

头相关传输函数数据库及其特性分析 *

谢菠荪** 钟小丽 饶 丹 梁志强

(华南理工大学物理科学与技术学院声学研究所, 广州 510641)

摘要 对 52 名受试者(男、女各半)进行测量, 建立了中国人样本的高空间分辨率的头相关传输函数(HRTF)以及受试者生理尺寸的数据库. 基于此数据库, 对与声源定位有关的因素(双耳时间差、双耳声级差、耳廓产生的谱特征等)进行了分析. 结果表明, 在统计意义上最大双耳时间差存在显著的性别差异, 且与国外的数据也存在显著性差异, 这是由头部尺寸的统计差异引起的; 另外, 耳廓引起的 HRTF 高频谱特征是非常个性化的物理量; 在高频($f \geq 5.5$ kHz)HRTF 的左、右对称性将受到破坏. 此工作为今后有关双耳听觉的研究和虚拟听觉的应用提供了数据基础.

关键词 头相关传输函数 测量 双耳时间差 双耳听觉

头相关传输函数(head-related transfer function, HRTF)描述了自由场声波从声源到双耳的传输过程 [1]. 它反映了头部、耳廓和躯干等构成的生理系统对声波散射(综合滤波)的结果. HRTF 是声源方向、距离、频率的连续函数, 并且与生理结构和尺寸密切相关, 是具有明显个性化特征的物理量. 由于 HRTF 包含了有关声源定位的主要信息(如双耳时间差、双耳声级差等), 所以它在双耳听觉的研究和虚拟听觉的应用中具有非常重要的意义.

近十年来, 国外有多个课题组对 HRTF 进行了测量并建立了数据库 [2-8], 部分结果已公开, 可作为科学研究用途. 然而, 有关 HRTF 数据库及其研究还存在以下的问题:

(i) 文献 [2] 和 [3] 的 HRTF 是在人工头上测量得到的, 其中文献 [2] 对 KEMAR 人工头测量所得的数据目前应用最多. 但人工头是根据一定人群的平均生理尺寸做成的理想模型, 不能反映不同个体的个性化特性.

收稿日期: 2005-11-09; 接受日期: 2006-07-12

* 国家自然科学基金(批准号: 10374031)和广州市科技计划基金(批准号: 2005Z3-D0071)资助项目

** E-mail: phbsxie@scut.edu.cn

(ii) HRTF的测量是非常耗时的工作, 通常对真人受试者测量都做了一定的简化. HRTF是空间方向的连续函数, 而实际测量只能在有限的方向上进行(空间采样). 有许多数据库的空间分辨率偏低, 如文献 [4]和 [5]只分别测量了 120 和 144 个空间方向. 文献 [6]和 [7]以及互联网上公布的 IRCAM 数据库 (<http://recherche.ircam.fr/equipes/salles/listen/>)虽然是对全空间(至少是上半空间)进行测量, 但其在水平面的测量间隔分别是 22.5° , 10° 和 15° , 并且文献 [7]略去了一些高纬度(仰角)的测量. 利用这些HRTF数据, 通过插值不足以恢复空间方向上连续的HRTF^[9]. 对于真人受试者, 只有文献 [8]给出的是高空间方向分辨率的数据库, 包括 43 名受试者. 但是, 由于测量采用的是双耳极坐标系统, 侧向的空间分辨率还是不足, 特别是没有测量正侧向的HRTF数据.

(iii) 有些数据库的受试者样本太少, 如文献 [4]只有 12 名受试者, 不足以给出有效的统计结果. 而且, 现有的真人受试者数据库中, 受试者的性别比例差别较大. 如果某些参量的统计结果存在显著的性别差异, 那么整体的统计结果就会出现偏差.

(iv) 除文献 [8]外, 大多数研究没有给出双耳定位的物理因素与生理尺寸的统计关系.

(v) 最重要的一点, 目前还没有以中国人受试者为样本的 HRTF 数据库, 国内也未见有对声源定位相关的物理数据的测量工作. 现有的有关数据都是来自国外的测量, 主要是在西方人群中挑选受试者的结果. 一直以来国内的研究都是直接引用这些数据. 然而, HRTF 与受试者的生理尺寸密切相关; 不同民族的生理外形、尺寸是有一定差别的, 所以国外的数据不一定完全适用于中国人.

从 2003 年开始, 本课题组开展了中国人样本的高空间分辨率 HRTF 测量工作, 目前已建立了包括 52 名受试者的数据库, 并对其特性进行了分析. 本文将报道这方面的结果.

1 HRTF 的测量与数据库的建立

HRTF 是声源到双耳的频域传输函数. 自由场的情况下, HRTF 定义为

$$H_L = H_L(r, \theta, \phi, f) = \frac{P_L(r, \theta, \phi, f)}{P_0(r, f)}, \quad H_R = H_R(r, \theta, \phi, f) = \frac{P_R(r, \theta, \phi, f)}{P_0(r, f)}, \quad (1)$$

其中 P_L , P_R 分别是简谐点声源在左、右耳产生的复数声压; P_0 是头不存在时头中心位置处的复数声压; r 为声源到头中心的距离; f 为声波的频率; 方位角 $0^\circ \leq \theta < 360^\circ$ 和仰角 $-90^\circ \leq \phi \leq 90^\circ$ 表示声源的方向, 其中 $\phi = 0^\circ$ 和 90° 分别表示水平面和正上方, 而 $(\theta = 0^\circ, \phi = 0^\circ)$ 和 $(\theta = 90^\circ, \phi = 0^\circ)$ 分别表示水平面上正前和正右方向. 在远场 $r > 1.2$ (或 1.3) m, HRTF 基本与 r 无关. HRTF 的时域表示是头相关脉冲响应 $h_l(r, \theta, \phi, t)$, $h_r(r, \theta, \phi, t)$, 简记为 HRIR, 它们与 H_L , H_R 互为 Fourier 变换.

