

虚拟听觉环境实时绘制系统平台

张承云^①, 谢菠荪^{①②*}

① 华南理工大学理学院物理系声学研究所, 广州 510640;

② 华南理工大学亚热带建筑科学国家重点实验室, 广州 510640

* 联系人, E-mail: phbsxie@scut.edu.cn

2012-02-16 收稿, 2012-05-10 接受

国家自然科学基金(11174087, 10774049)和华南理工大学亚热带建筑科学国家重点实验室课题资助

摘要 报道了基于个人计算机和 C++ 语言编程的虚拟听觉环境(VAE)系统平台的近期工作进展. 系统实时跟踪倾听者头部的空间位置和方向, 动态模拟声波从声源到双耳的传输, 通过耳机重放能够产生自由场和反射声环境的逼真听觉感知. 提出了一系列改善系统性能的方法, 包括基于主成分分析(PCA)的近场虚拟声源合成, 倾听者的 6 个自由度运动模拟等. 特别是在虚拟多声源时, 能明显减少运算量和数据量. 实际测量表明, 与现有系统相比, 性能得到提升. 在传统的工作模式下, 系统可以同时合成 280 个虚拟声源(包括直达声源和反射虚声源); 而在本文提出的 PCA 工作模式下, 系统可以同时合成 4500 个虚拟声源. 一组心理声学实验也证实了系统的实际感知性能, 同时得出了一些双耳听觉的初步研究结果. 随着其性能的扩展, 系统可作为双耳听觉研究和虚拟现实应用的一个功能灵活且强大的平台.

关键词

虚拟听觉环境(VAE)

动态实时绘制

头相关传输函数(HRTF)

主成分分析(PCA)

虚拟听觉环境是由人工产生或控制声学环境, 使倾听者产生犹如置身于自然声学环境的听觉感知. 这是声学(物理学)、信号与信息处理、计算机科学与人类感知等跨学科领域的高技术研究课题. 虚拟听觉环境不仅可以作为人类双耳听觉感知研究的一种重要基础实验手段, 也是多媒体与虚拟现实技术的一个重要组成部分, 并且在各种虚拟训练设备(如航天或航空训练、军事训练等)、声学辅助设计(如各种室内声学设计)和通信等方面有重要的应用^[1].

在自然的声学环境中, 声源辐射的声波通过直达声和各种环境边界反射和散射的途径传输, 形成空间声场, 它包含了声源和环境的时间和空间声学信息. 当倾听者进入声场后, 双耳所接收到的是包括其自身生理结构反射和散射的声压. 生理结构的反射和散射将声源和环境的声学信息转换成鼓膜处的双耳声(压)信号. 此外, 声源或倾听者头部运动都会引起双耳声压变化, 带来动态的听觉信息. 双耳声

信号经听觉系统(包括高层神经系统)处理后, 形成相应的空间听觉事件, 如声源定位、声学环境的听觉感知等.

由于双耳声信号包括了声音的主要信息, 虚拟听觉环境可以通过模拟双耳声信号并用耳机(或扬声器)重放的方法实现. 也就是根据给定的声源、声环境和倾听者信息, 通过对声源的物理特性、声传输特性(包括直达声和环境反射声)、倾听者对声波的反射/散射三部分的模拟, 从而模拟出声波从声源到双耳传输的物理过程. 另外, 动态信息对虚拟听觉环境的主观真实性是至关重要的, 同时也会带来重要的声源定位因素, 这对区分前后甚至上下镜像方向的声源是重要的. 因而信号处理应该将这些信息及其动态变化合成出来. 所以完整的虚拟听觉环境系统应该是实时、交互、动态且尽可能考虑倾听者个性化特性(至少是合理的人群平均特性)的绘制系统^[1].

从 20 世纪 90 年代中期起, 国际上对动态虚拟听

觉环境进行了大量研究,并且有几个课题组实现了此类系统,包括德国 Ruhr-Universitat Bochum 研制的第一代 SCATIS^[2]、第二代 IKA-SIM 系统^[3]、美国 NASA(航空航天局)的 SLAB 系统^[4,5]、芬兰 Helsinki University of Technology 的 DIVA 系统^[6]、美国波士顿大学的系统^[7]、德国 RWTH Aachen University 的基于扬声器重放的虚拟听觉环境系统等^[8]。早期的系统是采用专用的数字信号处理芯片(DSP)实现的,随着计算机技术的发展,现在主要是在 PC 机或服务器平台上用 C/C++ 语言编程实现。

虽然国际上在虚拟听觉环境方面的研究已取得较大进展,但提高虚拟听觉环境系统的性能始终是一个开放的问题。为了得到真实的虚拟听觉环境效果,需要动态、实时地对声源、声传输、倾听者的散射与双耳声信号转换等进行模拟,但相应算法的计算量是很大的^[9],在需要同时模拟多个声源(包括直达声源和反射声的虚声源)时,这问题显得更为突出。在软硬件资源一定的条件下,通常只能降低系统的精确程度,使得信号处理可以实现。例如,限制同时模拟声源的数目;采用少量的虚声源模拟环境反射声(一阶,最多二阶)的空间信息而略去高阶反射声信息;采用非常低阶的滤波器(头相关传输函数)模拟倾听者对声波的散射;只考虑远场声源而忽略近场头相关传输函数带来的距离定位信息;只考虑倾听者头部水平面的运动自由度而略去其他自由度;降低系统的动态性能等。这不可避免地会影响虚拟听觉环境的性能。虽然软硬件技术的发展可在一定程度上缓解上述矛盾,但解决问题的关键应该是从声学 and 听觉的角度对信息处理进行简化,在软硬件资源一定的条件下使系统的性能得到提高。事实上,近十多年国际上一直在进行各种虚拟听觉环境系统简化及性能改进的工作^[10]。但这是一个非常复杂的问题,并非一两项研究就能完全解决,目前仍然是一个重要的研究发展方向。另一点也非常重要,到目前为止,国内其他课题组还没有研制出虚拟听觉环境实时绘制系统的基本软硬件实验平台,因而也就无法开展相关的工作。

在国家自然科学基金的资助下,我们课题组从 2007 年开始进行虚拟听觉环境实时绘制系统的研究工作,建立了相应的软硬件系统,为国内进一步开展人类双耳听觉感知研究和虚拟现实应用提供了实验基础平台。同时对虚拟听觉环境的一些基本信号处

理方法作出改进,在保持合理的动态性能的前提下,使系统在处理多虚拟声源、模拟听觉距离感知、多自由度动态信息等关键方面有明显的改进。本文将报道这方面的工作,包括系统的原理、结构和设计,动态客观性能的测试结果和利用该平台进行心理声学实验的一些初步结果。

1 系统的基本结构与原理

1.1 系统结构

采用标准的虚拟听觉环境实时绘制系统结构,它可分为三大部分,如图 1 所示。

(i) 信息定义与输入。信息定义与输入部分通过用户界面对所虚拟的听觉环境信息进行设定,包含声源、声传输环境及倾听者特性三部分。其中声源部分的设定包括声源信号的类型、声源数目、空间位置(或运动声源的空间轨迹)、方向及指向性(辐射模式)、音量大小等。声传输环境的设定包括空间环境的结构与几何尺寸、边界及空气的吸声特性等。倾听者特性设定包括倾听者的初始空间位置、方向及倾听者的个性化特性等。

