17	18998.55	17151.83

重回帰モデルの統計量を Table 4 に, 推定した N1 周波数と初期 HRTF から抽出した N1 周波数との関係を Fig. 4 に示す. 7 方向の重相関係数は 0.56 - 0.90 であり, 概ね良好な結果が得られたが, 90° , 150° , 180° での重相関係数は他の方向と比較すると低い. 平均絶対残差

(MAR: mean absolute residual)は 7 方向で 303.9 - 613.4 Hz(0.04 - 0.09 oct.)であった. 絶対残差が弁別閾内となる耳の割合は 83 - 100% であった. ここで, 弁別閾は 0.15 oct.とした.

Table 4 Statistics of the multiple regression models.

	vertical angle [$^\circ$]						
	0	30	60	90	120	150	180
r	0.76	0.72	0.90	0.56	0.79	0.65	0.63
Mean Absolute Residual [Hz]	369.3	512.4	303.9	374.9	407.1	613.4	539.5
MAR [oct.]	0.07	0.08	0.04	0.05	0.05	0.08	0.09
MAR < jnd [%]	89	80	95	100	97	92	83

5 おわりに

個人に適合する HRTF を提供することを目的として, 受聴者の耳介形状から継続長 1ms の初期頭部伝達関数の N1 周波数の推定を試みた. その結果, 以下の知見が得られた.

(1) 日本人成人 54 耳の上半球正中面 7 方向(0° - 180° , 30° 間隔)の初期 HRTF から抽出した N1 周波数の個人差は約 2000 Hz から 4000 Hz, 0.26 - 0.70 [oct.]であった.

(2) 各方向の N1 周波数を目的変数, 耳介形状を説明変数として重回帰分析を行った. 重相関係数は 0.56 - 0.90 であり, 絶対残差が弁別閾内となる確率は 83 - 100% であった.

以上より, 耳介形状から上半球正中面 7 方向の初期 HRTF の N1 周波数を概ね弁別閾内の誤差で推定できる可能性が示唆された.

今後, 本重回帰モデルをナイーブな被験者により検証する必要がある. さらに, N1 のレベルおよびバンド幅の推定も進めたい.

謝辞

本研究の一部は科研費(基盤研究(A) 15H01790)により実施した.

参考文献

- [1] 飯田, 頭部伝達関数の基礎と 3 次元音響再生への応用, 82-100, コロナ社, 2017.
- [2] Iida *et al.*, Appl. Acoust., 68, 835-850, 2007.
- [3] Iida and Ishii, Appl. Acoust., 129, 239-247, 2018.
- [4] Iida *et al.*, JASA, 136, 317-333, 2014.
- [5] Takemoto *et al.*, JASA, 132, 3832-3841, 2012.
- [6] 飯田と太田, 建築音響研究会資料, AA2017-53, 2017.

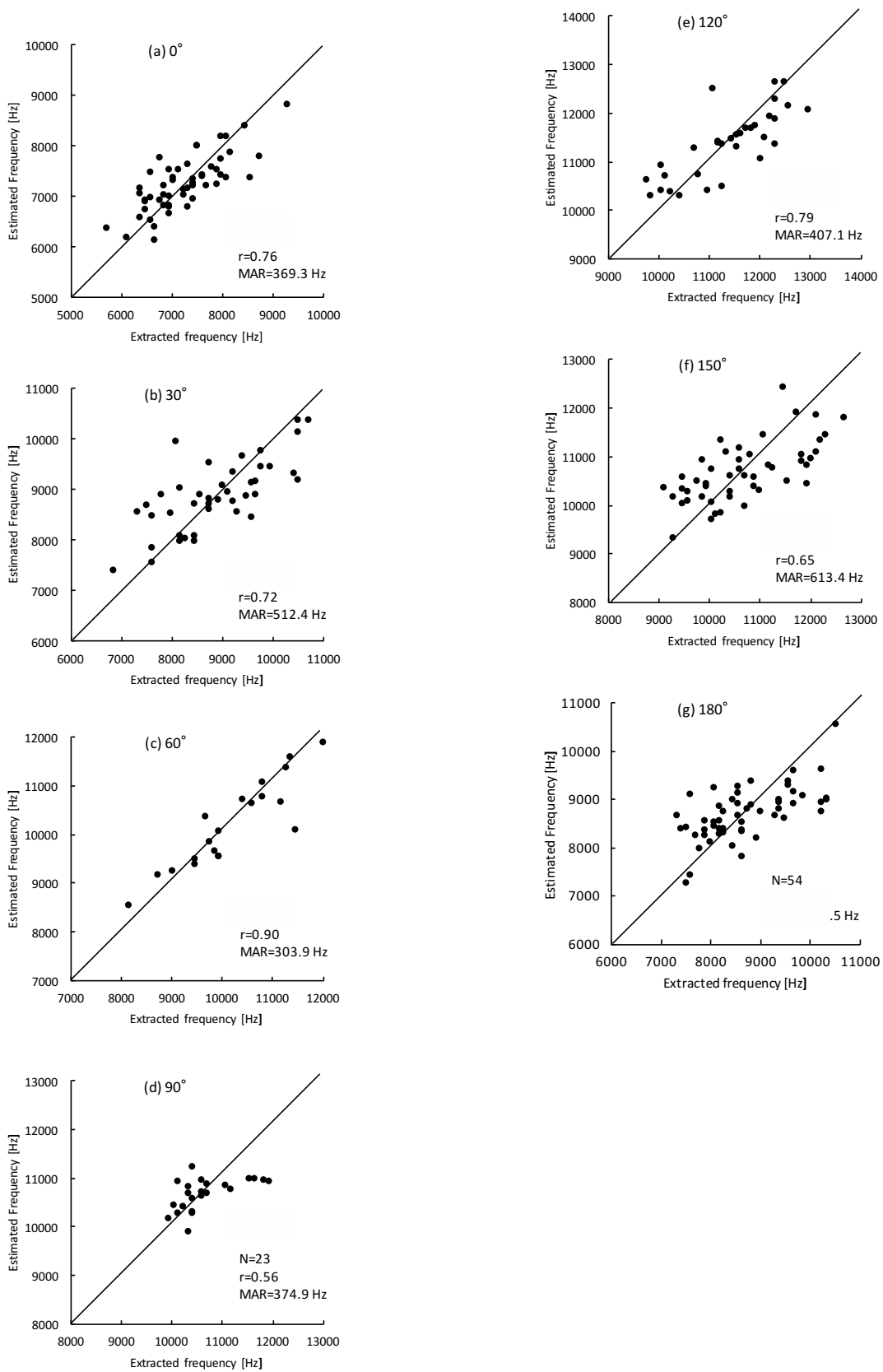


Fig. 4 Relationship between the frequencies extracted from the early HRTF and those estimated from the anthropometry of the listener's pinnae.