

初期頭部インパルス応答による正中面音像定位

飯田 一博（千葉工業大学先進工学部）

太田 正人（千葉工業大学大学院）

2017年10月6日

日本音響学会建築音響研究委員会

初期頭部インパルス応答による正中面音像定位

Median plane sound image localization using early head-related impulse response

飯田 一博† 太田 正人‡

Kazuhiro IIDA† and Masato OOTA‡

{†先進工学部, ‡大学院工学研究科} 千葉工業大学

{†Faculty of Advanced Engineering, ‡Graduate School} Chiba Institute of Technology

内容概要 いくつかの従来の知見を併せると、正中面音像定位において重要なスペクトル情報はN1, N2, P1, P2の概形であり、これらは耳介で生成され、頭部インパルス応答(HRIR)の初期部分に含まれていると推論される。本研究では、上半球正中面7方向(0°-180°, 30°間隔)において5人の被験者のHRIRを測定し、4種類の継続長(0.25ms, 0.5ms, 1ms, 2ms)の初期HRIRを作成した。その振幅スペクトルの分析と音像の上昇角および距離に関する音響心理実験を行った。その結果、初期1msのHRIRには主要なノッチやピークの概形の情報が含まれ、上半球正中面において全区間HRIRと同等の音像定位(上昇角および距離の知覚)を実現できることが示唆された。

1. はじめに

頭部伝達関数(HRTF: head-related transfer function)の5 kHz以上のノッチやピークが音像の上昇角知覚に貢献していることは広く知られている(Butler and Belendiuk 1977, Mehrgardt and Mellert 1977, Musicant and Butler 1984)。

また、これらのノッチとピークは、その微細構造よりもむしろ概形が重要であることも報告されている。例えばAsano *et al.* (1990) はARMAモデルで平滑化したHRTFを用いて正中面音像定位実験を行い、仰角判断の主要な手掛かりは5 kHz以上にあり、ピークやノッチの微細部分ではなく巨視的なパターン情報が音像定位に利用されていることを示した。Kulkarni and Colburn (1998) は水平面内(方位角: -45°, 0°, 45°, 180°)のHRTFの振幅スペクトルを7段階で平滑化して音像定位実験を行い、ノッチやピークの概形と比較すると、微細なスペクトル構造は相対的に重要ではないことを示した。ただし、ノッチが消滅するような平滑化を施すと、音像が上昇することも報告している。

さらに、多数のノッチやピークのうち、いくつかのノッチやピークはスペクトラルキューとして特に重要な役割を果たしていることも報告されている。Iida *et al.* (2007) とIida and Ishii (2018) は受聴者の実測HRTFから抽出したノ

ッチやピークの一部もしくは全部で再構成したパラメトリックHRTFを用いて正中面音像定位実験を行い、上半球正中面全体にわたって良好な音像定位を実現できるノッチとピークの最小構成は、4kHz以上において最も周波数の低い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。

以上の知見を併せると、正中面の音像定位においては、HRTFにおけるN1, N2, P1, P2のスペクトルの概形が重要であることが推論される。

一方、ノッチとピークの成因についても研究が進められ、これらは主に耳介で生成されていることが報告されている。Gardner and Gardner (1973) は耳介の窪みを順次埋めて正中面音像定位実験を行い、耳介の閉塞を進めると音像定位性能が低下することを示した。Musicant and Butler (1984) は4 kHzを遮断周波数とする高域通過ノイズを用いた音像定位実験により、耳介を埋めることにより多くの前後誤判定が発生することを報告した。Iida *et al.* (1998) は耳介の主要な3つの窪み、すなわち舟状窩、三角窩、耳甲介腔を粘土で埋めてHRTF測定と音像定位実験を行い、耳甲介腔を埋めるだけで4 kHzから10 kHzの間のピークとノッチが消滅し、このHRTFの音像定位精度は通常のHRTFの

それと比較して統計的に有意に低下することを報告した。Takemoto *et al.* (2012) は FDTD (finite-difference time-domain) 法を用いて頭部形状から HRTF を計算し、耳介が HRTF のピーク・ノッチパターンを決定していると報告した。

このような耳介の影響は、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、胴体からの反射波は HRIR の後期部分に含まれていると考えられる。すなわち、正中面定位に重要なノッチやピークは HRIR の初期で生成されると推測される。飯田他(2010)および Iida *et al.* (2011)は、正面方向の HRIR を用いて、矩型窓の継続長を変化させて初期部分を切り出して振幅スペクトルを観察し、N1, N2, P1, P2 のスペクトルの概形は HRIR の初期 0.5–1ms 程度で生成されることを報告した。竹本他(2011)も FDTD 法を用いた数値シミュレーションで同様の結果を報告している。

以上の知見は、正中面音像定位において重要な情報、つまり N1, N2, P1, P2 の概形は耳介で生成され、それは HRIR の初期部分に含まれていることを示唆している。したがって、HRIR の初期部分(以降、初期 HRIR とよぶ)により精度の高い正中面音像定位ができるのではないかと期待される。

しかしながら、正面以外の方向においても正面と同様に初期 HRIR で振幅スペクトルの概形が生成されるか否かは不明であり、また初期 HRIR で精度の高い正中面音像定位ができる否かも不明である。

このような背景のもと、本論文では上半球正中面において 4 種類の継続長(0.25ms, 0.5ms, 1ms, 2ms)の初期 HRIR を作成し、振幅スペクトルを分析した。さらに、2つの音響心理実験により、初期 HRIR により知覚される音像の上昇角と距離を検証した。

2. 初期 HRIR

2.1. HRIR の測定方法

20代の4人の男性被験者(HRM, ISI, OOT, TKH)および1人の女性被験者(SKW)の上半球正中面における7つの上昇角(0°–180°, 30°間隔)の HRIR を無響室で測定した。ここで上昇角は、矢状面(本実験では正中面)内の正面方向を