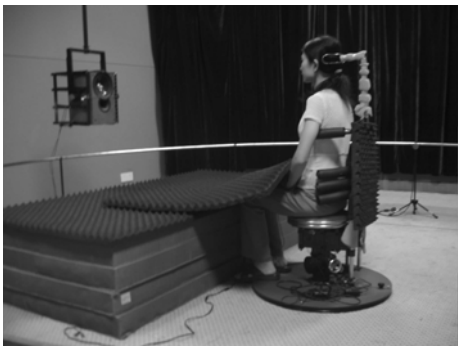


图 1 HRTF 测量环境图

测量在经过强吸声处理(混响时间 0.15 s)的听音室内进行(图 1), 本底噪声不大于 30 dB. 为了消除地面的反射声, 扬声器(声源)与被测对象之间的地面铺设 有 0.4 m 厚、容重由上至下递增的玻璃棉吸声材料, 同时天花也装上活动的吸声材料. 采用同轴扬声器(KEF-Q1)作为声源, 它距离头中心 1.5 m(远场测量). 受试者坐在特制的转椅上, 转椅的靠背和特殊的支撑可使受试者(特别是头部)固定. 为减少反射声, 转椅的靠背、支撑以及受试者的膝部都铺设 有吸声材料. 转椅的高度可以调节, 测量中受试者耳道入口的高度固定在 1.15 m. 通过调节固定扬声器的吊杆可改变声源的仰角 ϕ , 而通过转动转椅可改变受试者与声源之间的方位角 θ . 固定在转椅上的激光和固定在地面的半径 2.0 m 的圆坐标系(圆心与转椅的转轴重合)可准确指示出受试者与声源之间的方位角 θ .

分别在 9 个纬度面(仰角)上, 对均匀分布的 M 个方位角 θ 方向进行测量, 各纬度面的测量方位角数目 M 和测量间隔(分辨率) $\Delta\theta$ 如表 1. 我们最近的研究证明^[9], 采用表 1 给出的测量间隔, 可以通过空间插值的方法恢复各纬度面上对 θ 连续变化的 HRTF(空间采样定理). 低仰角($\phi < -45^\circ$)方向的 HRTF 测量较为困难, 并且与受试者的坐姿有很大的关系, 所以这里(以及国外现有的研究)不包括低仰角方向的测量.

表 1 各纬度面上的测量方位角数目 M 和测量间隔 $\Delta\theta$

$\phi/(^\circ)$	-30	-15	0	15	30	45	60	75	90
M	72	72	72	72	72	72	36	24	1
$\Delta\theta/(^\circ)$	5	5	5	5	5	5	10	15	360

HRTF 的测量原理和通常传输函数的测量类似. 这里采用最大长度序列法(maximum-length sequence, MLS)进行测量^[2], 图 2 是测量原理的方块图. 计算机先产生采样频率为 44.1 kHz、长度为 8191 点的 MLS 序列, 并采取 8 次重复平均的方法(可提高信噪比 9 dB), 此信号经声卡(Echo Layla 24)的 D/A 变换和功率放大器(B&K 2716C)后馈给扬声器. 测量时, 受试者移开后, 原头中心位置处的自由场声压约为 90 dB.

一对微缩传声器 (DPA 4060 High End Binaural Microphones)嵌入经过改装的

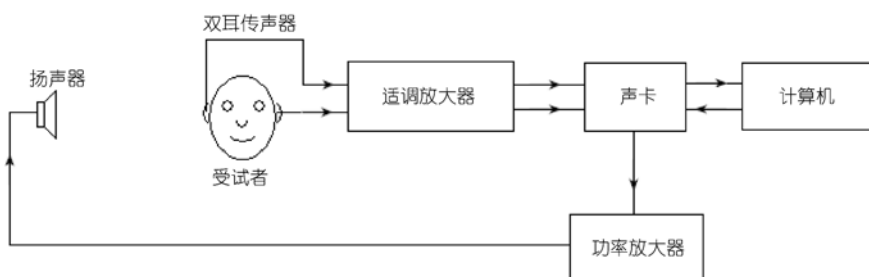


图 2 测量原理的方块图

游泳用耳塞, 并塞入被测对象的耳道入口处拾声信号 (图 3), 也就是采用 HRTF 的封闭耳道测量法。Møller 及其他的一些作者已经证明^[10], 该方法可以完全保留 HRTF 中有关声源的空间方向信息。传声器的输出经 B&K 2690A 0S4 适调放大器放大, 再经声卡 A/D 变换后输入到计算机。经用 MATLAB 语言自行编制的软件作去卷积运算, 得到原始的时域 8191 点的 HRIR。原始 HRIR 经过矩形时间窗截取 (用以消除房间反射声), 并对扬声器和传声器系统的传输特性均衡后, 最终得到相应声源方向的 512 点 HRIR。对它进行离散 Fourier 变换可以得到 512 点 HRTF。

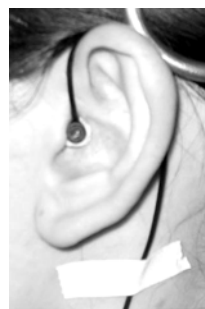


图 3 封闭耳道测量法

我们对男、女性各 26 名受试者进行了测量。受试者在华南理工大学的教师、研究生和本科生中选取, 年龄 18~45 岁。所建立的数据库共有 25636 对数据, 包括 52 名受试者每人 493 个空间方向 (表 1)。数据以时域 HRIR 的形式存储。每对 HRIR 数据存储为 44.1 kHz 采样, 16 bit 量化、长度 512 点的双通路数据 (pcm) 文件 (左、右耳各一通路)。

2 HRTF 数据的分析

对得到的 HRTF 数据进行分析, 可得出与声源定位有关的物理量的一些统计规律。

2.1 HRTF 的基本特性

虽然 HRTF 是具有个性化特征的物理量, 但在一些基本特性上是共同的。图 4 给出了一名典型的受试者 (编号 25) 在水平面 $\phi = 0^\circ$ 、左耳 (右耳也有类似的结果)、不同方位角 θ 的时域 HRIR 和相应的频域 HRTF。注意图 4(a) 和 4(b) 分别采用了线性和对数的振幅坐标。对时域 HRIR, 开始约 55~85 个采样值近似为零, 这对应声源到双耳的传播延时 (已用时间窗截取)。脉冲响应的主体部分长度约为 50 个采样, 反映了入射声波与头部、耳廓和躯干等的复杂相互作用, 然后脉冲响应近似

回复到零. 当声源处于左耳的异侧时($\theta = 60^\circ \sim 120^\circ$), 脉冲的幅度明显降低, 这是头部对声波的阴影作用所致. 而频域HRTF幅值表明, 在低频($f \leq 0.5$ kHz), 头部等的散射作用可以略去, $|H_L|$ 基本与频率无关. 但随着频率的增加, $|H_L|$ 表现出与 f 和 θ 复杂的函数关系, 这是头部、耳廓和躯干等的综合作用的结果. 而头部的阴影作用使得在 $f > 3 \sim 4$ kHz的高频, 声源处于耳的异侧时, HRTF幅值明显下降. 以上结论对其他 51 名受试者也基本适用(至少定性上), 并且和国外有关的研究结果是一致的 [2,6,7].

