

## 初期頭部インパルス応答による上半球正中面音像定位

☆太田正人, 飯田一博 (千葉工大)

## 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が広く知られている。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[1,2]。

さらに, 多数のノッチやピークのうち, 4 kHz 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)は特に重要な役割を果たしていることが報告されている[3,4]。

これらより, 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[5]。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[6]は正面方向の HRIR を用いて, 矩形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 の概形は HRIR の初期 0.5-1 ms 程度で生成されることを報告した。

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か否か, さらにできるとすればその最小の継続長はいくつであ

るのかも不明である。

本研究では正中面において通常の HRIR と同程度の定位性能となる初期 HRIR の最小の継続長を検証した。

## 2 初期 HRIR

20 代の 4 名の男性被験者(HRM, ISI, OOT, TKH)および 1 名の女性被験者(SKW)の上半球正中面における 7 つの上昇角(0°-180°, 30°間隔)の HRTF を無響室で測定した。測定用信号は swept sine wave で, サンプリング周波数は 48 kHz である。得られた HRTF をフーリエ逆変換して HRIR を求めた。

以下のアルゴリズムにより 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(Smax)を検出する。

(2) Smax を中心として 4 項, 2N サンプルの Blackman-Harris 窓で HRIR を切り出す。N は 12, 24, 48, 96 サンプルの 4 種類とした。これはそれぞれ, 0.25, 0.5, 1.0, 2.0ms に相当する。

比較のため, 上記(1)と同様に Smax を求め, 矩形窓で切り出した。ただし, 切り出しの終了点は Smax から時間的後ろへ 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。

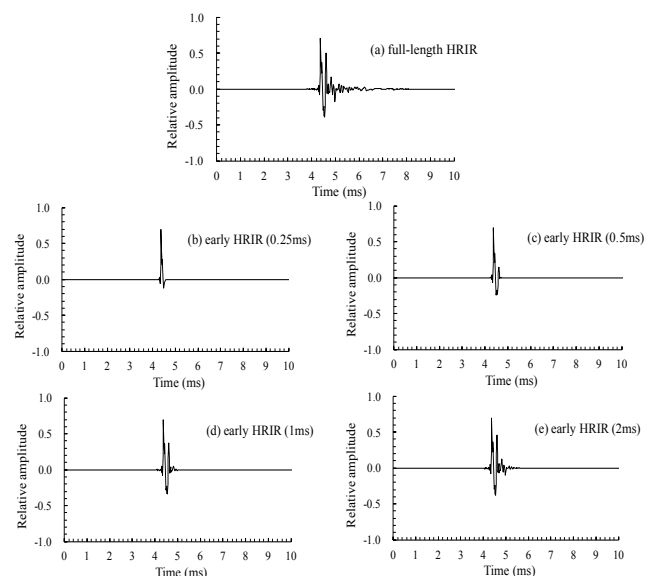


Fig. 1. Examples of a full-length HRIR and the early HRIRs for front direction of subject HRM.

\*Upper median plane sound image localization using early head-related impulse response, by OOTA, Masato and IIDA, Kazuhiro (Chiba Institute of Technology).

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 Smax 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だけが含まれている. 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はほとんど含まれている.

Fig.2 に被験者 HRM の正面方向の全区間 HRTF 及び初期 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)では多数のノッチとピークが観察される. 初期 HRTF(0.25ms)では, ノッチが 1 個生じたが, レベルは浅く, 周波数も全区間 HRTF とは一致しない. ピークは 2 個生じたが, そのレベルは全区間 HRTF より低い. 0.5ms では, ノッチが 3 個生じたが, 2 個目および 3 個目のノッチのレベルは全区間 HRTF より浅い. ピークは 4 個生じたが, レベルは全区間 HRTF より低い. 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 と同様であった. 2ms では, 微細構造も含めて全区間 HRTF とほぼ一致した.