起点とした角度で、前方が 0°, 上方が 90°, 後方が 180°である(Morimoto and Aokata 1984)。

直径 80mm のスピーカ(FOSTEX FE83E)を上半球正中面に 30°間隔で設置して測定用信号を提示した。スピーカから被験者の頭部中心までの距離は 1.2m である。測定用信号は swept sine wave (2^{18} サンプル)で、サンプリング周波数は 48 kHz である。被験者の外耳道入口で測定用信号を収録するため、耳栓型マイクロホン(飯田 2017a)を用いた。マイクロホンの振動板は被験者の外耳道入口に位置し、blocked-entrances condition (Shaw and Teranishi, 1968)とみなせる。このような測定条件において、式(1)により HRTF を求めた。

$$HRTF_{l,r}(\omega) = G_{l,r}(\omega)/F(\omega) \quad (1)$$

ここで、 $F(\omega)$ は被験者のいない自由空間において被験者の頭部中心に相当する位置で測定したインパルス応答 $f(t)$ のフーリエ変換であり、 $G_{l,r}(\omega)$ は自由空間において被験者の外耳道入口で測定したインパルス応答 $g_{l,r}(t)$ のフーリエ変換である。添字の l,r はそれぞれ左耳、右耳を表す。このようにして求めた HRTF を逆フーリエ変換して HRIR を求めた。

2.2. 初期 HRIR の作成方法

以下のアルゴリズムにより初期 HRIR を作成した。

(1) HRIR の振幅の絶対値が最大となるサンプル(S_{max})を検出する。

(2) S_{max} を中心として 4 項、 $2N$ (時間軸上で前に N 、後ろに N)サンプルの Blackman-Harris 窓で HRIR を切り出す。本研究では、 N は 12, 24, 48, 96 サンプルの 4 種類とした。これらはそれぞれ 0.25, 0.5, 1, 2ms に相当する。

さらに、矩型窓で切り出すことにより比較用の HRIR を作成した。切り出しの終了点は S_{max} から時間軸上で後ろへ 256 サンプル以降で初めてゼロクロスするサンプルとした。以降、これを全区間 HRIR とよぶ。

Fig.1 に全区間 HRIR と初期 HRIR の算出例(被験者 HRM の正面方向)を示す。全区間 HRIR では S_{max} から約 4ms で応答が収束している。初期 HRIR(0.25ms)では、応答の最初の正のピークだけが含まれ、0.5ms では最初の正のピークと負のピークが含まれている。しかし、窓の影響で負のピークの絶対値は全区間 HRIR より小さい。1ms では2つ

めの正のピークまで含まれるが、その絶対値は全区間 HRIR より小さい。2ms では後期の微細な応答以外ほとんど含まれている。

また、Fig.2 に被験者 HRM の正中面 7 方向の全区間 HRTF と初期 HRTF の振幅スペクトルを示す。振幅スペクトルは 512 サンプルの FFT により求めた。周波数分解能は 93.75 Hz である。全区間 HRTF(実線) では、いずれの上昇角でも多数のノッチとピークが観察される。従来報告されているように(飯田 2017b), 前方や後方のノッチは深く、上方では浅い。ノッチ周波数は前方から上方になるにつれ高くなり、ピーク周波数は上昇角に依存せずほぼ一定である。

初期 HRTF(0.25ms)では、ノッチが 1 個(0°-120°)もしくは 2 個(150°,180°)生じたが、その多くは浅く、周波数も全区間 HRTF とは一致しない。ピークは 2 個(0°-120°)もしくは 3 個(150°,180°)生じたが、そのレベルは全区間 HRTF より低くなる傾向がある。

初期 HRTF(0.5ms)では、1-3 個のノッチが生じたが、全区間 HRTF のノッチとは一致しない。ピークは 3-4 個生じた。最も周波数の低いピークについては、全区間 HRTF のピーク(P1)とほぼ一致するが、それより周波数の高いピークについては、多くの場合一致しない。

初期 HRTF(1ms)では、スペクトルの微細構造は異なるが、ノッチとピークの概形は全区間 HRTF と同様であり、これらの周波数は全区間 HRTF とほぼ一致した。ただし、ピークのレベルはやや低くなる傾向がある。60°において非常に鋭いノッチが発生したが、これは他の被験者では見られなかった。

初期 HRTF(2ms)のスペクトルは微細構造も含めて全区間 HRTF とほぼ一致した。

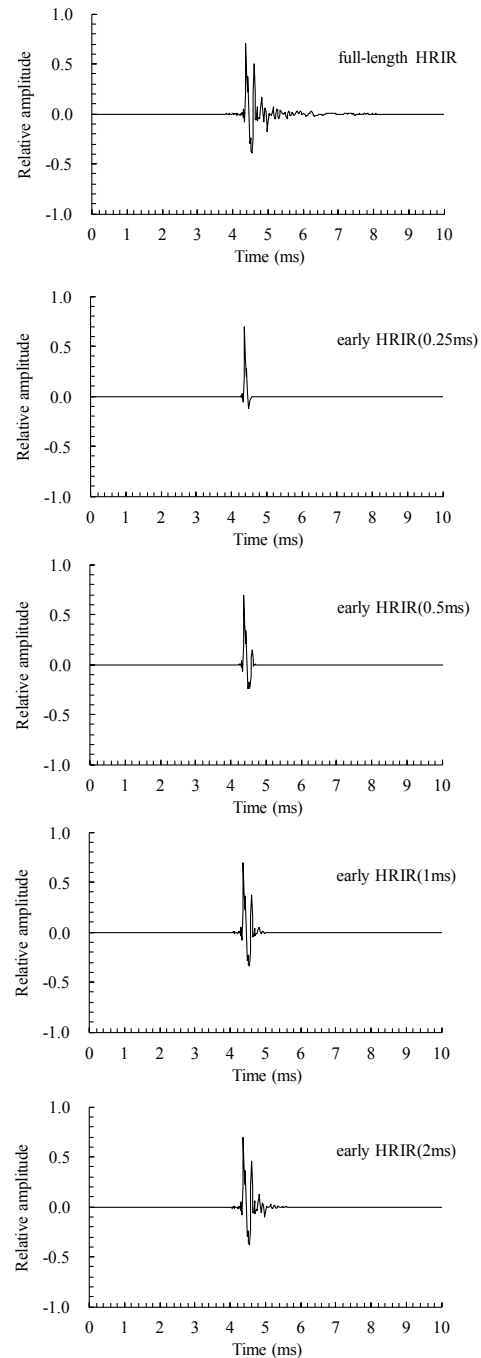


Fig. 1. Examples of the full-length HRIR and the early HRIRs for front direction of subject HRM.