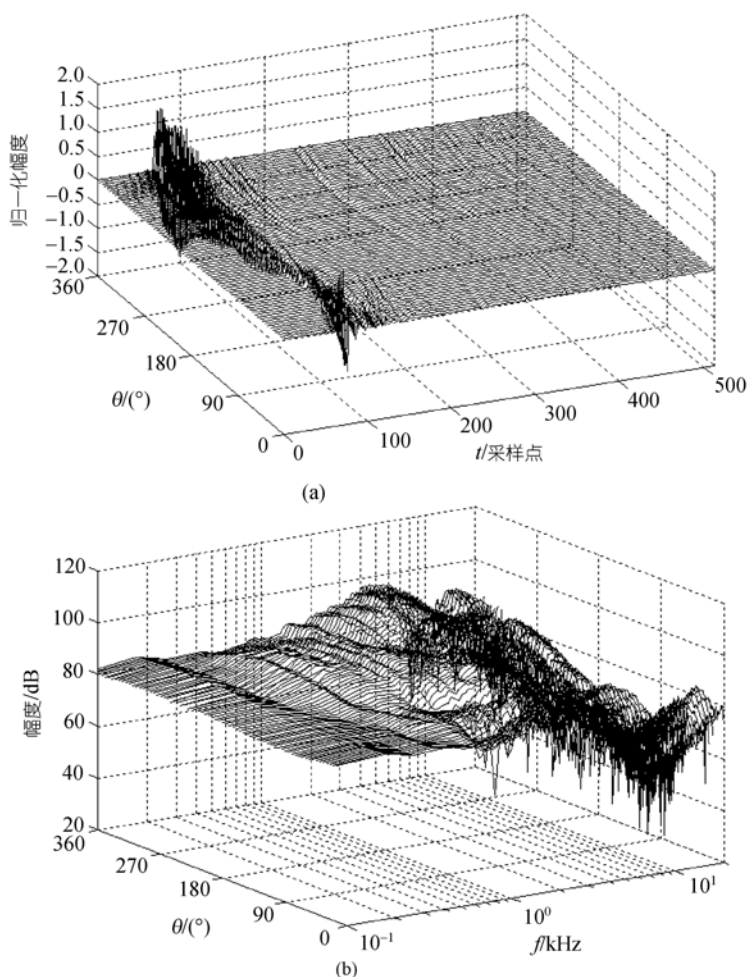


图 4 典型受试者的水平面 HRIR 和 HRTF(左耳)

(a) 时域 HRIR, (b) 频域 HRTF

2.2 双耳时间差(ITD)

双耳声压的时间差(interaural time difference, ITD)是声源定位的一个主要因

素. 在 $f < 1.5$ kHz 的低频, 双耳声压的相延时是定位的主要因素; 而在 $f > 1.5$ kHz, 双耳声压的包络延时可作为一个定位的因素^[11]. 过去国内的研究采用的是国外测量得到的 ITD 数据, 其适用性也有待验证. 利用 HRTF 数据, 可计算出 ITD. 这里采用相关法计算 ITD. 设声源在左、右耳产生的 HRIR 分别为 $h_l(\theta, \phi, t)$, $h_r(\theta, \phi, t)$, 它们的归一化互相关函数为

$$\Psi_{lr}(\tau) = \frac{\int h_l(\theta, \phi, t + \tau) h_r(\theta, \phi, t) dt}{\sqrt{\int h_l^2(\theta, \phi, t) dt \int h_r^2(\theta, \phi, t) dt}}. \quad (2)$$

由(2)式可计算出 $|\Psi_{lr}(\tau)|$ 在 $|\tau| \leq 1$ ms 范围内的最大值, 与此相应的 $\tau = \tau_{\max}$ 即为双耳声压的时间差 ITD. 由于受耳廓等的影响, 不同受试者的 HRTF 高频部分变化很大, 同时考虑到 ITD 主要是作为中、低频的定位因素, 因而在(2)式的计算中, 先对 $h_l(\theta, \phi, t)$ 和 $h_r(\theta, \phi, t)$ 进行了 $f < 2.7$ kHz 的低通滤波处理.

图 5(a)和 5(b)分别给出了 26 名男性和 26 名女性受试者的平均 ITD 曲线, 其中 $\text{ITD} > 0$ 表示右耳超前, $\text{ITD} < 0$ 表示右耳落后. 可以看出, 不同性别受试者的平均 ITD 随空间方向的变化规律是类似的. 对确定的纬度面 ϕ , 右半平面 θ 方向 ($0^\circ < \theta < 180^\circ$) 与左半平面 ($360^\circ - \theta$) 方向的 ITD 的绝对值近似相等, 特别是在中垂面 $\theta = 0^\circ$ 或 180° 上, $\text{ITD} \approx 0$. 这是由于在 $f < 2.7$ kHz 的中、低频情况下, 声波波长相对较长, 生理结构细微不对称的影响较少. 随着声源接近侧向, $|\text{ITD}|$ 增加, 在侧向 $\theta = 90^\circ$ 和 270° 附近达到最大. 对比不同的纬度面, $\phi = 0^\circ$ 的 ITD 变化幅度最大, 而随着仰角 ϕ 偏离水平面, ITD 的变化幅度将减少. 这些结果基本上和现有文献是一致的^[6~8].

如果对每名受试者的 493 个不同空间方向的 ITD 数据进行分析, 可得到 $|\text{ITD}|_{\max}$ 及其对应的空间声源方向 ($\theta_{\max}, \phi_{\max}$), 其结果特别值得注意:

(i) 对于 52 名男、女性受试者的绝大多数 (47 名), $\phi_{\max} = 0^\circ$ (水平面), $\theta_{\max}$ 在 $90^\circ \pm 10^\circ$ (或 $270^\circ \pm 10^\circ$) 的范围内, 其中有 39 名受试者的 $\theta_{\max} = 90^\circ \sim 95^\circ$ (或 $265^\circ \sim 270^\circ$). 另有少数 (5 名) 受试者的 $\phi_{\max} = 15^\circ$ (比水平面略高的纬度), $\theta_{\max} = 90^\circ \sim 95^\circ$ (或 $265^\circ \sim 270^\circ$). 但对于这些受试者, 在侧向 $\theta = 90^\circ$ (或 270°) 附近, $\phi = 0^\circ$ 和 15° 两个纬度面的 ITD 非常接近 (差别 $< 10 \mu\text{s}$). 因而可以认为在水平面正侧向附近 $|\text{ITD}|$ 达到其最大值, 少数的反常是由实验误差或生理结构引起的. 对 52 名男、女性受试者的统计表明, $\theta_{\max}$ 的平均值为 91.9° (标准差 3.6°), 或平均值为 267.9° (标准差 3.9°), 这与现有的一些结果不同, 第 3 节还会讨论这个问题.