(ii) 动态虚拟听觉环境信号处理。根据设定的信息,采用一定的物理算法模拟从声源到倾听者的传输过程,包括直达声和反射声的传输过程。采用头踪迹跟踪器检测倾听者头部方向和位置数据,刷新倾听者自身散射特性的模拟参数,从而实时、动态地模拟出双耳声信号。

(iii) 重放部分。对双耳声信号进行耳机-耳道传输特性预均衡后,采用耳机重放。

1.2 虚拟自由场声源处理

虚拟自由场声源是最简单的情况。用球坐标 (r, θ, ϕ) 定义声源 S 相对倾听者头部中心的空间位置,其中

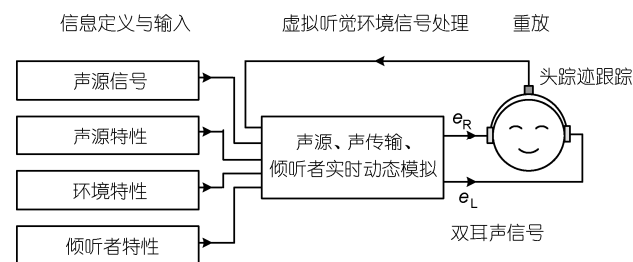


图 1 虚拟听觉环境实时绘制系统结构图

r 为声源距离; $-90^\circ \leq \phi \leq 90^\circ$ 和 $0^\circ \leq \theta < 360^\circ$ 分别表示仰角和方位角; $\phi = 0^\circ$ 和 $+90^\circ$ 分别表示水平面和正上方; 在水平面, $\theta = 0^\circ$ 和 90° 分别表示正前和正右方. 头相关传输函数(HRTF)定义为自由场点声源到双耳的声学传输函数:

$$H_\alpha(r, \theta, \phi, f) = \frac{P_\alpha(r, \theta, \phi, f)}{P_0(r, f)}, \quad (1)$$

其中, $\alpha = L$ 或 R 代表左耳或右耳; P_α 是位置为 (r, θ, ϕ) 的声源在左耳(或右耳)产生的频域声压; P_0 是头部移开后头中心位置处的声压. 一般情况下, HRTF 与声源位置以及频率 f 有关, 并且与个体有关. 对声源距离 $r \geq 1.0$ m 的远场, HRTF 近似与距离无关. 但对于 $r < 1.0$ m 的近场, HRTF 与距离有关, 因而包含距离定位的因素. HRTF 的时域表示是头相关脉冲响应(HRIR), 它们由傅里叶变换相联系.

在传统的信号处理中, 为了合成虚拟自由场点声源, 可以将单路的时域信号 $e_0(t)$ 作适当的延时处理和幅度标度后和一对相应 HRIR 卷积(或一对相应 HRTF 滤波), 得到双耳声信号:

$$e_\alpha(t) = \frac{1}{r} h_\alpha(r, \theta, \phi, t) * e_0(t - T), \quad (2)$$

其中 h_α 是声源到左耳(或右耳)的 HRIR, t 是时间. $T = r / c$ 是声源到倾听者的传输延时, c 是声速; 幅度标度 $1/r$ 模拟了自由场中球面声波的幅度随距离衰减.

对合成多虚拟声源的情况, 假设有 M 个虚拟声源, 其输入信号为 $e_{0,i}(t)$, 且分别位于空间位置 (r_i, θ_i, ϕ_i) , $i = 1, 2, \dots, M$. 那么根据声波的线性叠加原理, (2)式的双耳声信号可推广为 M 个虚拟声源贡献的叠加:

$$e_\alpha(t) = \sum_{i=1}^M \frac{1}{r_i} h_\alpha(r_i, \theta_i, \phi_i, t) * e_{0,i}(t - T_i), \quad (3)$$

其中 r_i 是第 i 个虚拟声源的距离, $T_i = r_i / c$ 为第 i 个虚拟声源到倾听者的传输延时.

如果 HRIR 可以作最小相位近似^[11], 每个 HRIR 可以近似为其最小相位函数 $h_{\min}$ 的纯延时:

$$h_\alpha(r, \theta, \phi, t) = h_{\min,\alpha}(r, \theta, \phi, t - \tau_\alpha), \quad (4)$$

其中 $\tau_\alpha = \tau_\alpha(r, \theta, \phi)$ 是和声源位置有关但与频率无关的纯延时. 将(4)式代入(3)式并利用卷积的延时性质后, 可以得到

$$e_\alpha(t) = \sum_{i=1}^M \frac{1}{r_i} h_{\min,\alpha}(r_i, \theta_i, \phi_i, t) * e_{0,i}(t - T_{\alpha,i}), \quad (5)$$

其中 $T_{\alpha,i} = T_i + \tau_\alpha(r_i, \theta_i, \phi_i)$.

由(3)式可以看出, 在传统的信号处理中, 虚拟 M 个声源需要将 M 个输入信号延时与标度后与 M 对 HRIR 进行卷积运算并将所得信号叠加. 如果 HRIR 的长度是 N 个采样点, 在不考虑幅度标度的情况下, 计算一个样本点的双耳信号需 $2MN$ 次乘法和 $2(MN-1)$ 次加法的运算. 随着声源数目 M 的增加, 其运算量是很大的, 可能会超出计算机的运算能力. 采用最小相位近似和时间窗截断的方法可减少 HRIR 长度, 从而减少运算量. 但在一定的采样频率下, 最小相位 HRIR 的长度减少到一定的下限以下(如 44.1 kHz 采样频率, 减少到 64 点以下时), 就会产生明显的可感知的畸变. 因此在实际系统中, 通常是采用限制声源数目的方法, 但这不可避免地会影响虚拟听觉环境的性能.

可以采用 HRTF 或 HRIR 基函数分解实现多虚拟声源联合处理, 从而减少运算量^[1,12]. 在现有的多种基函数分解方法中, 主成分分析(PCA)是其中较为有效的一种, 它将 HRIR 分解为少数基函数的权重组合. 但现有的主成分分析方法主要是针对远场 HRTF 或 HRIR 的, 不包括声源距离的变化信息. 为了虚拟不同距离的声源, 我们将主成分分析的方法推广到近场 HRIR. 以最小相位 HRIR 为例, 空间位置为 (r, θ, ϕ) 的最小相位 HRIR 可以分解为 Q 个谱形状基函数的权重组合:

$$h_{\min,\alpha}(r, \theta, \phi, t) = \sum_{q=1}^Q w_{q,\alpha}(r, \theta, \phi) g_q(t) + h_{\min,\text{av}}(t), \quad (6)$$

其中 $g_q(t)$, $h_{\min,\text{av}}(t)$ 分别是基函数与平均最小相位 HRIR, 它们都只是时间的函数, 与声源位置以及左右耳无关. 而 $w_{q,\alpha}(r, \theta, \phi)$ 是相应的权重系数, 它只和声源位置以及左右耳有关, 与时间无关. 因此在主成分分解中, 只要改变权重即可得到不同声源位置的 HRIR. 对每名受试者, 只要通过测量得到一组不同空间方向的 HRTF 数据, 即可通过统计方法求出 $h_{\min,\text{av}}(t)$ 和 $g_q(t)$, 并用函数正交投影的方法得到 $w_{q,\alpha}(r, \theta, \phi)$. 具体的计算方法是和远场 HRIR 或 HRTF 的 PCA 分解类似的^[13], 由于篇幅关系, 这里不再详述, 我们将在另文讨论.