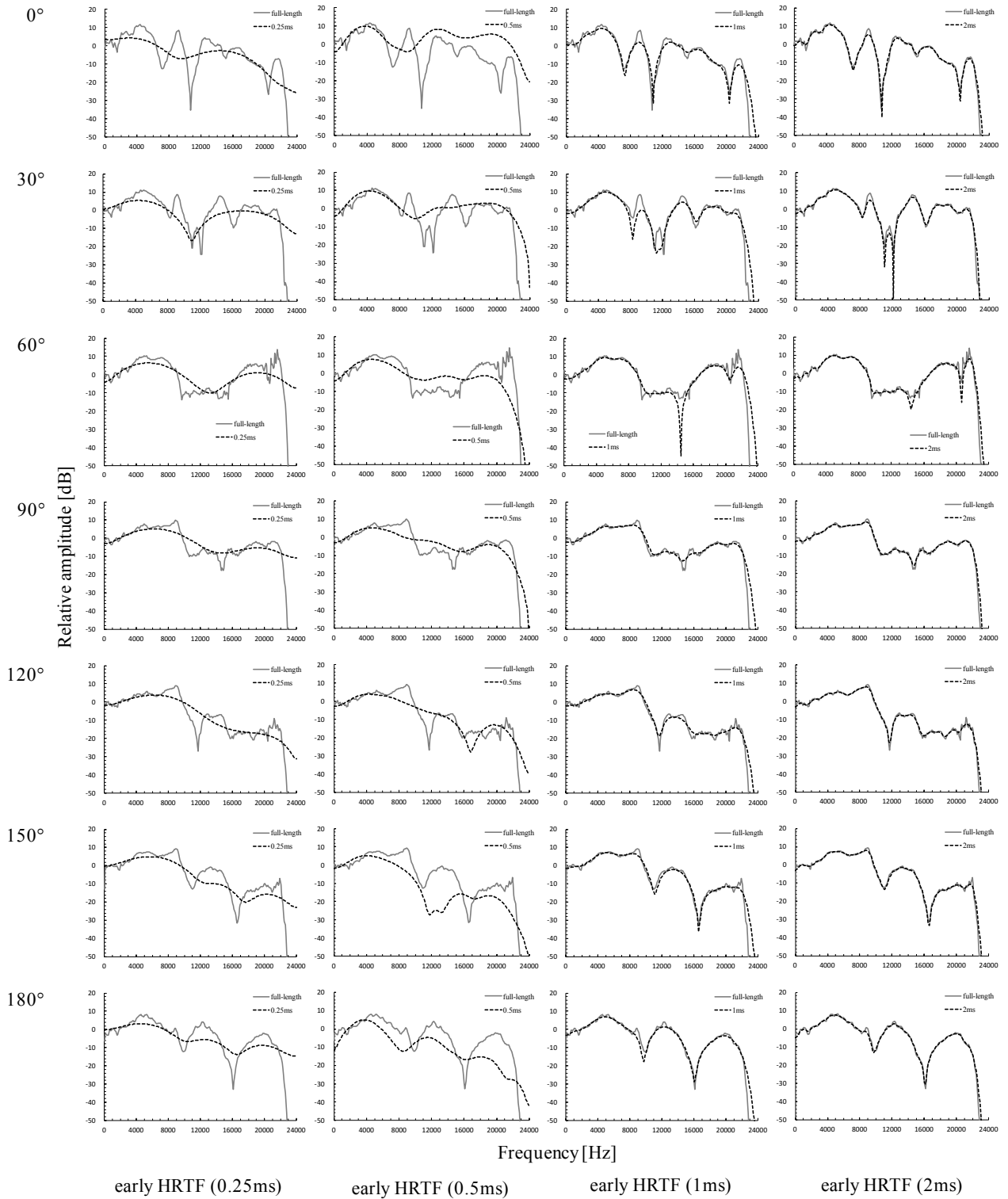


Fig. 2. Examples of the amplitude spectra of the full-length HRTFs and the early HRTFs for the seven vertical angles for subject HRM. The solid lines and the broken lines indicate the full-length HRTFs and the early HRTFs, respectively.

3. 実験 1: 音像の上昇角の知覚精度

上半球正中面において, 初期 HRIR により知覚される音像の上昇角を検証した.

3.1. 実験方法

実験システムは, ノート PC (Panasonic CF-R8), オーディオインタフェース(RME Fireface 400), アンプ(Marantz PM4001), オープンタイプヘッドホン(AKG K1000), 耳栓型マイクロホン, マイクロホンアンプ(Roland M-10MX)で構成した. 実験は防音室で行った. 室の内法は 4.6m (W) × 5.8m (D) × 2.8m (H)で, 暗騒音は 19.5 dB(A)であった.

被験者本人の全区間 HRIR および 2.2 節の方法で作成した 4 種類の継続長(0.25, 0.5, 1.0, 2.0ms)の初期 HRIR を用いた. 目標方向は, 上半球正中面内 7 方向(0°–180°, 30°間隔)である. 被験者は 2.1 節で HRIR の測定を行った 5 人(HRM, ISI, OOT, SKW, TKH)である.

外耳道入口を開放した状態における鼓膜での音圧 P を再現するため, 閉塞した状態の外耳道入口での信号 ($S \times HRTF$)に以下に示す補正フィルタ G (Møller *et al.*, 1995) を施してヘッドホンから提示した.

$$P(\omega) = S(\omega) \times HRTF(\omega) \times G(\omega), \quad (2)$$

$$G(\omega) = \frac{1}{M(\omega) \times PTF(\omega)} \times \frac{Z_{ear\ canal}(\omega) + Z_{headphone}(\omega)}{Z_{ear\ canal}(\omega) + Z_{radiation}(\omega)} \quad (3)$$

$$\triangleq \frac{1}{M(\omega) \times PTF(\omega)} \times PDR(\omega), \quad (4)$$

ここで, S は音源信号, $M \times PTF$ はヘッドホンから耳栓型マイクロホンまでの伝達関数, $Z_{ear\ canal}$ と $Z_{headphone}$ はそれぞれ外耳道およびヘッドホンのインピーダンス, $Z_{radiation}$ は外耳道から外側をみた放射インピーダンスである. G の第 2 項は PDR (pressure division ratio)とよばれ, Møller *et al.*はこれが 1 となるヘッドホンを FEC(free-air equivalent coupling to the ear)ヘッドホンと定義した. 本実験では, 彼らが FEC ヘッドホンとみなせるとした K1000 (AKG)を用いた.