(ii) 分别对男性和女性受试者的 $|\text{ITD}|_{\max}$ 进行统计, 结果如表 2. 进一步用数理统计的方差分析方法分析表明 ($F = 43.6, p = 2.54 \times 10^{-8}$), 在 0.05 的水平下, 男、女性受试者的平均 $|\text{ITD}|_{\max}$ 存在显著性差别, 男性受试者的较大. 这是由生理尺寸的差别引起的. 我们的研究指出, 采用不适合倾听者生理尺寸的 HRTF 进行虚

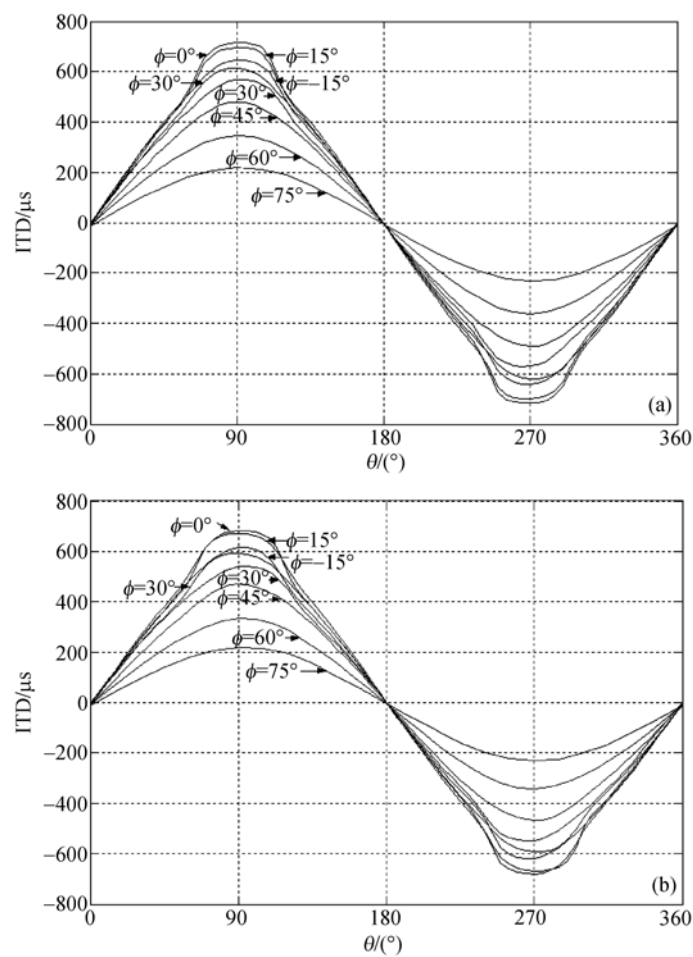


图 5 不同纬度面 ϕ 的 ITD 与方位角 θ 的关系
(a) 26 名男性受试者的平均 ITD, (b) 26 名女性受试者的平均 ITD

表 2 中国人受试者最大 ITD 的统计结果

性别	人数/名	最大 $ \text{ITD} _{\text{max}}$ / μs	最小 $ \text{ITD} _{\text{max}}$ / μs	平均 $ \text{ITD} _{\text{max}}$ / μs	标准差 σ / μs
男	26	761	671	722	22
女	26	734	659	686	17

拟声信号处理, 会对主观重放效果有明显的影响^[12]. 事实上, 图 5 表明, 在侧向 ITD 随 θ 变化较慢, 或反过来, 小的 ITD 改变将对应较大的 θ 变化. 以上的 $|\text{ITD}|_{\text{max}}$ 的差别, 足以引起虚拟声重放中明显的侧向声像畸变. 因而用统计方法建立双耳听觉模型的时候应考虑性别上的差别, 理想的情况应分别建立适合男性和女性的模型. 如果必须建立通用的模型, 则至少要选择等量的男、女性受试者

的数据进行统计平均, 否则就会出现偏差. 但现有的研究很少注意到这一点.

2.3 双耳声级差

当 $f > 1.5$ kHz, 双耳声压的声级差(interaural level difference, ILD)是定位的另一个主要因素. 按(1)式, 可直接利用 HRTF 计算出 ILD 作为声源方向和频率的函数. 但高频时 ILD 随频率变化很快, 特别是由于生理结构和尺寸(主要是耳廓)的差别, 当 $f > 5 \sim 6$ kHz 时, 在特定频率上不同受试者的 ILD 差别较大, 因而不适宜直接在特定的频率上计算所有受试者的平均 ILD, 应该对 HRTF 进行某种频率上的平滑处理后再计算.

心理声学 and 生理声学的研究指出, 人类的听觉系统可看成为一系列的带通滤波器, 称为临界频带听觉模型. 如果用矩形的滤波器等效, 则称为等效的矩形听觉滤波器. 这些滤波器对应于基底膜上不同的位置. 声音的响度感觉、掩蔽等心理声学现象都与听觉滤波器密切相关. 滤波器的带宽 ERB(单位 Hz)与其中心频率 f (单位 kHz)由以下公式决定^[13]

$$\text{ERB} = 24.7(4.37f + 1). \quad (3)$$

利用等效矩形听觉滤波器的概念, 可引入一个新的频率标度——ERB 数(number of ERB, 简记为 N), 它与频率 f (单位 kHz)的关系为

$$N = 21.4 \log_{10}(4.37f + 1). \quad (4)$$

对特定受试者和特定方向的每一对 HRTF, 可分别在 ERB 的带宽内求出其能量(这也符合听觉的信号处理模式), 然后再算出每个频带的双耳声级差作为方向和以 N 标度的频率的函数(其中的积分在各 ERB 内进行)

$$\text{ILD}(\theta, \phi, N) = 10 \log_{10} \frac{\int_{\text{ERB}} |H_L(\theta, \phi, f)|^2 df}{\int_{\text{ERB}} |H_R(\theta, \phi, f)|^2 df} \quad (\text{dB}). \quad (5)$$

最后对 52 名受试者取平均, 得到平均的双耳声级差 ILD_{ave} . 图 6 是水平面 $\phi = 0^\circ$ 的 ILD_{ave} .

对正前和正后方($\theta = 0^\circ$ 和 180°), 在 $N \leq 31$ ($f \leq 6.2$ kHz) 频率范围内, $|\text{ILD}_{\text{ave}}| \approx 0$ dB (≤ 1.0 dB); 对 $N \geq 32$, $|\text{ILD}_{\text{ave}}|$ 有所增加, 但也不超过 4.5 dB. 事实上, 虽然从每名受试者的数据来看, HRTF 并不一定是严格左、右对称的, 中垂面上正前和正后方声源所产生的 ILD 并不一定严格为零, 但对 52 名受试者的 ILD 取平均后, 这种左、右不对称的影响将部分被消去, 使得 ILD_{ave} 接近 0 dB. 高频 $|\text{ILD}_{\text{ave}}|$ 的增加是由于生理结构上的细微不对称以及实验误差所引起. 可以证明, 对于中垂面其他方向的 HRTF 也有类似的结论.