把(6)式的最小相位近场 HRIR 的 PCA 分解表示代入(5)式, 可以得到简化的多虚拟声源联合处理方法.

$$e_\alpha(t) = \sum_{i=1}^M \frac{1}{r_i} \left[\sum_{q=1}^Q w_{q,\alpha}(r_i, \theta_i, \phi_i) g_q(t) + h_{\min,\text{av}}(t) \right] * e_{0,i}(t - T_{\alpha,i})$$

$$= \sum_{q=1}^Q g_q(t) * \left[\sum_{i=1}^M w_{q,\alpha}(r_i, \theta_i, \phi_i) \frac{1}{r_i} e_{0,i}(t - T_{\alpha,i}) \right] + h_{\min,av}(t) * \left[\sum_{i=1}^M \frac{1}{r_i} e_{0,i}(t - T_{\alpha,i}) \right]. \quad (7)$$

图 2 是根据(7)式画出的信号处理的方块图. 可以看出, M 个声源共用 $(Q+1)$ 个卷积函数 $g_q(t)$ 和 $h_{\min,av}(t)$, 相当于 $(Q+1)$ 个频域滤波器, 而虚拟声源方位由 $w_{q,\alpha}(r, \theta, \phi)$, $1/r_i$ 以及 $T_{\alpha,i}$ 来控制. 因此不管虚拟多少个声源, 合成双耳声信号的卷积固定为 $2(Q+1)$ 个. 由于 PCA 分解使得 M 个虚拟声源可以共用一组平行的卷积(或滤波器)处理, 因而在多声源处理中比传统处理方法效率更高. 事实上, 如果 HRIR 的长度是 N 个采样点, 则利用(7)式同时虚拟 M 个声源, 不考虑幅度标度时计算一个样本点的双耳信号需 $2(NQ + MQ + N)$ 次乘法和 $(2N + 2M - 4)(Q + 1)$ 次加法运算. 与前面的传统信号处理方法比较, 只要满足 $M > N(Q + 1)/(N - Q)$ 的条件, 计算即可得到简化. 且 M 越大, 效果越明显.

本研究将采用以下三组不同的 HRIR 数据库.

(i) MIT 媒体实验室测量得到的 KEMAR 人工头(配 DB 61 小号耳廓)的远场 HRTF^[14], $r = 1.4$ m, 共 710 个声源方向的 HRTF 数据. 测量 HRIR 数据的采样

频率是 44.1 kHz, 长度 512 点. 这是目前应用最多的非个性化 HRTF 数据, 本研究采用它是为了便于对比.

(ii) 本实验室测量得到的中国人受试者样本的个性化远场 HRIR 数据^[15]. 该数据是采用封闭耳道法测量得到的. 数据包括 52 名受试者, $r = 1.5$ m, 每名受试者共有 493 个声源方向的 HRTF 数据, 它们在各纬度面的分布如表 1 所示. 测量 HRIR 数据的采样频率是 44.1 kHz, 长度 512 点.

(iii) 本实验室测量得到的 KEMAR 人工头的近场 HRIR 数据^[16]. 该数据的双耳声压测量点是在耳道模拟器的末端. 数据包括 r 从 0.2 m 到 1.0 m, 间隔 0.1 m 共 9 个声源距离; 在每个声源距离上包括 493 个声源方向, 它们在各纬度面的分布和表 1 的情况相同. 测量 HRIR 数据的采样频率是 44.1 kHz, 长度 512 点. 由于测量真人受试者的近场 HRTF 非常困难, 目前国际上也没有这方面的数据库(即使是人工头的近场 HRIR 数据库也只有少数几个), 所以本研究暂时采用人工头的数据. 本实验室正在进行中国人受试者样本的个性化近场 HRIR 数据的测量工作, 待该项工作完成后, 将所得到的数据换上即可.

虚拟声源可选择采用两种信号处理模式.

(i) 传统的信号处理模式, 如(3)式或(5)式所示.

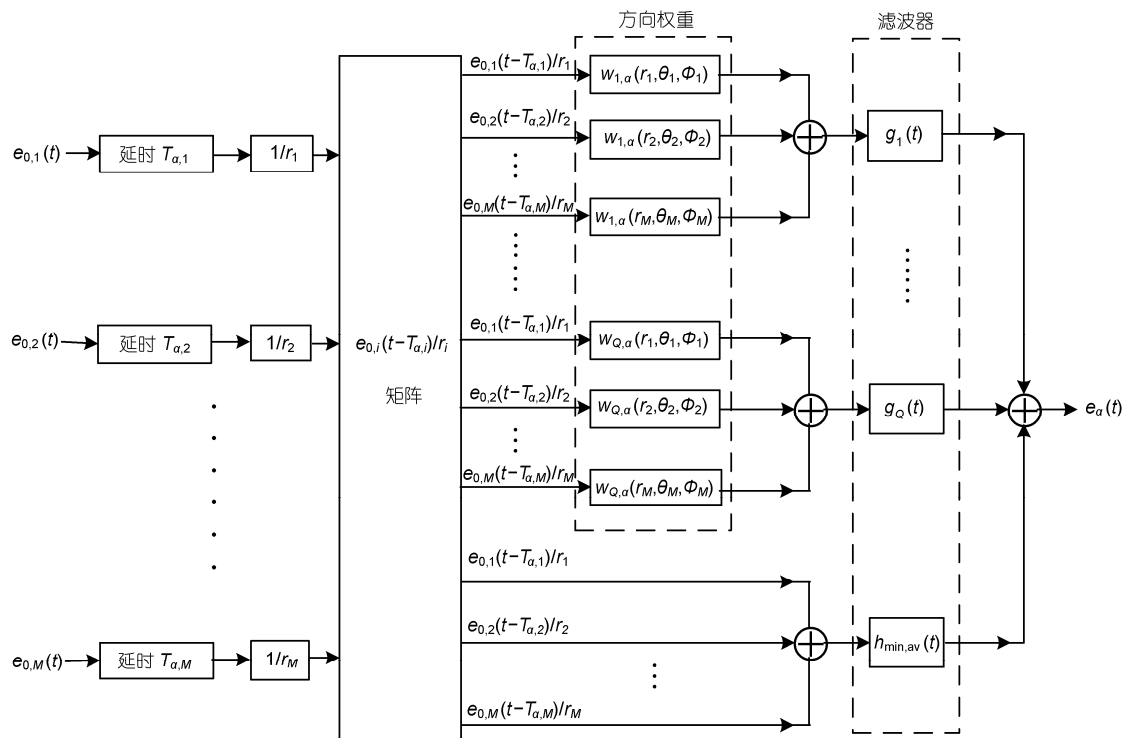


图 2 基于 PCA 模式的多虚拟声源联合处理框图(单耳)

表 1 测量 HRTF 的声源方向分布

纬度面 ϕ	方位角间隔 $\Delta\theta$
-30°	5°
-15°	5°
0°	5°
15°	5°
30°	5°
45°	5°
60°	10°
75°	15°
90°	360°

所用 HRIR 的长度为 128 点, 用矩形时间窗对原始的 512 点 HRIR 截取得到. 如果采用适当的 HRTF 幅度平滑和最小相位近似^[17], HRIR 长度可进一步减少到 64 点.

(ii) 基于 PCA 的多虚拟声源联合处理模式, 如(7)式和图 2 所示. 这是本系统与现有系统的一个不同之处. 将原始的 512 点 HRIR 作最小相位近似后, 用矩形时间窗截取 128 点. 然后用 PCA 的方法进行分解. 具体计算结果表明, 无论对真人受试者的个性化远场 HRIR 数据, 还是 KEMAR 人工头的近场 HRIR 数据, 采用 $Q = 15$ 个时间基函数可代表 HRIR 能量变化的 97.4% 以上, 因而可获得足够的准确性.