$M \times PTF$ の補正は 200 Hz から 17 kHz の周波数帯域で以下のように行った. 被験者は防音室の中央に座り, 耳栓型マイクロホンを外耳道に装着した. マイクロホンの振動板は 2.1 節の HRIR 測定と同様に外耳道入口に位置した. 被験者にヘッドホンを装着して M 系列信号(48kHz サンプリング, 12 次, 繰返しなし)を提示し, 耳栓型マイクロホン

までの伝達関数 $M \times PTF$ を測定した. さらに, ヘッドホンを装着したまま, その位置を変えないようにして耳栓型マイクロホンだけを取り外した(耳介はヘッドホンに覆われていない). このようにして求めて補正フィルタ G により, ヘッドホンから耳栓型マイクロホンまでの 200 Hz から 17 kHz までの代表的な peak-to-peak レンジは約 20 dB から約 3 dB まで低減した.

音源信号は 200Hz から 17kHz の広帯域ホワイトノイズである. 刺激は被験者の左右の外耳道入口で 63 dB SPL となるように提示した. 刺激の提示時間は 1.2 秒(立上がり, 立下がりそれぞれ 0.1 秒を含む)である.

回答方法は, 被験者が目標方向を推測することを避けるため, マッピング法とした. ノート PC の画面上に正中面を示す円周と前後上下を示す矢印を表示した. 被験者のタスクは, 知覚した上昇角を円周上にマウスでクリックすることである. 頭内定位した場合はチェックボックスに印を入れた.

1 回の試行で 35 刺激(5 種類の継続長 × 7 方向)をランダムな順で被験者に提示した. 1 試行の所要時間は約 7 分である. 各被験者に対して 10 試行ずつ実験を行った. つまり, 被験者は各刺激に対して 10 回ずつ回答した.

3.2. 回答分布

Fig. 3 に全区間 HRIR および初期 HRIR に対する 5 人の被験者の回答を示す. 縦軸は回答した上昇角, 横軸は目標上昇角(以降, 目標方向とよぶ)である. 円の直径は回答の頻度に比例し, 回答方向は 5°間隔で丸めている.

全区間 HRIR では, いずれの被験者においても, ほとんどの回答は目標方向付近に分布した. ただし, 被験者 OOT の回答は目標方向 60°と 120°において, TKH の回答は目標方向 150°において, それぞれ 90°付近に分布した.

初期 HRIR(0.25ms)では, 多くの回答は全区間 HRIR と同様の分布をしなかった. 被験者 HRM のほぼすべての回答は後方に分布し, ISI, OOT, SKW の回答は一部を除き広い範囲に分布した. 被験者 TKH のほぼすべての回答は前方から上方に分布した. 一方で, HRM 以外の 4 名の被験者では, 120°の回答は全区間 HRIR とほぼ同様の分布をした. 他にも, HRM の 180°, ISI の 0°, SKW の 90°と 150°, TKH の 150°では全区間 HRIR と同様の分布をした.