当水平面的声源偏离正前(后)方时, ILD_{ave} 表现出复杂的与频率的关系; 且当

声源靠近两侧时, 由于头部的阴影作用, $|\text{ILD}_{\text{ave}}|$ 增加. 但对于低频 $N \leq 8$ ($f \leq 0.3$ kHz), $|\text{ILD}_{\text{ave}}|$ 不大于 4.7 dB.

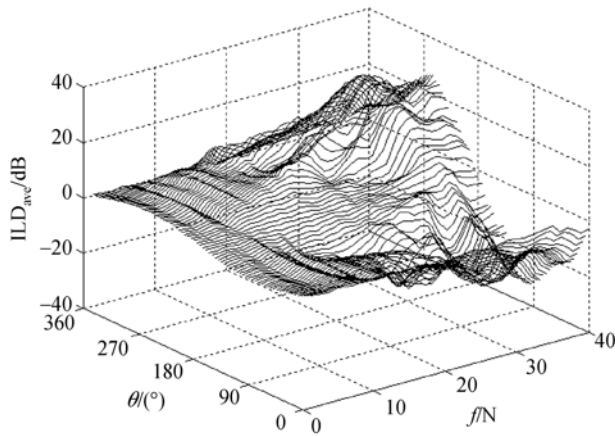


图 6 水平面上 52 名受试者的 ILD_{ave}

2.4 耳廓引起的高频谱特征

在 $f > 5$ (或 6) kHz, 耳廓的尺度与波长可以比拟, 它产生的反射或散射声与直接传到耳道口的声波迭加. 当迭加的声波反相时, 将在 HRTF 的幅度谱上产生谷点. 谷点的频率与声源的方向、耳廓的生理结构有关. 耳廓反射所产生的谷点对中垂面上的声源定位有重要的作用 [14].

图 7 给出了一名受试者(编号 25)在中垂面 $\theta = 0^\circ$ 上, 仰角 $\phi = -30^\circ, 0^\circ, 30^\circ$ 的 HRTF 幅度谱曲线. 在 6~10 kHz 频率范围内, 幅度谱上有一个谷点, 且左、右耳的谷点频率大致相等(差别小于 0.3 kHz). 随着仰角的上升, 谷点的频率也上升, $\phi = -30^\circ, 0^\circ, 30^\circ$ 对应的右耳谷点频率分别为 6.5, 8.1 和 8.8 kHz. 在高的仰角 $\phi \geq 60^\circ$, 谷点逐渐变得不明显, 这是由于不同仰角的入射声由耳廓不同部位反射, 与直达声声程差不同所致. 这和文献 [15] 的结果一致, 并且许多研究把这看成是普遍的规律.

但值得注意, 不同受试者的结果相差很大. 部分受试者的定性结果和图 7 类似, 但定量结果(包括谷点的频率和深度)不同. 也有部分受试者左、右耳的谷点频率不相等, 如一名受试者(编号 4)在 $(\phi = 0^\circ, \theta = 0^\circ)$ 的左、右耳的谷点的频率分别为 8.6 和 9.3 kHz. 这是由于左右耳廓在细微生理结构上的不对称引起的, 对这类受试者, 左右对称的耳廓模型将不适用. 更有小部分受试者的谷点并不明显, 甚至几乎完全消失, 这是由于耳廓形状不同引起的. 当耳廓不能很好地将反射声聚焦到耳道入口时, 直达声与反射声的干涉效果将减弱.

从 52 名受试者的 HRTF 结果来看, 没有两名(或两名以上)受试者的耳廓引起

的频谱特征是完全相同的, 且没有特别的规律可循. 因而耳廓对HRTF(以及听觉)产生的影响是最具有个性化特征的, 对其计算统计平均结果意义不大. 这也就不难解释在虚拟听觉的应用中, 采用非个性化HRTF进行信号处理所带来的声像畸变和重放效果因人而异的问题. 更值得指出的是, 有研究提出在采用非个性化HRTF进行虚拟声信号处理时, 人为地加深HRTF幅度谱上的谷点, 从而减少非个性化HRTF函数带来的声像畸变^[16]. 但从上面的分析可以看出, 不同受试者的HRTF的差别不只是幅度谱上的谷点深度, 谷点频率及其他特征也不相同. 因而加深HRTF幅度谱上谷点的方法是值得商讨的.

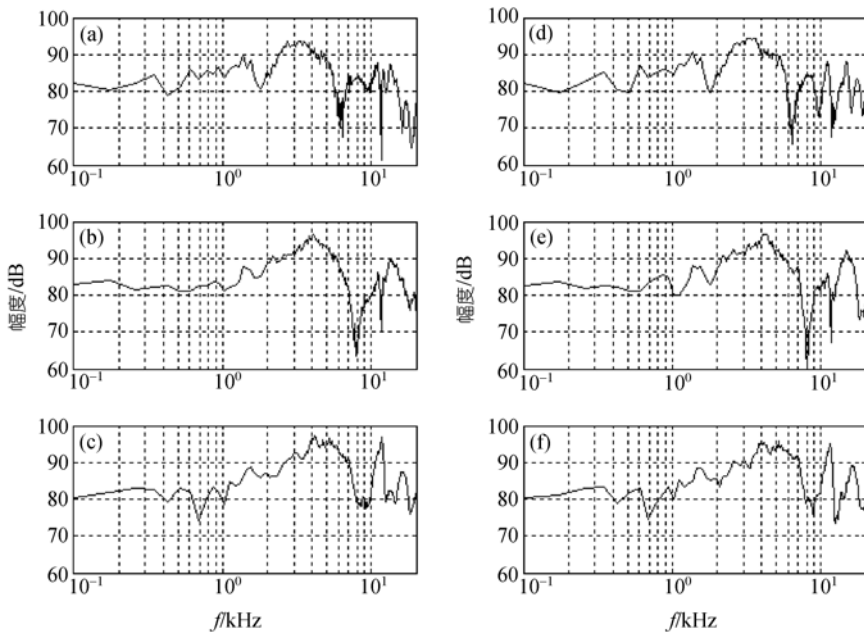


图7 中垂面 $\theta = 0^\circ$ 上, 不同仰角 HRTF 幅度谱曲线

(a)~(c)依次为左耳 $\phi = -30^\circ, 0^\circ, 30^\circ$ 的情况; (d), (e)和(f)依次为右耳 $\phi = -30^\circ, 0^\circ$ 和 30° 的情况