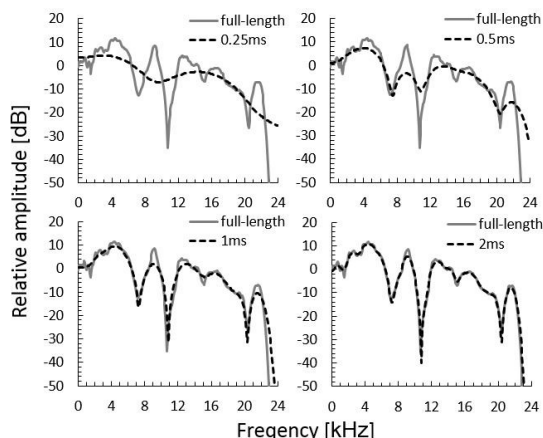


Fig. 2 Examples of the amplitude spectra of the full-length HRTF and the early HRTFs for front direction for subject HRM

### 3 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

実験は防音室にて行った. 実験システムの構成は, ノート PC(Panasonic CF-R8), オーディオインターフェース(RME Fireface UFX), アンプ

(Marantz PM4001), A/D コンバータ(Roland M-10MX), FEC ヘッドホン[7](AKG K1000), 耳栓型マイクロホンである. 白色雑音(200 Hz -17 kHz)に初期 HRIR 及び全区間 HRIR を畳み込みヘッドホンから提示した. 使用した HRIR の種類は 0.25 ms, 0.5 ms, 1 ms, 2 ms, 全区間の 5 種類である. 目標上昇角は上半球正中面 0°-180°を 30°間隔で区切った 7 方向である. 刺激の種類は HRIR 5 種×目標方向 7 方向=35 種類である. 実験直前にヘッドホンから外耳道入口までの伝達関数を測定し, その逆特性を刺激に畳み込んだ[3]. 提示音圧は被験者の左右外耳道入口で 63dB SPL で提示時間は 1.2 s (立ち上り立ち下りに 0.1 s を含む)である. 回答方法はマッピング法である. 被験者は HRIR を測定した 5 名である. 回答回数は各刺激 10 回である.

#### 3.2 実験結果

Fig.3 に代表例として被験者 HRM の回答分布図を示す. 各パネルの横軸は目標上昇角, 縦軸は回答した上昇角である. 円の直径は回答の頻度に比例し, 回答は 5°間隔で丸めている. 全区間 HRIR では回答は概ね目標上昇角付近に分布した. 初期 HRIR(0.25ms)ではほとんどの回答が 180°付近に分布した. 初期 HRIR(0.5ms)では, 0°, 90°, 120°では目標方向と 180°付近の両方に回答が分布したが, 60°と 180°では全区間 HRIR と同様の分布をした. 初期 HRIR(1ms, 2ms)ではいずれの目標方向においても全区間 HRIR と同様の回答傾向であった.

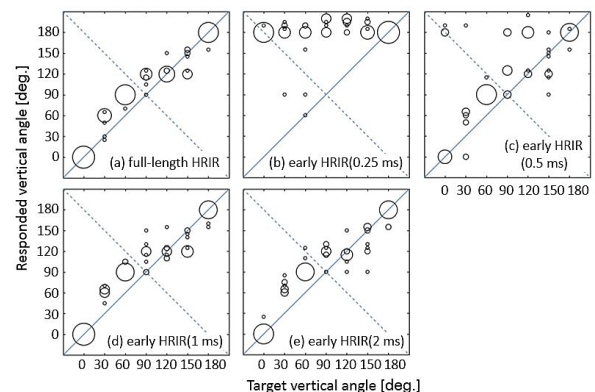


Fig. 3 Responses of subject HRM to the full-length HRIR and the early HRIRs.

被験者, HRIR の継続長, 目標上昇角ごとに平均上昇角定位誤差を算出した. ここで平均定位誤差とは, 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. さらに, 全区間 HRIR と初期 HRIR に統計的有意差があるか否かを Tukey

の多重比較検定によって求めた。結果を Table 1 に示す。

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。

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。

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ても、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。

Table 1 Mean vertical localization error for each subject, HRIR and target vertical angle [°].  
\*\*:  $p < 0.01$ , \*:  $p < 0.05$ .

| 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    | 25.6   | 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    |

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。

### 3.3 考察

被験者 HRM を除く 4 名の被験者では目標方向 120°においては初期 HRIR の継続長が 0.25ms, 0.5ms であっても平均定位誤差に全区間 HRIR との有意差は認められなかった。この点について考察する。

Fig.4 に各被験者の 120°における全区間 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(実線)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。

初期 HRTF(0.25ms)をみるといずれの被験者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 のもの同等となった理由であると考えられる。