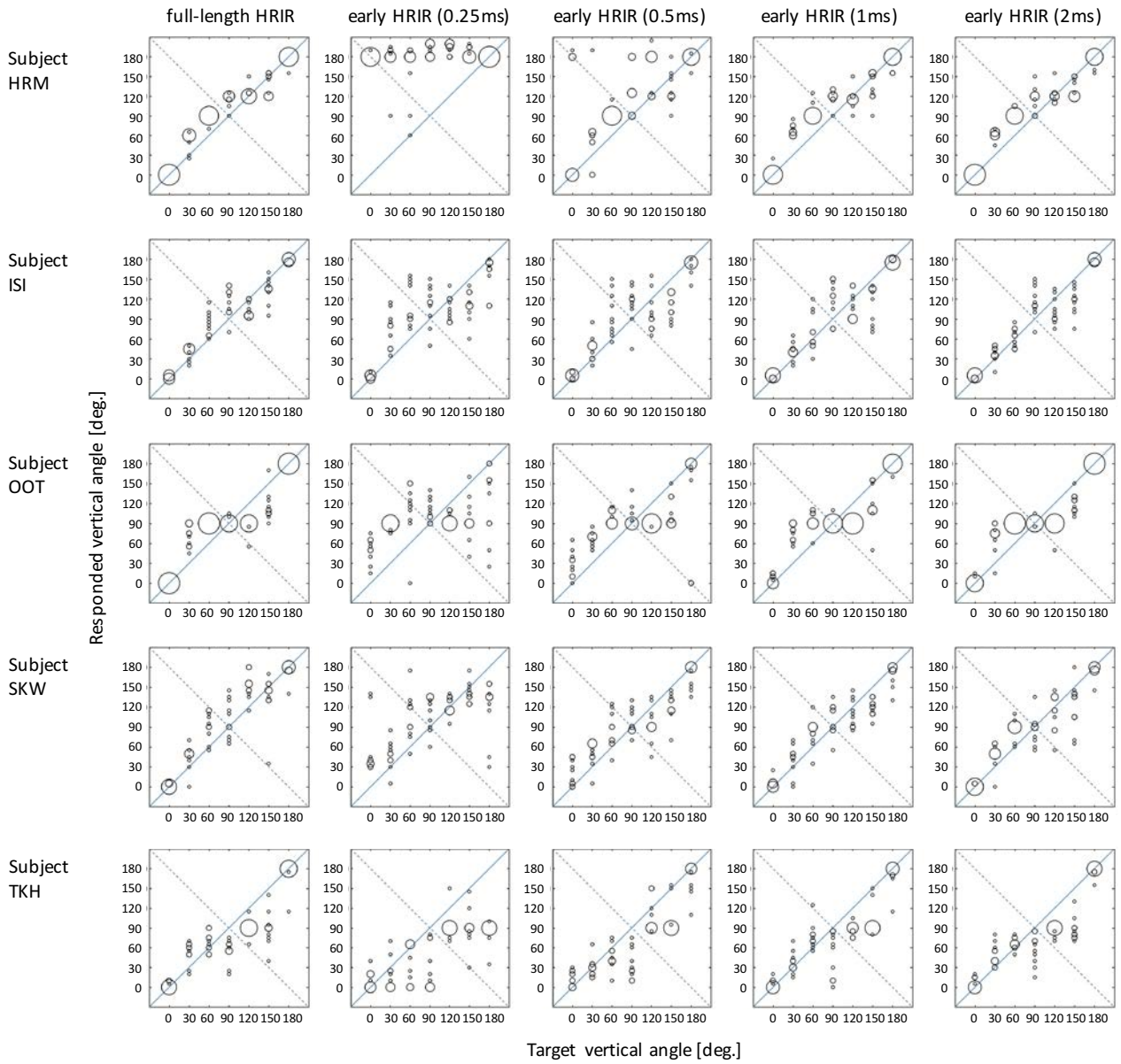


Fig. 3. Responses to the full-length HRIRs and the early HRIRs. The ordinate represents the responded vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the number of responses with a resolution of 5° .

初期 HRIR(0.5ms)でも、多くの回答は全区間 HRIR と同様の分布をしなかった。被験者 HRM では 0°, 90°, 120° において、それぞれ目標方向と 180° 付近の両方に回答が分布し、ISI では 60°-120° において全区間 HRIR と比較して広い範囲に回答が分布した。被験者 OOT では、0° の回答は前方から上方に分布し、180° の回答は 180° と 0° の両方に分布した。被験者 SKW の 0° では前方から斜め上方に分布した。一方で、被験者 HRM の 60° と 180°, ISI の 0° と 90°, OOT の 30° と 120°, SKW の 30°-90°, TKH の 90° では全区間 HRIR と同様の分布をした。このように、0.25ms と 0.5ms では全区間 HRIR と同様の分布をしない場合が多いが、120° を中心とする上方では全区間 HRIR と同様の分布をする傾向がみられた。

初期 HRIR(1ms)と初期 HRIR(2ms)では、ほとんどの回答は全区間 HRIR と同様の分布をした。

3.3. 平均定位誤差

被験者、HRIR の継続長、および目標方向ごとに平均定位誤差を求めた(Table 1)。また、全区間 HRIR と初期 HRIR の間に有意差があるか否かを検証するため Tukey の多重比較検定を行った(Table 2)。ここで、平均定位誤差は目標上昇角と回答した上昇角の差の絶対値の平均と定義する。

全区間 HRIR では、平均定位誤差は水平面上の 0°, 180° では小さく、水平面から上昇した 30° から 150° では大きくなる傾向がみられた。これは、従来報告されている傾向 (Carlile *et al.* 1997; Majdak *et al.* 2010) と同様である。

初期 HRIR(0.25ms)では、多くの被験者および目標方向において全区間 HRIR より大きな値となった。ただし、HRM 以外の 4 名の被験者では、120° においては全区間 HRIR と同等であった。いずれの被験者においても 1-6 方向で全区間 HRIR との間に有意差があり、特に 0° と 180° では多くの被験者に共通して有意差があった。全被験者の回答においては、0°-90° および 180° で全区間 HRIR と有意差があった($p < 0.01$)。

初期 HRIR(0.5ms)でも、多くの被験者および目標方向において全区間 HRIR より大きく、特に 0° では ISI を除く 4 名の被験者では大きな値となった。しかし、120° においては被験者 OOT, SKW, TKH では全区間 HRIR と同等であった。被験者 HRM, ISI, OOT では、0° を共通として 1-2 方

Table 1 Mean vertical localization error for each subject, HRIR and target vertical angle.

Subject	HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
HRM	full-length	0.5	24.2	27.7	25.5	4.5	16.5	3.1
	0.25ms	178.3	144.9	98.5	100.0	73.4	35.3	0.7
	0.5ms	70.9	43.3	32.3	40.8	39.3	24.6	3.3
	1ms	2.8	37.8	35.4	28.6	9.7	17.6	5.8
	2ms	0.5	30.7	32.8	26.6	7.1	20.2	4.7
ISI	full-length	2.8	11.8	23.5	29.3	16.4	18.4	3.1
	0.25ms	3.3	45.8	51.7	32.3	18.3	41.0	21.6
	0.5ms	5.2	18.2	36.9	30.6	31.2	43.1	9.6
	1ms	3.3	14.3	22.6	36.0	18.9	38.7	2.8
	2ms	3.3	11.6	11.8	31.0	22.8	31.6	3.4
OOT	full-length	0.6	40.3	30.5	3.1	34.1	37.8	0.4
	0.25ms	50.1	57.7	60.1	20.4	24.3	54.8	59.1
	0.5ms	29.2	37.0	39.3	9.6	30.3	43.5	40.3
	1ms	5.1	45.3	34.1	2.5	30.2	34.5	2.2
	2ms	3.0	41.9	30.4	2.5	34.3	28.9	0.5
SKW	full-length	1.9	20.3	33.3	23.5	30.2	20.3	6.3
	0.25ms	55.5	23.3	46.7	28.3	11.8	13.3	63.3
	0.5ms	20.7	24.5	28.9	15.3	28.8	28.9	14.4
	1ms	4.4	20.8	25.6	19.8	20.1	31.0	11.4
	2ms	1.7	22.9	27.3	15.4	22.2	33.8	6.0
TKH	full-length	2.7	25.4	12.7	35.1	30.2	61.7	7.2
	0.25ms	11.4	21.6	30.1	60.5	33.2	60.9	96.7
	0.5ms	14.1	9.4	19.4	54.0	25.7	48.3	16.9
	1ms	4.7	12.3	18.6	40.1	32.6	49.8	10.2
	2ms	5.7	20.6	8.3	33.2	34.4	61.6	3.6
All subjects	full-length	1.7	24.4	25.5	23.3	23.1	30.9	4.0
	0.25ms	59.7	58.7	57.4	48.3	32.2	41.1	48.3
	0.5ms	28.0	26.5	31.4	30.1	31.1	37.7	16.9
	1ms	4.1	26.1	27.3	25.4	22.3	34.3	6.5
	2ms	2.8	25.5	22.1	21.7	24.2	35.2	3.6

Table 2 Results of Tukey's multiple comparison test between the full-length HRIRs and the early HRIRs for mean vertical localization error.