2.5 HRTF 的左右对称性

过去许多研究和实际应用中假定 HRTF 是左、右对称的, 即左、右耳之间的 HRTF 满足关系 $H_R(\theta, \phi, f) \approx H_L(360^\circ - \theta, \phi, f)$. 从 2.3 节的讨论, 在高频 HRTF 可能会出现左、右不对称. 为了检验 HRTF 左、右对称假设的适用范围, 在水平面 $\phi = 0^\circ$ 定义各 ERB 频带范围内的 HRTF 左右能量比值

$$\Delta(N) = \frac{1}{72} \sum_{\theta=0^\circ}^{355^\circ} \left| 10 \log_{10} \frac{\int_{\text{ERB}} |H_R(\theta, 0^\circ, f)|^2 df}{\int_{\text{ERB}} |H_L(360^\circ - \theta, 0^\circ, f)|^2 df} \right| \quad (\text{dB}). \quad (6)$$

上式右面的分子表示与 θ 方向声源对应的右耳 HRTF 在频率标度为 N 的 ERB 频带内的能量, 分母表示与 $(360^\circ - \theta)$ 方向的声源对应的左耳 HRTF 在标度为 N 的 ERB 频带范围内的能量, 绝对值号内的计算求出以 dB 为单位的能量比值, 求和号及除以 72 表示对水平面 72 个 θ 方向取平均. 在左、右对称的情况下 $\Delta = 0$ dB, $\Delta > 0$ dB 表示左、右不对称. Δ 只考虑了 HRTF 在能量(振幅)上的左、右对称性, 没有考虑 HRTF 相位上的左、右对称性. 主要是考虑测量过程中受试者的微小移动带来的主要是相位误差, (6)式可排除这种误差.

用每名受试者的数据分别算出 $\Delta(N)$, 并对 52 名受试者进行统计平均, 得到 $\Delta_{\text{ave}}(N)$ 结果, 如图 8(注意, 对每名受试者, 都有 $\Delta \geq 0$ dB, 因而对 52 名受试者进行统计平均不会导致的各受试者的贡献相抵消). 在低频 $N \leq 16$ ($f \leq 1$ kHz), $\Delta_{\text{ave}}(N) \leq 1$ dB, 左、右对称是一个非常好的近似. 但随着频率的增加, 当 $N \geq 30$ ($f \geq 5.5$ kHz), $\Delta_{\text{ave}}(N)$ 明显增加. 这是由于在高频耳廓和头部细微结构对声波的作用逐渐明显(当然也包括一定的实验误差), 它们的左、右不对称将导致 HRTF 的左、右不对称. 值得指出的是, 不同受试者的 HRTF 左、右不对称的起始频率和程度是不同的, 即使对同一受试者, 对不同的方向 θ , 结果也不相同. 这里给出的是对水平面 72 个方向和 52 名受试者的平均结果, 但已足以说明在高频 HRTF 的左、右对称性将受到破坏. 但目前有关 HRTF 和虚拟声的研究没有注意到这个问题, 经常沿用 HRTF 左、右对称的假设.

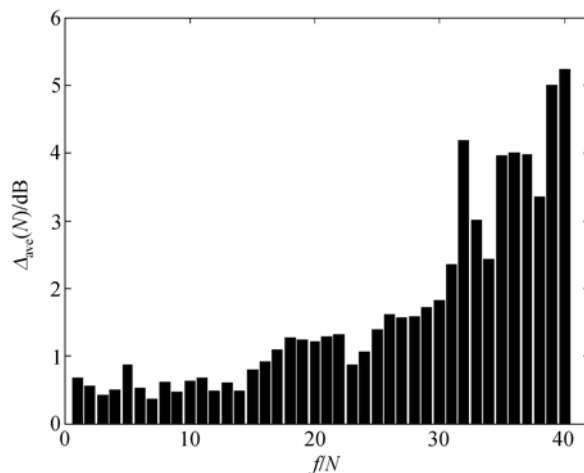


图 8 52 名受试者的 $\Delta_{\text{ave}}(N)$

3 与现有的 HRTF 数据比较

引言部分指出, 在现有的真人受试者 HRTF 数据库中, 只有文献 [8] 给出的

CIPIC数据库是高空间方向分辨率且包括较多受试者样本(共 43 人, 27 男, 16 女). 因而这里侧重将我们的结果和CIPIC数据库进行比较.

文献 [8]和 [17]给出了CIPIC数据库 43 名受试者的最大ITD的统计结果: 平均 $|\text{ITD}|_{\max} = 646 \mu\text{s}$, 标准差 $\sigma = 33 \mu\text{s}$. 但该统计结果存在以下问题:

(i) 文献 [8]和 [17]是通过检查左、右耳HRIR的起始时间之差来求ITD的. 然而, 过去有关的研究指出 [18], HRTF可表示为最小相位函数 $H_{\min}(\theta, \phi, f)$ 、全通相位函数 $\exp[j\psi_{\text{all}}(\theta, \phi, f)]$ 以及线性相位函数 $\exp[-j2\pi fT(\theta, \phi)]$ 的乘积(其中 $T(\theta, \phi)$ 表示声波从声源到耳的传播距离引起的纯延时, 对应于图 4(a)HRIR的起始零值):

$$H(\theta, \phi, f) = H_{\min}(\theta, \phi, f) \exp[j\psi_{\text{all}}(\theta, \phi, f)] \exp[-j2\pi fT(\theta, \phi)]. \quad (7)$$

利用 HRIR 起始时间之差求 ITD 的方法只计及线性相位函数项对 ITD 的贡献, 而略去了最小相位函数项和全通相位函数项的贡献, 所求出的 ITD 较其他方法的小.

(ii) 由于文献 [8]的HRTF测量时采用的是双耳极坐标系统, 侧向的空间方向分辨率还是不足, 特别是没有测量正侧向的HRTF数据, 而最大ITD经常出现在侧向, 所以它所得到的最大ITD不一定准确.

(iii) 文献 [8]的统计中没有考虑性别的差异, 且男、女性受试者的人数不等. 为了比较, 我们利用(2)式的相关法重新估算了 CIPIC 数据的最大 ITD. 首先利用每名受试者的 HRTF 数据计算所有已测方向的 ITD, 并找出其中 $|\text{ITD}|$ 的最大值, 作为该受试者最大 ITD 的下限(实际的 $|\text{ITD}|_{\max}$ 一定大于或等于此下限). 分别对男、女受试者的结果进行统计, 结果如表 3. 进一步用数理统计的方差分析方法分析表明($F = 17.9, p = 1.26 \times 10^{-4}$), 在 0.05 的水平下, CIPIC 数据的男、女性受试者的 $|\text{ITD}|_{\max}$ 下限也存在显著性差异. 这进一步说明了按不同性别进行统计平均的重要性.