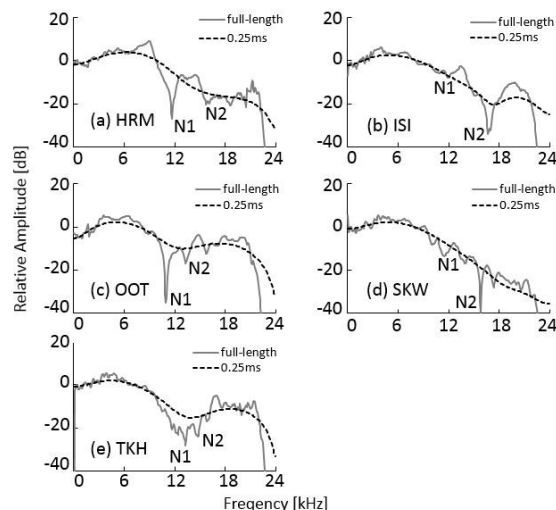


Fig.4 Amplitude spectra of the full-length HRTFs and the early HRTFs(0.25ms) for the vertical angle of 120°.

## 4 音像距離の知覚精度

平均上昇角定位誤差が全区間 HRIR と同程度であった初期 HRIR(1ms, 2ms)について音像距離も全区間 HRIR と同程度であるか否かを検証した。

#### 4.1 実験方法

被験者本人の全区間 HRIR および 2 種類の継続長(1ms, 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。刺激の種類は HRIR 3 種×目標上昇角 3 方向=9 種類である。2 つの刺激を対にして提示するので組み合わせ数は  $9P_2 = 72$  通りである。回答回数は各刺激対につき 5 回である。

#### 4.2 実験結果

Fig.5 に全被験者の回答から得た音像距離の尺度値を示す。図中のエラーバーは 95%信頼区間を示す。音像距離は HRIR の継続長よりも目標方向に強く影響を受け、前方、上方、後方になるにつれて近くなった。継続長の影響をみると、初期 HRIR の音像距離は全区間 HRIR と比べて 0°では近く、90°と 180°では遠くに知覚した。しかしその差は 95%信頼区間の範囲内であった。

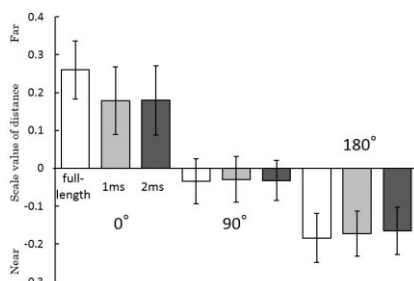


Fig.5 Scale value of the distance of a sound image. The error bars denote 95% confidence interval.

継続長および目標方向を要因とした 2 元配置分散分析を行った。その結果、HRIR の応答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。

目標方向の違いに主効果が認められたため、Tukey の多重比較検定を行った。その結果、全区間、1ms, 2ms いずれにおいても、0°, 90°, 180°それぞれ互いに有意差が認められた( $p<0.01$ )。ただし、1ms および 2ms における 90°と 180°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、初期 HRIR(1ms, 2ms)により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。

#### 5 まとめ

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) 1ms の初期 HRTF の振幅スペクトルは全区間 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 の平均定位誤差は多くの被験者及び目標方向において全区間 HRIR より大きくなった。1ms と 2ms の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 の音像距離は全区間 HRIR とほぼ同等であり、統計的に有意な差は認められなかった。

これらの結果は、初期 HRIR(1ms)は主要なノッチ・ピークの概形の情報を含んでいること、および上半球正中面において全区間 HRIR と同等の上昇角と距離を提供できることを示唆している。今後、初期 HRIR を使用することで受聴者の耳介形状に基づく 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。

#### 参考文献

- [1] Asano *et al.*, J Acoust Soc Am 88, 159-68, 1990
- [2] Kulkarni and Colburn., Nature 396, 747-49, 1998
- [3] Iida *et al.*, Applied Acoustics, 68, 835-50, 2007
- [4] Iida *et al.*, Applied Acoustics, 129, 239-47, 2018
- [5] Takemoto *et al.*, J. Acoust. Soc. Am. 132, 3832-41, 2012
- [6] 飯田他, 音講論(秋), 473-76, 2010
- [7] Møller, Applied Acoustics 36, 172-18, 1992