Subject	Early HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
HRM	0.25 ms	**	**	**	**	**	*	
	0.5 ms	**				**		
	1 ms							
	2 ms							
ISI	0.25 ms		**					*
	0.5 ms	*						
	1 ms							
	2 ms							
OOT	0.25 ms	**	*	**	*			*
	0.5 ms	**						
	1 ms							
	2 ms							
SKW	0.25 ms	**						**
	0.5 ms							
	1 ms							
	2 ms							
TKH	0.25 ms							**
	0.5 ms							
	1 ms							
	2 ms							
All subjects	0.25ms	**	**	**	**			**
	0.5ms	**						
	1ms							
	2ms							

向で全区間 HRIR との間に有意差があった。全被験者の回答においても 0° で全区間 HRIR と有意差があった($p < 0.01$)。

初期 HRIR(1ms)と初期 HRIR(2ms)では、ほとんどの被験者および目標方向において、平均定位誤差は全区間 HRIR と同等であった。いずれの被験者、および全被験者におい

でも、すべての目標方向において全区間 HRIR との間に有意差は認められなかった。

以上まとめると、被験者と目標方向を問わず全区間 HRIR と統計的有意差のない平均定位誤差が得られる HRIR の継続長は 1ms および 2ms といえる。0.25ms および 0.5ms においては、多くの被験者および目標方向において、全区間 HRIR より大きな平均定位誤差となった。ただし、HRM を除く 4 名の被験者では、120°においては全区間 HRIR と同等の平均定位誤差が得られた。

3.4. 前後誤判定率

Table 3 に全被験者の回答から求めた前後誤判定率を示す。ここで、前後誤判定率は上半球正中面において目標方向と異なる四分円に音像を知覚した率と定義する。さらに、全区間 HRIR と初期 HRIR の間に統計的有意な差があるか否かを検証するためにカイ 2 乗検定を行った(Table 4)。

全区間 HRIR では、目標方向 0°, 180°においては 0%であった。原理上、前後誤判定率は上方付近で高くなるが、60°で 40%、120°で 20%であった。

初期 HRIR(0.25ms)では、すべての方向において全区間 HRIR より高かった。0°, 180°では 24%, 26%であった。これは 3.2 節で述べたように 0°を後方に回答したり、180°を前方から上方に回答したりしたためである。他の 4 方向では 24%–58%であり、30°($p < 0.01$), 60°($p < 0.05$), 150°($p < 0.05$)で有意差があった(0°と 180°では、全区間 HRIR で 0%であったため検定ができない)。

初期 HRIR(0.5ms)でも 0°, 180°で前後誤判定が発生し、それぞれ 8%, 4%であった。120°, 150°では全区間 HRIR より高く、150°で有意差($p < 0.05$)があった。

初期 HRIR(1ms)では、0°と 180°においては 0%であった。しかし、30°, 120°, 150°で全区間 HRIR より高く、30°, 120°で有意差($p < 0.05$)があった。

初期 HRIR(2ms)でも 0°と 180°においては 0%であった。しかし、120°, 150°で全区間 HRIR より高く、120°で有意差($p < 0.01$)があった。

Table 3 Ratio of front-back confusion for each HRIR and target vertical angle.

Subject	HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
All subjects	full-length	0.00	0.02	0.40	-	0.20	0.14	0.00
	0.25ms	0.24	0.32	0.58	-	0.24	0.24	0.26
	0.5ms	0.08	0.02	0.40	-	0.28	0.26	0.04
	1ms	0.00	0.06	0.38	-	0.34	0.16	0.00
	2ms	0.00	0.02	0.40	-	0.38	0.20	0.00

Table 4 Results of chi-square tests between the full-length HRIRs and the early HRIRs for ratio of front-back confusion.

Subject	Early HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
All subjects	0.25ms	-	**	*	-	-	*	-
	0.5ms	-	-	-	-	-	*	-
	1ms	-	*	-	-	*	-	-
	2ms	-	-	-	-	**	-	-

3.5. 頭内定位率

全被験者の回答から求めた頭内定位率を Table 5 に示す。全区間 HRIR では、120°および 150°で頭内定位が発生したが、他の方向では発生しなかった。初期 HRIR(0.25ms)ではすべての方向で発生した。1ms および 2ms においては全区間 HRIR とほぼ同等であった。

Table 5 Ratio of inside-of-head localization for each HRIR and target vertical angle.

Subject	HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
All subjects	full-length	0.00	0.00	0.00	0.00	0.12	0.06	0.00
	0.25ms	0.06	0.16	0.10	0.06	0.08	0.14	0.18
	0.5ms	0.06	0.04	0.02	0.08	0.02	0.04	0.04
	1ms	0.02	0.00	0.00	0.02	0.04	0.02	0.00
	2ms	0.02	0.00	0.00	0.02	0.02	0.00	0.00

4. 実験 2: 音像距離の知覚精度

実験 1 終了後、被験者から音像距離が異なる場合があるとの内観報告があった。そこで、1ms および 2ms の初期 HRIR は音像距離においても全区間 HRIR と同等か否かを検証した。

4.1. 実験方法

被験者本人の全区間 HRIR および 2.2 節の方法で作成した 2 種類の継続長(1, 2ms)の初期 HRIR を用いた。目標方向は正中面内 3 方向(0°, 90°, 180°)である。実験システム、音源信号および被験者は実験 1 と同様である。

シェッフェの一対比較の浦の変法を用いて音像距離を評価した。被験者に 2 つの刺激を対にして提示し、先に提示された刺激の音像距離に対し、後に提示された刺激の音

像距離が近いかわ遠いかを回答させた。ここで、非常に近い場合は-2.0、どちらでもない場合は0.0、非常に遠い場合は+2.0とした。ノートPCの画面上に-2.0から+2.0まで0.1刻みの目盛りのついた数直線とスライダバーを表示し、音像距離の違いに応じて、マウスでスライダバーを目盛りに合わせるよう被験者に指示した。

9刺激(3種類の継続長×3方向)から72種類(9P₂)の刺激対を作成した。そのうち36対をランダムな順に並べたものを1試行とし、各被験者は練習試行1試行を含む11試行の実験を行った。ただし、練習試行は分析から除外し、各刺激対に対する5回の回答を分析した。