无论采用哪一种信号处理模式, 都需要采用个性化耳机-耳道传输函数进行均衡处理^[18].

1.3 HRIR 空间插值与虚拟运动声源

HRIR 是声源空间位置的连续函数. 但实验测量是在离散的空间位置进行的, 得到的是有限空间方向的 HRIR 数据. 为了虚拟空间任意位置的声源, 需要通过空间插值的方法得到未测量方向的 HRIR 数据. 现有的多种不同空间插值方法都可以用于 HRIR 的方向插值^[1], 别外还需要对 HRIR 进行距离插值. 本系统采用最简单的双线性方向插值和线性距离插值的方法, 也就是将球面上的空间方向分为四边形网格, 先采用相邻四个测量方向的 HRIR 的线性组合来近似未测量方向的 HRIR, 然后采用两个相邻距离的已知 HRIR 的线性组合来近似未测量距离的 HRIR.

由于 HRIR 通常是 N 点有限长度的离散时间序列, 在传统的虚拟声源信号处理模式下, 需要对 HRIR 进行逐点插值, 比较复杂. 在基于 PCA 的处理模式下, (7)式给出的时间基函数 $g_q(t)$ 是与声源位置无关的, HRIR 与声源位置的关系是由 $2Q$ 个权重系数 $w_{q,a}(r, \theta, \phi)$ 所代表. 因而空间插值只需要对 $2Q$ 个权重

系数进行, 比直接对 HRIR 插值方便得多, 这是基于 PCA 分解的处理模式的另一个优点, 在下面讨论的动态信息处理中尤为突出.

对于运动声源, 其空间的轨迹由以下参数方程所决定:

$$r = r(t), \quad \theta = \theta(t), \quad \phi = \phi(t). \quad (8)$$

传统的虚拟运动声源处理模式是根据时变声源位置不断地刷新(3)式的 HRIR、标度因子 $1/r_i$ 和延时 T_i , 但这比较复杂, 容易产生可察觉的噪声. 在基于 PCA 分解的处理模式下, 只要根据(8)式连续改变(7)式的权重 $w_{q,a}(r, \theta, \phi)$ 、标度因子 $1/r_i$ 和延时 $T_{a,i}$ 即可产生运动虚拟声源, 信号处理比较简单, 同时避免了不断切换 HRIR 所带来的可察觉的噪声. 对虚拟高速运动的声源, 还可以增加多普勒效应的模拟, 详细见文献[19].

1.4 动态信息处理

对于动态情况, 假定虚拟声源在空间的绝对位置不变, 当倾听者头部运动后, 虚拟声源与倾听者头部的相对位置会变. 因此需根据虚拟声源新的球坐标 (r', θ', ϕ') 不断刷新(3)式或(7)式的稳态虚拟声源合成参数, 包括标度因子、延时及 HRIR 或权重系数等.

在三维空间中, 头部共有 6 个运动自由度, 包括 3 个平移自由度和 3 个转动自由度. 如图 3 所示, 需要 3 个直角坐标 (x, y, z) 及 3 个角度转动参数 (α, β, γ) 来完整地描述倾听者头部位置和方向. 假设开始时头部中心位于坐标原点 $(x = 0, y = 0, z = 0)$, x, y, z 轴分别指向正右、正前和正上方. 当头部运动后, 平移量用 $(\Delta x, \Delta y, \Delta z)$ 来表示. 转动的角度用 (α, β, γ) 表示, 分别对应绕 z 轴、 x 轴、 y 轴的转动. 则由几何分析可得到头部运动前后声源相对头部的位置关系为

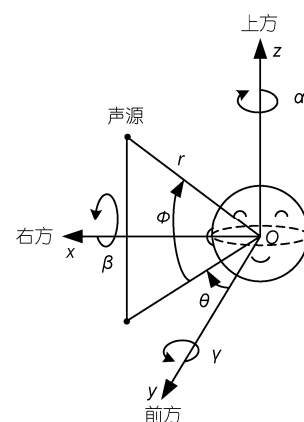


图 3 坐标系统

$$\begin{bmatrix} r' \cos \phi' \sin \theta' \\ r' \cos \phi' \cos \theta' \\ r' \sin \phi' \end{bmatrix} \begin{bmatrix} \cos \alpha \cos \gamma + \sin \alpha \sin \beta \sin \gamma & -\sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma & \cos \beta \sin \gamma \\ \sin \alpha \cos \beta & \cos \alpha \cos \beta & -\sin \beta \\ -\cos \alpha \sin \gamma + \sin \alpha \sin \beta \cos \gamma & \sin \alpha \sin \gamma + \cos \alpha \sin \beta \cos \gamma & \cos \beta \cos \gamma \end{bmatrix} \begin{bmatrix} r \cos \phi \sin \theta - \Delta x \\ r \cos \phi \cos \theta - \Delta y \\ r \sin \phi - \Delta z \end{bmatrix}. \quad (9)$$

利用(9)式的关系计算出虚拟声源相对倾听者的位置参数后,即可实现动态信号处理.与虚拟运动声源的情况类似,传统的处理模式是根据(9)式的时变声源相对倾听者头部位置不断地刷新(3)式或(5)式的 HRIR、标度因子 $1/r_i$ 和延时 T_i ,但这比较复杂.由于在实际中,只能以一定的时间间隔对倾听者位置进行检测(时间采样),直接从一个位置的 HRIR 切换到另一个位置的 HRIR 容易产生可察觉的切换噪声.解决办法是通过同时进行两个位置的 HRIR 卷积并从一个卷积的输出过渡到另一输出,但这需要双倍的 HRIR 卷积运算,在虚拟多声源的情况下更加耗费计算资源.在基于 PCA 分解的处理模式下,连续改变(7)式的权重 $w_{q,a}(r, \theta, \phi)$ 、标度因子 $1/r_i$ 和延时 $T_{a,i}$ 即可实现动态信息处理,信号处理比较简单,也避免了可察觉的切换噪声.

在本系统将对头部运动的 6 个自由度进行跟踪,可选择采用传统或基于 PCA 分解的处理模式实现动态信息处理.

1.5 环境反射声的模拟

完整的虚拟听觉环境应包括反射声,这对重放的主观真实性非常重要.环境反射声的物理模拟是室内声学的重要内容,有很多方法可以用于环境反射声的模拟^[20].因为环境反射声的物理模拟已形成一个相对独立的研究领域,因而为简单起见,不在这方面作详细的探讨.这里仅采用简单的虚源法模拟房间反射声.

以最简单的矩形房间为例,一阶反射包含有 6 个虚声源,二阶反射包含有 30 个虚声源, L 阶反射包含有 $6 \times 5^{(L-1)}$ 个虚声源,前 L 阶反射共包含 $3 \times (5^L - 1)/2$ 个虚声源,各虚声源的位置可由镜像几何关系确定.另外,用虚源法模拟室内反射声,必要时还需对虚声源进行可视性检验.