表 3 CIPIC HRTF 数据库的 $|\text{ITD}|_{\max}$ 统计结果

性别	人数/名	最大 $ \text{ITD} _{\max}$ / μs	最小 $ \text{ITD} _{\max}$ / μs	平均 $ \text{ITD} _{\max}$ / μs	标准差 σ / μs
男	27	≥ 814	≥ 706	≥ 748	27
女	16	≥ 758	≥ 681	≥ 714	24

与表 2 给出的统计结果比较, 并采用方差分析方法分别对男、女性受试者的结果进行分析, 结果表明, 在 0.05 的水平下(对男性受试者数据, $F = 15.2, p = 2.88 \times 10^{-4}$; 对女性受试者数据, $F = 20.4, p = 5.46 \times 10^{-5}$), 中国人样本的平均 $|\text{ITD}|_{\max}$ 与 CIPIC 数据库的统计结果存在显著性差异, 中国人样本的平均 $|\text{ITD}|_{\max}$ 相对小些. 这是由不同种族生理尺寸的统计差异引起的. 值得注意的是, CIPIC 数据库已包含了少量的亚裔受试者, 如果全采用西方人受试者进行统计, 差别可能会更大.

因此, 这里的结果已充分说明, 建立基于中国人样本的 HRTF 数据库的必要性, 直接引用国外数据进行中国人双耳听觉的研究可能会出现偏差.

由于 CIPIC 数据库的 HRTF 测量采用的是双耳极坐标系统, 因而无法计算出对应最大 ITD 的空间方向、水平面上的双耳声级差 ILD 以及左、右对称系数, 因而也就无法和我们的结果比较. 但文献 [7] 给出的(采用另一 HRTF 数据库)结果表明, 最大 ITD 出现在水平面上 $\theta_{\max} = 110^\circ \sim 120^\circ$ 或 $250^\circ \sim 260^\circ$, 也就是两侧偏后的方向, 并认为这是由于双耳位于头部两侧偏后的位置所致. 而 2.2 节的结果表明, 采用我们 HRTF 数据库的结果是 $\theta_{\max}$ 在 90° 或 270° 附近. 这是由于 ITD 的计算方法不同所致. 虽然文献 [7] 也是采用(2)式的相关法计算 ITD, 但所用的是全频带的 HRIR, 包含耳廓等高频不确定因素. 具体计算表明, 如果将头部简化为一刚性球体, 双耳简化为球面上两侧偏后(如水平面 $\theta = 100^\circ, 260^\circ$)的两点, 利用计算得到的 HRTF 数据和(2)式可以得到, 最大 ITD 依然是在水平面 $\theta = 90^\circ$ 和 270° 附近(但稍为偏前), 与双耳位于球面上两侧($\theta = 90^\circ$ 和 270°)的结果差别很少.

4 与受试者生理数据的关系

为了探讨双耳时间差等声源定位因素与受试者生理尺寸的关系, 对每名受试者头面部的部分生理尺寸进行了测量. 在文献 [8] 的研究中, 对 17 项生理尺寸进行了测量, 统计结果表明只有其中的三个参数, 即头高、头宽、头深, 与 HRTF 中的一些声源定位因素有一定的相关性. 生理尺寸参数和测量位置是密切相关的, 如在头部不同位置测量头宽可得到差别很大的结果, 但是文献 [8] 并没有给出头宽测量的位置. 国内已制定了人体生理尺寸的测量方法, 并对成年人的头面部尺寸进行了测量, 公布了相应的国家标准¹⁾. 按 GB/T 2428-1998 标准, 我们测量了受试者的头全高 x_1 、最大头宽 x_2 、两耳屏间宽 x_3 三项参数, 另参照文献 [8] 增加测量了头深 x_4 . 表 4 给出了我们测量结果, 并和相关的结果进行了比较.

在我们的测量结果中, 男性受试者的平均头全高 x_1 、最大头宽 x_2 、两耳屏间宽 x_3 、头深 x_4 都较女性受试者的大. 进一步用数理统计的方差分析方法分析表明, 在 0.05 的水平下, 男、女性受试者的四项参数存在显著性差别(例如对参数 x_2 , $F = 14.95$, $p = 3.19 \times 10^{-4}$). 这一点是和 GB/T 2428-1998 的结果类似的. 在 GB 的结果中, 男性受试者的平均 x_1, x_2 也较女性受试者的大, 但男、女性受试者的平均 x_3 却相等. 值得注意的是, 我们所得到的男性平均 x_1, x_3 , 男性和女性的平均 x_2 都较 GB 的结果大(虽然存在统计误差). 这可能是由于 GB/T 2428-1998 是以 1988 年测量的数据为基础, 至今已有 17 年. 而人体生理尺寸数据是有一定时效性的. 事实上, 如果将我们的测量结果和 1999 年前后对山东汉族大学生的测量结果比较^[19], 可

1) GB/T 2428-1998, 成年人头面部尺寸, 中华人民共和国国家标准

以发现它们之间是比较吻合的.

表 4 生理尺寸参数的测量结果

数据来源	头全高 x_1/mm	最大头宽 x_2/mm	两耳屏间宽 x_3/mm	头深 x_4/mm
男性测量平均值与标准差	231 (9.1)	160 (6.6)	145 (5.1)	188 (6.8)
女性测量平均值与标准差	213 (11.0)	153 (5.3)	138 (5.0)	180 (7.0)
GB 男性平均值与标准差	223 (10.68)	154 (5.86)	140 (3.99)	
GB 女性平均值与标准差	216 (9.77)	149 (5.30)	140 (3.75)	
文献 [19]男性平均值与标准差	231.3(10.8)	161.3 (5.5)	149.8 (5.3)	
文献 [19]女性平均值与标准差	221.1(10.4)	156.6 (5.5)	140.2 (6.0)	

如果对 2.2 节得到的全部 52 名受试者的最大双耳时间差 $|\text{ITD}|_{\max}$ 与生理尺寸参数进行相关分析, 计算得到的归一化相关系数如表 5. 可以看出, $|\text{ITD}|_{\max}$ 与最大头宽 x_2 以及两耳屏间宽 x_3 都有较大相关性, 与头全高 x_1 有一定的相关性, 与头深 x_4 的相关性较小. 但 4 个生理尺寸参数并非完全独立, 例如 x_1 与 x_2 , x_2 与 x_3 之间也有一定的相关性.