4.2. 実験結果

Fig. 4に全被験者の回答を用いて求めた音像距離の尺度値を示す。エラーバーは標準誤差を示す。音像距離は、継続長よりもむしろ目標方向に強く影響を受け、前方から上方、後方となるにつれ近くなった。継続長の影響についてみると、初期HRIRの音像距離は全区間HRIRのそれと比較して0°では近くなり、90°と180°では遠くなった。しかし、その差は標準誤差の範囲内であった。

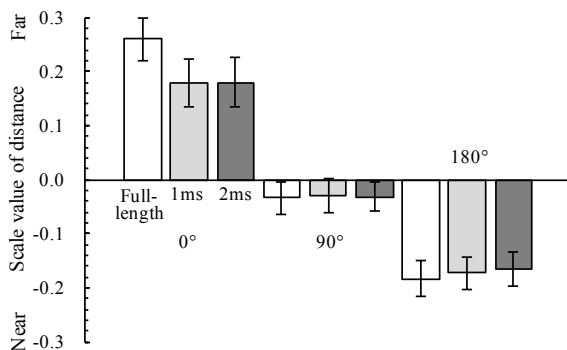


Fig. 4. Scale value of the distance of a sound image. The error bar denotes the standard error.

継続長および目標方向を要因とした2元配置の分散分析を行った(Table 6)。その結果、継続長には主効果が認められず、目標方向には主効果が認められた($p < 0.01$)。要因間の交互作用は認められなかった。Tukeyの多重比較検定を行った結果、全区間、1ms、2msのいずれにおいても、目標方向間(0°と90°、0°と180°、90°と180°)で音像距離に有意な差が認められた($p < 0.01$)。ただし、1msおよび2msにおける90°と180°の間の有意水準は5%であった。

以上より、目標方向0°、90°、180°においては、1msおよ

び2msの初期HRIRで知覚される音像距離は全区間HRIRと同等であると考えられる。なお、目標方向間で音像距離に有意な差が認められたことについては、今後の研究課題としたい。

Table 6 Two-way ANOVA table.

Variance Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Statistics	P-value
Main Effect A (Length of HRIR)	2	0.0220	0.0110	0.3453	0.7084
Main Effect B (Target vertical angle)	2	5.5260	2.7630	86.9025	2.0E-28
A vs.B Interaction	4	0.0915	0.0229	0.7196	0.5794
Error	216	6.8675	0.0318		
Total	224	12.5070			

5. 考察

初期HRIRの継続長と上昇角知覚精度の関係について考察する。

5.1. 考察1

実験1において、継続長0.25msと0.5msの初期HRIRの平均定位誤差は0°を含む複数の方向で全区間HRIRと統計的有意差が認められた。一方、1msと2msでは被験者と目標方向にかかわらず、全区間HRIRと有意差は認められなかった。まず、この理由について考察する。

被験者HRMの目標方向0°を例にとると、Fig4に示したように0.25msと0.5msでは後方への回答が多く、1msと2msでは全区間HRIRと同様の回答であった。HRTFの振幅スペクトル(Fig.2)をみると、0.25msと0.5msでは全区間HRTFに存在する明確なノッチ(N1, N2)が生じなかった。一方、1msではスペクトルの微細構造は異なるものの、主要なノッチやピークの概形が全区間HRTFと一致し、2msでは微細構造も含めて全区間HRTFとほぼ一致した。このように0.25msと0.5msではスペクトル上に上昇角知覚の手掛かりが生じず、1msと2msでは生じたことが上昇角知覚精度に影響を与えたと考えられる。

5.2. 考察2

HRMを除く4名の被験者については、120°では0.25msでも全区間HRIRと同等の平均定位誤差が得られた。この理由を考察する。

従来、上方のHRTFのノッチは前方や後方と比較して浅く、不明瞭であることが報告されている(飯田 2017b)。

Fig.5 に 5 人の被験者の 120°の全区間 HRTF と初期 HRTF(0.25ms)の振幅スペクトルを示す。全区間 HRTF のスペクトルをみると、被験者 ISI と SKW においては N1 が浅く、N2 も 16 kHz–17 kHz 付近にあるため、刺激の周波数帯域(200 Hz–17 kHz)には実質的にはノッチは存在しない。OOT と TKH では、N1 と N2 は存在するが、OOT では N1 の帯域幅が狭く N2 は浅い。TKH では、N1 と N2 が繋がって広い窪みとなった。一方、被験者 HRM のスペクトルには明確な N1 と N2 が存在する。

0.25ms の HRTF では、いずれの被験者においても顕著なノッチは生じなかったが、ノッチ以外については全区間 HRTF と同様のスペクトル形状であった。したがって、全区間 HRTF で明確なノッチが存在しない場合は、0.25ms のスペクトルは全区間 HRTF と類似のものになる。これが 4 名の被験者の平均定位誤差が全区間 HRIR と同等となった理由であると考えられる。

試みに、0.25ms の 120°の平均定位誤差の順で被験者を並べると式(5)のようになる。全区間 HRTF でノッチが浅い被験者では平均定位誤差が小さく、ノッチが明確な被験者では平均定位誤差が大きくなる傾向がみられる。

$$SKW(11.8^\circ) < ISI(18.3^\circ) < OOT(24.3^\circ) < TKH(33.2^\circ) < HRM(73.4^\circ) \quad (5)$$

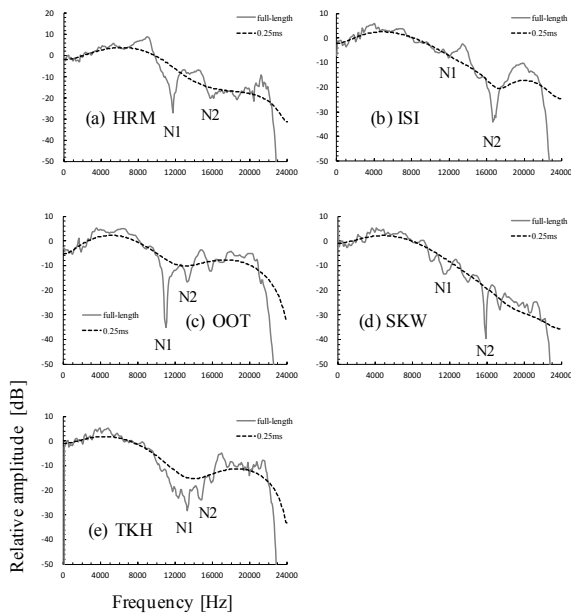


Fig. 5. Amplitude spectra of the full-length HRTFs and the early HRTFs(0.25ms) for the vertical angle of 120°. The solid lines and the broken lines indicate the full-length HRTFs and the early HRTFs, respectively.