一旦虚声源的空间位置确定后,相应的反射声就可以看成是由虚声源位置发出,像直达声或自由场虚拟声源一样采用(3)式或(7)式分别对每个反射声进行处理.当倾听者头部的位置改变后,虚声源相对头部的位置随之改变,因而需要对每一个虚声源都

实施 1.4 节的动态信息处理.但是,因为虚声源的数目随反射声阶数的增加而快速增加,可能由于运算量太大而超过系统的计算能力.因而采用传统的信号处理模式时,通常只能模拟到一阶最多是二阶的反射,现有的研究也大部分是如此.如果采用基于 PCA 的多虚拟声源联合处理模式,则可缓解运算量负担,因为卷积运算不随虚拟声源数目而增加.

对于壁面的吸收,当与频率有关的反射或吸收系数给定后,可以设计出相应低阶滤波器模拟.而对于声源的指向性(非点声源),也可以用相应的指向性滤波器来模拟.我们的系统采用虚声源来模拟早期反射声(最高为 3 阶).而后期混响则通过人工混响算法实现,这些都有现成的算法^[21],虽然效果并不完全理想,但可以直接应用.

2 系统软硬件设计

2.1 硬件结构

如图 4 所示,硬件部分包括一台 4 核个人计算机(Intel Q9550 CPU@2.83 GHz, 4 G 内存)、支持 ASIO 功能的声卡(ESI UGM96)、耳机(Sennheiser HD250)、电磁头踪迹跟踪器(Polhemus FASTRAK).

虚拟听觉环境实时绘制系统有两个重要的动态参数:场景刷新率和系统滞后时间.场景刷新率指的是单位时间内的信号处理(场景)刷新次数,主要取决于头踪迹跟踪器的性能.系统滞后时间指的是从倾听者运动到某个位置与系统输出相应的响应信号之间的时间差,受头踪迹跟踪器、声卡、信号处理、数据通信、数据缓存的影响,而前两项占大部分.因此头踪迹跟踪器和声卡的性能对系统的动态性能有重

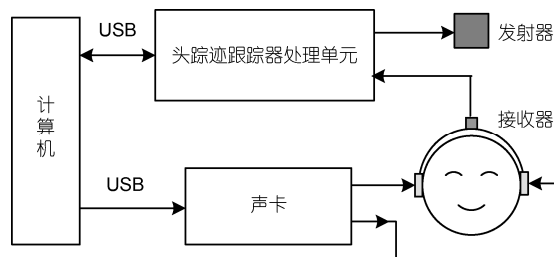


图 4 硬件结构示意图

要影响,需慎重选取。

本文采用 Polhemus FASTRAK 作为头踪迹跟踪器,其包含处理单元(1个)、发射器(1个)、接收器(最多4个),可输出平移和转动共6个自由度参数,只接1个接收器时数据刷新率为120 Hz,数据输出的滞后时间为4 ms,距离测量精度0.08 cm、分辨率0.0005 cm,角度测量精度0.15°、分辨率0.025°。

声卡与计算机操作系统之间的驱动程序对延时会有很大影响,目前声卡的驱动模式主要有 MME, WDM, DirectSound, ASIO 等,其中 ASIO 的延时最短,但要求声卡上要集成专用电路。ASIO 是由 Steinberg 公司提出来的,可从其网站上免费下载 API 函数。本文选用的是支持 ASIO 驱动的 ESI UGM96 声卡,缓冲区大小选用128个样本,数据字长24比特。

计算机是整个系统的信息处理中心,系统开始工作前计算机先通过 USB 接口下载参数完成头踪迹跟踪器及声卡的初始化工作,以及通过软件界面设置声源、声环境和倾听者信息。系统开始工作后头踪迹跟踪器将倾听者的位置和方向信息通过 USB 接口传给计算机,信号处理软件据此信息计算出双耳信号,然后通过声卡传送给耳机重放。

2.2 软件结构

软件部分在 Microsoft Visual Studio.NET 2003 平台上用 C++ 语言实现,完成了人机交互、头踪迹跟踪数据接收、信号处理参数计算、信号处理、音频数据输入输出等功能,整个系统包括3个线程,实现5个功能模块,如图5所示。

(i) 人机接口模块:实现声源、声环境、倾听者信息的设置与定义,实时显示虚拟声源方向和距离等。

(ii) 头踪迹跟踪器接口模块:实现计算机与头踪迹跟踪器的连接,实时接收跟踪器传送过来的倾听者头部位置和方向信息。

(iii) 声卡接口模块:实现计算机与声卡的连接,对声卡进行参数设置,将处理好的音频数据发送给声卡。此模块最重要的是 ASIO 驱动的使用,Steinberg 公司提供了免费 API 函数。

(iv) 信号处理参数计算模块:根据倾听者头部运动数据,根据(9)式计算出(实或虚)声源的方向和距离;根据声源距离计算延时、衰减系数。如果有需要,该模块还对虚声源进行可视性检验。

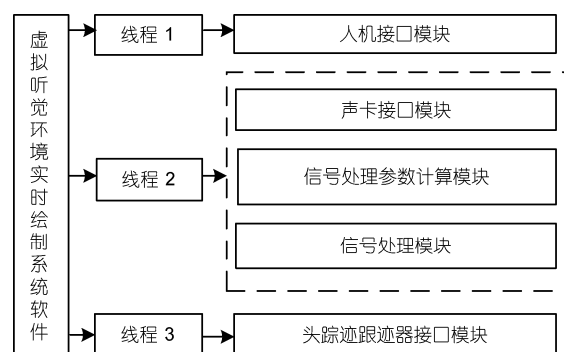


图5 软件模块组成框图

(v) 信号处理模块:对声源、声环境及倾听者进行模拟,并做耳机-耳道传输特性均衡处理。为了降低系统的滞后时间,输入信号数据块的长度越短越好。但是其取值又受声卡性能限制,实验验证 UGM96 声卡的 ASIO 数据缓冲区取64个样本点时可正常工作,但为稳定起见,取128点,带来的延时为2.9 ms(采样率为44.1 kHz)。

3 系统客观性能参数

场景刷新率、系统滞后时间及系统能同时处理的声源个数(实和虚)是虚拟听觉环境实时绘制系统的重要客观性能参数,我们对本文所构建系统的这些参数进行了测试。

场景刷新率取决于头踪迹跟踪器数据刷新率,Polhemus FASTRAK 在只有一个接收通道的情况下刷新率为120 Hz。由于计算机内部有高精度定时器,所以可以很方便地测量出这个参数。经测量,两次场景刷新之间的时间间隔平均值为8.33 ms,折算成刷新率即为120 Hz,与厂家提供的参数一致。系统滞后时间的测试采用了一种较简便的方法^[22],测量得到的平均系统滞后时间为25.4 ms。

系统能同时处理的声源个数与计算机软硬件均有关系,本文系统的测量结果如下:

(i) 在传统的信号处理模式下,可同时处理约280个声源;

(ii) 基于 PCA 的多虚拟声源联合处理模式下,可同时处理约4500个声源;

(iii) 在 PCA 模式下,如果用于虚拟矩形房间的情况,系统可以模拟前4阶的早期反射声虚声源,但考虑到可能需要与传统处理方法做对比实验,因而本研究暂时模拟前3阶的早期反射声,共包括186个

虚声源.

为了与现有系统比较,表 2 列出了几个现有系统和本系统的性能参数,为方便对比,只列出了采用耳机重放的此类系统.现有系统的参数来自文献[2~6].由于系统的软硬件结构不同,因此性能上有些差异.但总体来说,本系统在性能上显示出其优势.