表 5 $|\text{ITD}|_{\max}$ 与生理尺寸参数之间的相关系数

	头全高 x_1	最大头宽 x_2	两耳屏间宽 x_3	头深 x_4
最大头宽 x_2	0.516			
两耳屏间宽 x_3	0.373	0.544		
头深 x_4	0.367	0.246	0.290	
最大双耳时间差 $ \text{ITD} _{\max}$	0.580	0.741	0.748	0.435

综合上面讨论可以得出结论, 最大双耳时间差 $|\text{ITD}|_{\max}$ 与最大头宽(或两耳屏间宽)等生理尺寸参数有关, 头部平均生理尺寸的统计差异是造成平均 $|\text{ITD}|_{\max}$ 存在性别差异的重要原因. 同样, 这也是不同种族的平均 $|\text{ITD}|_{\max}$ 存在统计差异的原因(不同种族的平均生理尺寸存在统计差异是熟知的). 因此, 这进一步说明建立中国人样本 HRTF 数据库是必要的.

5 结论

对 52 名受试者(男、女各 26 名)进行测量, 建立了基于中国人样本、高空间分辨率的头相关传输函数以及受试者生理尺寸参数的数据库. 对数据库进行分析, 得到了与声源定位有关的物理量和生理尺寸参数之间的统计规律. 结果表明:

(i) 对于不同的纬度面, 水平面的 ITD 随 θ 的变化幅度最大, 而随着仰角 ϕ 的上升或下降, ITD 随 θ 的变化幅度都减少. 对于绝大多数受试者, 当声源位于水平面 $\theta = 90^\circ \sim 95^\circ$ (或 $265^\circ \sim 270^\circ$) 时, 它所产生的 $|\text{ITD}|$ 最大.

(ii) 最大双耳时间差的统计平均值 $|\text{ITD}|_{\max}$ 存在着显著的性别差异, 因而应分别建立适合男性和女性的模型. 如果必须建立通用的模型, 则至少要选择等量

的男、女性受试者的数据进行统计平均, 否则就会出现偏差. 但现有的研究很少注意到这一点.

(iii) 中国人样本的平均 $|ITD|_{\max}$ 与国外(CIPIC)数据库的统计结果存在着显著性差异. 因此建立中国人样本的 HRTF 数据库是必要的, 直接引用国外数据进行中国人双耳听觉的研究可能会出现偏差.

(iv) 头全高、最大头宽、两耳屏间宽、头深等生理尺寸参数的统计平均值存在着显著的性别差异, 而最大双耳时间差 $|ITD|_{\max}$ 与最大头宽或两耳屏间宽有高的相关性, 因而生理尺寸性别上的统计差异是导致 $|ITD|_{\max}$ 差异的原因.

(v) 耳廓产生的 HRTF 高频谱特征是非常个性化的物理量, 不同受试者的差别很大.

(vi) 在低频 HRTF 近似左、右对称. 但随着频率的增加, 在 $f \geq 5.5$ kHz, 由于耳廓、头部外形等在细微结构上的不对称的影响, HRTF 左、右对称的假设将不成立.

(vii) 本文工作作为国内今后开展双耳听觉的基础研究和虚拟声方面的应用提供了数据基础.

致谢 作者感谢北京邮电大学管善群教授的指导, 感谢华南理工大学谢志文、同济大学毛东兴、俞悟周三位副教授的帮助, 同时感谢协助和参与实验的华南理工大学的教师、研究生和本科生.

参 考 文 献

- 1 钟小丽, 谢波菰. 头相关传输函数的研究进展(一). 电声技术, 2004, 12: 44—46
- 2 Gardner W G, Martin K D. HRTF measurements of a KEMAR. J Acous Soc Am, 1995, 97(6): 3907—3908[DOI]
- 3 Bovbjerg B P, Christensen F, Minnaar P, et al. Measuring the head related transfer functions of an artificial head with a high directional resolution. Presented at AES 109th Convention. USA: Los Angeles, California, 2000. 5264
- 4 Blauert J, Brueggen C, Bronkhorst A W, et al. The AUDIS catalog of human HRTFs. J Acoust Soc Am, 1998, 103 (5): 3082[DOI]
- 5 Wightman F L, Kistler D J. Headphone simulation of free-field listening, I: stimulus synthesis. J Acoust Soc Am, 1989, 85(2): 858—867[DOI]
- 6 Møller H, Sørensen M F, Hammershøj D, et al. Head-related transfer functions of human subjects. J Audio Eng Soc, 1995, 43(5): 300—321
- 7 Riederer K A J. Head-related transfer function measurement. Thesis of Master Degree of Science. Helsinki: Helsinki University of Technology, 1998. 82—113
- 8 Algazi V R, Duda R O, Thompson D M, et al. The CIPIC HRTF database. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New York: IEEE Press, 2001. 99—102
- 9 Zhong X L, Xie B S. Spatial characteristics of head related transfer function. Chin Phys Lett, 2005, 22(5): 1166—1168[DOI]
- 10 Møller H. Fundamentals of binaural technology. Appl Acoust, 1992, 36(3/4): 171—218[DOI]

- 11 Blauert J. *Spatial Hearing*. 2nd ed. Cambridge, England: MIT Press, 1997. 140—155
- 12 谢波荪. 头部尺寸对虚拟声像定位的影响. *应用声学*, 2002, 21(5): 1—7
- 13 Moore B C J. *An Introduction to the Psychology of Hearing*. San Diego, USA: Academic Press, 2003. 73
- 14 Kulkarni A, Colburn H S. Role of spectral detail in sound-source localization. *Nature*, 1998, 396 (24/31): 747—749[DOI]
- 15 Raykar V C, Duraiswami R, Yegnanarayana B. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *J Acoust Soc Am*, 2005, 118(1): 364—374[DOI]
- 16 Zhang M, Tan K C, Er M H. Three-dimensional sound synthesis based on head related transfer functions. *J Audio Eng Soc*, 1998, 46(10): 836—844
- 17 Algazi V R, Avendano C, Duda R O. Estimation of a spherical-head model from anthropometry. *J Audio Eng Soc*, 2001, 49(6): 472—478
- 18 Kulkarni A, Isabelle S K, Colburn H S. Sensitivity of human subjects to head-related transfer functions phase spectra. *J Acoust Soc Am*, 1999, 105(5): 2821—2840[DOI]
- 19 金利新, 丁士海, 侯海青. 山东汉族大学生头面部测量. *青岛大学医学院学报*, 2000, 36(1): 43—45