6. 結論

本研究では、上半球正中面 7 方向(0°–180°, 30°間隔)において 5 人の被験者の HRIR を測定し、4 種類の継続長(0.25ms, 0.5ms, 1ms, 2ms)の初期 HRIR を作成した。その振幅スペクトルの分析と音像の上昇角および距離に関する音響心理実験を行い、以下の結果を得た。

(1) 0.25ms および 0.5ms の初期 HRTF では 1–3 個のノッチが生じたが、全区間 HRTF のノッチとは一致しなかった。1ms では全区間 HRTF と同様のノッチとピークが生じたが、スペクトルの微細構造は異なった。2ms の振幅スペクトルは微細構造も含めて全区間 HRTF とほぼ一致した。

(2) 0.25ms および 0.5ms の初期 HRIR では、平均定位誤差は多くの被験者および目標方向において全区間 HRIR より大きくなった。0.25ms では 0°–90°および 180°で、0.5ms では 0°で統計的有意差が認められた。一方、1ms と 2ms では平均定位誤差は全区間 HRIR とほぼ同等であり、7 方向のいずれにおいても有意差は認められなかった。

(3) 1ms および 2ms の初期 HRIR の音像距離は全区間 HRIR のそれとほぼ同等であり、統計的有意な差は認められなかった。

以上より、初期 1ms の HRIR には主要なノッチやピークの概形の情報が含まれ、上半球正中面において全区間 HRIR と同等の音像定位(上昇角および距離の知覚)を実現できることが示唆された。

最後に、実用面において重要な点を付記しておきたい。初期 HRIR は耳介の特性を主とし、胴体からの反射が含まれていないため、受聴者個人のノッチやピークの特性(周波数、レベル、先鋭度)を受聴者の耳介形状から推定できる可能性がある。従来の研究では、正面方向 HRTF のノッチとピークの周波数を耳介形状から推定する方法が提案されているが(Iida *et al.* 2014, Spagnol and Avanzini 2015, Mokhtari *et al.* 2015, 2016)、正面以外の方向の周波数を推定する方法は確立されていない。また、ノッチやピークのレベルや先鋭度の推定方法も確立されていない。初期 HRIR を利用することで、これらの問題が解決され、HRTF の個人化が進展することが期待される。

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。

文献

- Asano F, Suzuki Y, Sone, T. Role of spectral cues in median plane localization. *J Acoust Soc Am* 1990; 88:159–68.
- Carlile S, Leong P, Hyams S. The nature and distribution of errors in sound localization by human listeners. *Hear Res* 1997;114:179–96.
- Butler A, Belendiuk K. Spectral cues utilized in the localization of sound in the median sagittal plane. *J Acoust Soc Am* 1977;61:1264–9.
- Gardner B, Gardner S. Problem of localization in the median plane: effect of pinna cavity occlusion. *J Acoust Soc Am* 1973;53:400–8.
- Iida K, Yairi M, Morimoto M. Role of pinna cavities in median plane localization. In: Proceedings of 16th international congress on acoustics 1998;845–6.
- Iida K, Itoh M, Itagaki A, Morimoto M. Median plane localization using parametric model of the head-related transfer function based on spectral cues. *Appl Acoust* 2007;68:835–50.
- Iida K, Gamoh N, Ishii Y, Morimoto M. Contribution of the early part of the head-related impulse responses to the formation of two spectral notches of vertical localization cues. In: Proceedings of FORUM ACUSTICUM 2011 (Aalborg-Denmark).
- Iida K, Ishii Y, Nishioka S. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. *J Acoust Soc Am* 2014;136:317–33.
- Iida K, Ishii Y. Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization. *Appl Acoust* 2018;129:239–47.
- Kulkarni A, Colburn HS. Role of spectral detail in sound-source localization. *Nature* 1998;396:747–9.
- Majdak P, Goupell MJ, Laback B. 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. *Atten Percept Psychophys* 2010;72:454–69.
- Mehrgardt S, Mellert V. Transformation characteristics of the external human ear. *J Acoust Soc Am* 1977;61:1567–76.
- Mokhtari P, Takemoto H, Nishimura R, Kato H. Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry. *J Acoust Soc Am* 2015;137: 690–701.
- Mokhtari P, Takemoto H, Nishimura R, Kato H. Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry. *J Acoust Soc Am* 2016;140: 814–31.
- Møller H, Hammershøi D, Jensen CJ, Sørensen MF. Transfer characteristics of headphones measured on human ears. *J Audio Eng Soc* 1995;43:203–17.
- Morimoto M, Aokata H. Localization cues of sound sources in the upper hemisphere. *J Acoust Soc Jpn (E)* 1984;5:165–73.
- Musicant A, Butler R. The influence of pinnae-based spectral cues on sound localization. *J Acoust Soc Am* 1984;75:1195–200.
- Shaw EAG, Teranishi R. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J Acoust Soc Am* 1968;44:240–9.
- Spagnol S, Avanzini F. Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model. In: Proceedings of the 18th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, Nov 30–Dec 3 (2015).
- Takemoto H, Mokhtari P, Kato H, Nishimura R, Iida K. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *J Acoust Soc Am* 2012;132:3832–41.
- 飯田一博, 蒲生直和, 石井要次: 頭部伝達関数の第1・第2ノッチの検出方法に関する一考察, 日本音響学会講演論文集, 473-6 (2010年9月).
- 飯田一博: 頭部伝達関数の基礎と3次元音響システムへの応用, pp.224-9, コロナ社 (2017a).
- 飯田一博: 頭部伝達関数の基礎と3次元音響システムへの応用, pp.29-30, コロナ社 (2017b).
- 竹本浩典, Parham M, 加藤宏明, 西村竜一, 飯田一博: 耳介における初期反射波の解析, 日本音響学会講演論文集, 521-4 (2011年3月).