4 心理声学实验

本文的系统是作为双耳听觉和虚拟现实研究的一个平台.心理声学实验的目的一方面是对系统的主观性能进行验证,另一方面是应用系统平台对双耳听觉进行一些初步研究,给出相应研究结果.我们已设计和完成了一系列的实验,由于篇幅所限制,这里仅给出有代表性的 3 个心理声学实验的初步结果,更详细的结果和发现将在另文发表.

所有实验均采用全可听频带白噪声信号作为单路原始信号,白噪声信号是在计算机中通过软件产生(采样率 44.1 kHz、量化精度 16 bit).实验在一间混响时间为 0.15 s、本底噪声不大于 30 dBA 的听音室内进行,受试者坐在听音室中心位置的椅子上.

4.1 远场虚拟声源定位实验

采用非个性化的 HRTF 数据、在传统的信号处理模式下合成自由场的远场虚拟声源是最简单的情况,也是对系统主观性能最基本的验证.这里采用 MIT 媒体实验室测量的 KEMAR 人工头的远场 HRTF 数据,用传统的信号处理模式分别作稳态和动态自由场虚拟声源信号处理,对比虚拟声源的方向和距离定位结果.

共选择 28 个虚拟声源方向,考虑到对称性,这些方向均分布在右半球 4 个纬度面 $\phi = -30^\circ, 0^\circ, 30^\circ, 60^\circ$, 在每个纬度面取 7 个方位角 $\theta = 0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ$. 共有 6 名受试者(男女各 3 名)参加

实验.对稳态和动态信号处理,每名受试者对每个虚拟声源方向分别作 3 次判断,6 名受试者共有 18 次判断.受试者报告虚拟声源方向和距离的方法有多种,本文采用跟踪器的虚拟声源定位方法^[23],实验中受试者只需将跟踪器的接收器指到虚拟声源位置,计算机即可实时显示其方向和距离,同时将数据保存到文件中.

处理实验数据时先统计出前后和上下混乱率,也就是将前半球空间的虚拟声源判断成后半球空间的虚拟声源(或相反)、将上半球空间的虚拟声源判断成下半球空间的虚拟声源(或相反)的比例.统计结果表明,稳态和动态信号处理对应的前后混乱率分别为 30.3%和 2.5%,上下混乱率分别为 24.1%和 13.7%,其中动态信号处理的前后混乱主要发生在 $\phi = 60^\circ$ 的高纬度面.

采用稳态信号处理时存在前后和上下镜像方向的虚拟声源混乱现象.采用动态信号处理后,即使是采用非个性化的 HRTF,也可以基本消除前后方向的虚拟声源混乱现象,明显降低上下混乱率.这是因为听觉定位因素是有冗余的,动态处理带来了稳定的定位因素,减少了对个性化 HRTF 谱因素的依赖,这也是和过去的一些实验结果相一致的.由于这里对倾听者头部 3 个转动自由度信息进行了处理,因此前后及上下混乱率都得到明显下降,而过去有些实验仅考虑头部绕垂直轴的转动,所以只能降低前后混乱率.

为了更加直观地描述虚拟声源方向的实验结果,采用 Leong 等人^[24]提出的图示化方法来显示实验数据统计结果.图 6 是经过对镜像方向混乱的数据进行空间反演后感知声源方向的统计结果,为了方便观察各个方向的虚拟声源分布,从正前方、正右方及正后方三个视点将实验结果显示出来.“+”表示虚拟声源方向的期望值,椭圆中心的小黑点代表平均的虚

表 2 现有系统性能参数

系统名称	系统滞后时间(ms)	场景刷新率(Hz)	能同时处理的声源个数	滤波器长度(点)
SCATIS	66	60	32	80
IKA-SIM	60	60	40~60	64/96/128
SLAB	24(不包括跟踪器的延时)	120	24	128(直达声) 32(反射声)
DIVA	110~160	20	10~20	30~35(FIR) 15~18(IIR)
本文系统	25.4	120	280(传统模式) 4500(PCA 模式)	128

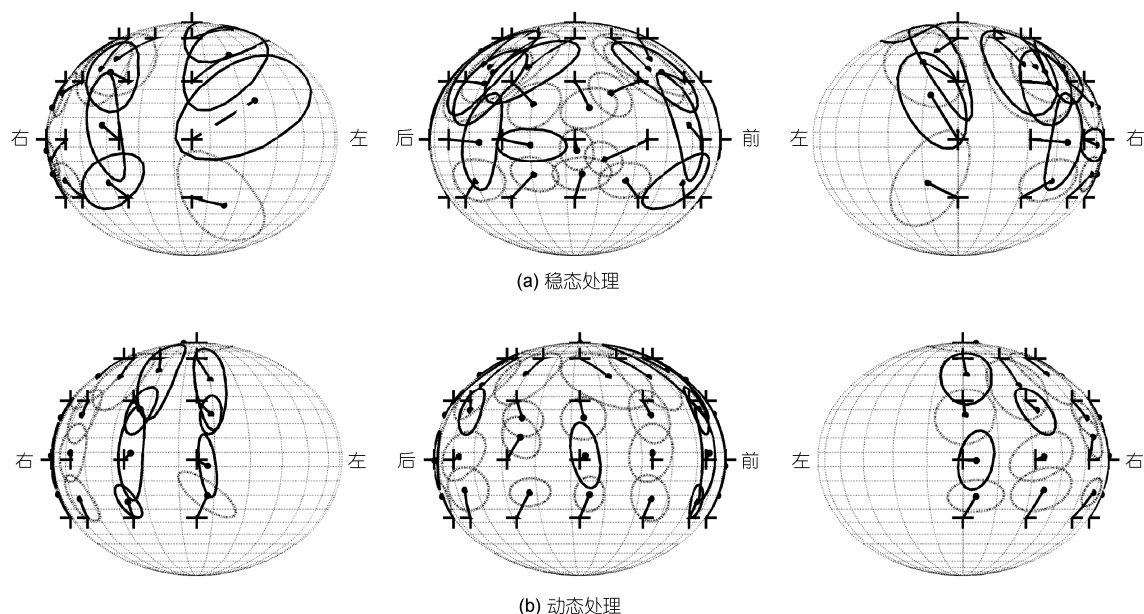


图6 远场虚拟声源的方向定位实验结果统计图

拟声源感知方向,而椭圆是显著性水平为 α 时的置信区间(本文取 $\alpha = 0.05$).若实验数据对称性好,则椭圆用虚线表示,否则用实线.从图中可以看出,动态处理时虚拟声源方向定位结果的离散性小很多,与理论值更为接近.

关于距离定位,对稳态处理的情况,在每一个方向上,6名受试者共18次判断的平均感知声源距离在0.16~0.62 m之间(靠近头表面).而对于动态处理的情况,平均感知声源距离在0.86~1.23 m之间.因此动态处理对虚拟声源的距离感知有很重要的作用.

以上关于稳态与动态处理的结论还可以通过T检验($\alpha = 0.05$)进一步进行证明,从而验证了虚拟听觉环境实时绘制系统的基本性能.

我们还进一步采用个性化HRTF进行动态信号处理,结果表明,这可以进一步减少定位的错误.我们也进行了包括环境反射声情况的定位实验研究.但这方面实验结果的分析比较复杂,由于篇幅所限,详细结果将在另文讨论.

4.2 近场虚拟声源的定位实验

本实验的目的一方面是验证自由场情况下的近场虚拟声源的合成,同时也对动态处理时采用传统处理模式与基于PCA的处理模式的实际效果进行对比.

采用我们课题组测量得到的KEMAR人工头近场HRTF数据,分别在传统的处理模式和PCA处理

模式下,合成动态的自由场虚拟声源.取5种不同的虚拟声源距离 $r = 0.2, 0.4, 0.6, 0.8, 1.0$ m;每个距离上取 $\phi = -30^\circ, 0^\circ, 30^\circ, 60^\circ$ 的4个纬度面,其中 $\phi = 60^\circ$ 时取3个方位角 $\theta = 0^\circ, 90^\circ, 180^\circ$,其余3个纬度面中每个纬度面取5个方位角 $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ$.

共有10名受试者(男女各5名)参加实验,实验流程与4.1节相同.对每种处理模式和每个虚拟声源位置,每名受试者分别作4次判断,因此每种情况共有40次判断,而每名受试者共需判断720次.

处理虚拟声源的方向数据与4.1节类似,表3给出了不同虚拟声源距离时感知方向的平均前后和上下混乱率,传统处理模式和PCA处理模式下的结果基本相同.除了 $r = 0.2$ m时的混乱率高一些外,其余几个距离对应的结果与4.1节类似.

将镜像方向混乱的数据进行空间反演,然后再进行虚拟声源方向统计.结果除了 $r = 0.2$ m时的数据离散性大一些外,其余几个距离对应的结果与图6

表3 采用近场HRTF的虚拟声源定位混乱率统计表

设定的距离 (m)	前后混乱率(%)		上下混乱率(%)	
	传统模式	PCA模式	传统模式	PCA模式
0.2	9.0	8.2	20.4	24.8
0.4	2.7	2.0	13.5	14.6
0.6	1.6	2.3	11.5	16.3
0.8	2.3	1.6	10.8	16.0
1.0	2.9	2.3	11.5	16.7

比较类似, 且传统的处理模式和 PCA 处理模式下的结果基本相同. 作为例子, 图 7 给出了 $r = 0.8 \text{ m}$ 时的虚拟声源方向统计图, 包括传统处理模式与 PCA 处理模式下的结果.

对虚拟声源距离, 实验结果表明, 采用传统处理模式与 PCA 处理模式所得结果类似. 在偏离中垂面方向(特别是水平面侧向附近)和 $r \leq 0.6 \text{ m}$ 的距离, 采用不同距离的近场 HRTF 处理可得到不同的感知虚拟声源距离. 当 $r > 0.6 \text{ m}$ 时, 则基本分不清虚拟声源的感知距离. 这是因为头部阴影作用使侧向 HRTF 随声源距离明显变化, 从而带来距离定位因素. 作为距离定位实验结果的一个例子, 图 8 给出了两种处理模式下, 水平面 $\phi = 0^\circ$, $\theta = 90^\circ$ 的平均距离定位结果(10 名受试者, 每名受试者 4 次判断)和标准差. 但实际感知距离的离散性较大, 且平均值与期望(目标)距离也不一致. 这是声源(包括虚拟和真实)距离定位的普遍结果, 与现有的研究是一致的^[25]. 实际中, 距离定位是多种因素综合作用的结果, 而不仅是 HRTF 的距离变化. 例如环境反射声也是距离定位的一个因素, 但引入环境反射声后情况将更复杂, 这里略去这方面的讨论.

上面的实验结果表明, 在一定的条件(侧向)下, 采用近场 HRTF 可部分地产生不同的听觉距离感知. 且传统的处理模式和 PCA 处理模式可得到等价的虚拟声源定位结果, 这可以用数理统计的 T 检验($\alpha =$

0.05)证明^[26].

4.3 主观对比和选择实验

为了进一步检验 PCA 信号处理模式是否会影响其他的听觉感知效果(包括非空间感知, 如音色等), 进行了主观对比和选择实验.

采用三间隔、两强制选择(3I-2AFC)主观对比和选择实验的方法. 传统处理模式得到的信号作为参考信号 A, 用 PCA 模式处理得到的信号作为目标信号 B. 每次重放的信号包括 3 段, 第 1 段为参考信号, 第 2 和第 3 段分别是参考和比较信号, 有 AAB 和 ABA 两种次序. 每段信号长 10 s, 循环播放. 实验中, 受试者必须判断第 2 和第 3 段信号中哪一段与第 1 段在听觉上不同(包括方向、距离、音色), 若无法判断则以随机的方式强制选择. 18 个方向、5 种距离组合起来共有 90 组信号, 每组播放 6 次(“AAB”及“ABA”各 3 次), 所有信号按随机顺序播放. 参加实验的受试者共 10 名(男 6 名女 4 名).

对每个虚拟声源距离和方向, 10 名受试者共判断 60 次. 判断结果的正确率 p 是一个服从(0~1)分布的随机变量. 当样本足够大时, 若受试者无法判断而随机选择答案, 则正确率趋近于 $p_0 = 0.5$. 对于 3I-2AFC 实验的正确率, 一般认为当 $p_0 \geq 0.75$ 时可以辨别目标信号与参考信号的区别. 对实验数据按 $p_0 = 0.5$ 和 $p_0 \geq 0.75$ 分别作双边和单边假设检验, 可得到相

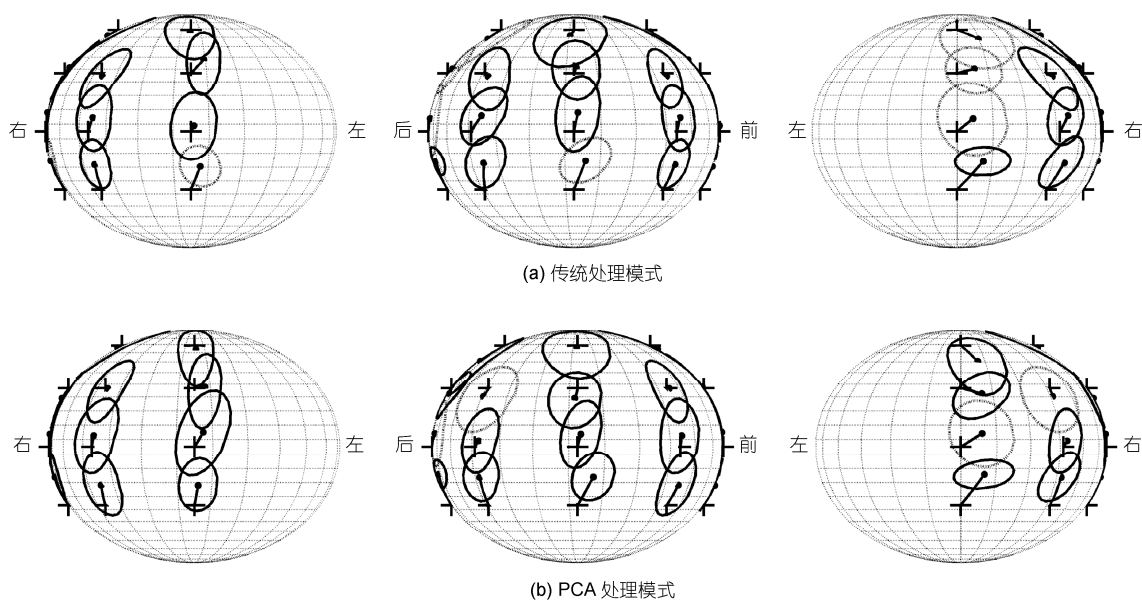


图 7 传统处理模式和 PCA 处理模式的虚拟声源方向定位实验结果统计($r = 0.8 \text{ m}$)

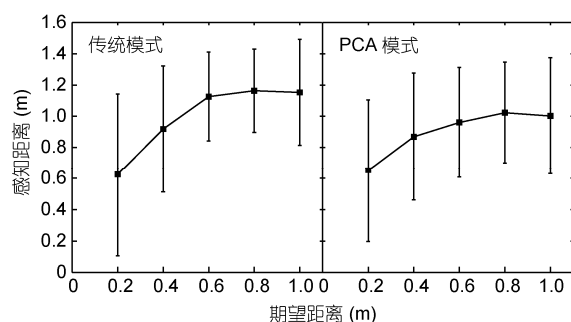


图8 水平面侧向距离定位实验统计结果($\theta = 90^\circ$, $\phi = 0^\circ$)

应的统计结果. 取显著性水平 $\alpha=0.05$, 则 $0.373 \leq p \leq 0.627$ 代表不可辨, $0.627 < p \leq 0.658$ 表示不可定论, $p > 0.658$ 表示可以分辨. 其中 $p=0.373$ 是实验失败临界值, $p=0.627$ 是不可辨别上限值, $p=0.658$ 是可辨别下限值. 实验结果显示, 除了 $(\theta=135^\circ, \phi=-30^\circ, r=0.2 \text{ m})$, $(\theta=135^\circ, \phi=30^\circ, r=0.4 \text{ m})$ 和 $(\theta=0^\circ, \phi=60^\circ, r=0.8 \text{ m})$ 不可定论之外, 其余均为不可辨别. 此结果进一步说明了 PCA 方法确实不会影响虚拟听觉效果.

5 结论

本研究构建了基于个人计算机和 C++ 语言编程的虚拟听觉环境实时绘制系统平台, 系统可灵活选

用传统的或 PCA 多声源联合处理工作模式, 采用个性化或非个性化, 远场或近场 HRIR 处理, 动态地模拟自由场和反射声环境的听觉感知, 采用虚源法模拟早期反射声. 系统在多声源动态处理、虚拟近场声源、模拟倾听者多自由度运动等方面的性能得到了提升. 实际测量表明, 该系统的动态参数为: 场景刷新率 120 Hz, 系统滞后时间 25.4 ms. 在传统的工作模式下, 系统能同时处理的声源个数(实和虚)约为 280 个. 若采用 PCA 多声源联合处理, 系统能同时处理约 4500 个声源. 整体性能上, 本文的系统已较现有的系统有明显的提高.

而心理声学实验结果验证了系统的实际感知性能, 得到一些虚拟听觉研究的初步结果. 结果表明引入多自由度动态因素后, 可明显减少上下虚拟声源混乱现象, 且基本消除前后虚拟声源混乱. 在一定的条件下, 采用近场 HRTF 可部分地产生不同的听觉距离感知. 实验同时证明, PCA 多声源联合处理可得到与传统处理等价的主观感知效果.

目前, 该系统的功能正在不断地扩展, 它将作为双耳听觉心理声学实验和多媒体研究的基础平台, 也可以作为其他虚拟现实系统的一个核心部分.

致谢 感谢钟小丽博士的帮助.

参考文献

- 1 谢淑萍. 头相关传输函数与虚拟听觉. 北京: 国防工业出版社, 2008
- 2 Blauert J, Lehnert H, Sahrhage J, et al. An interactive virtual-environment generator for psychoacoustic research I: Architecture and implementation. Acta Acust United Acust, 2000, 86: 94-102
- 3 Silzle A, Novo P, Strauss H. IKA-SIM: A system to generate auditory virtual environments. In: AES 116th Convention, Berlin, Germany, 2004. 6016
- 4 Wenzel E M, Miller J D, Abel J S. Sound Lab: A real-time, software-based system for the study of spatial hearing. In: AES 108th Convention, Paris, France, 2000. 5140
- 5 Miller J D, Wenzel E M. Recent developments in SLAB: A software-based system for interactive spatial sound synthesis. In: Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, 2002
- 6 Savioja L, Huopaniemi J, Lokki T, et al. Creating interactive virtual acoustic environments. J Audio Eng Soc, 1999, 47: 675-705
- 7 Scarpaci J W, Colburn H S, White J A. A system for real-time virtual auditory space. In: Proceedings of the 11th International Conference on Auditory Display, Limerick, Ireland, 2005. 241-246
- 8 Lentz T, Assenmacher I, Vorländer M, et al. Precise near-to-head acoustics with binaural synthesis. J Virtual Reality Broadcast, 2006, 3: urn:nbn:de:0009-6-5890
- 9 Sandvad J. Dynamic aspects of auditory virtual environments. In: AES 100th Convention, Copenhagen, Denmark, 1996. 4226
- 10 Begault D R, Wenzel E M, Godfroy M, et al. Applying spatial audio to human interfaces: 25 years of NASA experience. In: AES 40th Conference, Tokyo, Japan, 2010
- 11 Kulkarni A, Isabelle S K, Colburn H S. Sensitivity of human subjects to head-related transfer-function phase spectra. J Acoust Soc Am, 1999, 105: 2821-2840

- 12 Larcher V, Jot J M, Guyard J, et al. Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition of HRTF data. In: AES 108th Convention, Paris, France, 2000. 5097
- 13 Kistler D J, Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J Acoust Soc Am*, 1992, 91: 1637–1647
- 14 Gardner W G, Martin K D. HRTF measurements of a KEMAR. *J Acoust Soc Am*, 1995, 97: 3907–3908
- 15 谢波菽, 钟小丽, 饶丹, 等. 头相关传输函数数据库及其特性分析. *中国科学 G 辑: 物理学 力学 天文学*, 2006, 36: 464–479
- 16 Yu G Z, Xie B S, Rao D. Characteristics of near-field head-related transfer function for KEMAR. In: AES 40th Conference, Tokyo, Japan, 2010
- 17 Xie B S, Zhang T T. The audibility of spectral detail of head-related transfer functions at high frequency. *Acta Acust United Acust*, 2010, 96: 328–339
- 18 Møller H. Fundamentals of binaural technology. *Appl Acoust*, 1992, 36: 171–218
- 19 Krebber W, Gierlich H W, Genuit K. Auditory virtual environments: Basics and applications for interactive simulations. *Signal Processing*, 2000, 80: 2307–2322
- 20 Lehnert H, Blauert J. Principle of binaural room simulation. *Appl Acoust*, 1992, 36: 259–291
- 21 Gardner W G. Reverberation algorithms. In: Kahrs M, Brandenburg K, eds. *Applications of Signal Processing to Audio and Acoustics*. Boston: Kluwer Academic Publishers, 1998
- 22 Zhang C Y, Xie B S, Rao D, et al. Dynamic parameters measurement of virtual auditory environment real-time rendering system. In: *Proceedings of CISP*, Yantai, China, 2010. 3852–3855
- 23 张承云, 谢波菽. 用跟踪器报告虚拟声源位置的精度分析与验证. *华南理工大学学报(自然科学版)*, 2011, 39: 100–105
- 24 Leong P, Carlile S. Methods for spherical data analysis and visualization. *J Neurosci Meth*, 1998, 80: 191–200
- 25 Zahorik P. Assessing auditory distance perception using virtual acoustics. *J Acoust Soc Am*, 2002, 111: 1832–1846
- 26 盛骤, 谢式千, 潘承毅. *概率论与数理统计*. 第 3 版. 北京: 高等教育出版社, 2001