

# 虚拟环境中真实感空间声合成\*

张爱东 石教英 潘志庚

(浙江大学 CAD&CG 国家重点实验室 杭州 310027)

**摘要** 虚拟环境是用多种媒体构造的一个仿真视听等多方面感知效果的高级人机界面. 本文提出全面表达声音时空信息的思想, 讨论了用有限个音箱表现任意空间声的时间真实感和空间真实感. 通过对声音空间感实质的分析, 提出了基于三维空间向量分析的真实感空间声合成算法和基于限时计算的系统实现方法. 与基于双耳相关函数的空间声合成方法相比, 本方法能在较少的计算时间内合成更真实的空间声.

**关键词** 虚拟环境, 空间声, 时间真实感, 空间真实感, 限时计算.

虚拟环境作为人们可以预见的最完美的人机交互界面, 其特点就是要用计算机建造一个三维虚拟世界, 人在其中以自己的自然感知方式进行活动, 与计算机交换各种信息, 却不过多地感受到各种硬件设备的存在, 不过多地运用预定规则去理解各种信息. 以感知代替理解, 不仅可以减少人的工作量, 而且还可以更加丰富真实地表达信息, 因此可以说这是计算机表达信息的发展方向.

为了创造符合人类感知习惯的虚拟环境, 一方面需要提供多种媒体形式的表达方式, 比如视、听、触、味、嗅等感觉信息的载体; 另一方面则需要提高各种媒体信息的表达效果, 即真实感问题. 这里的真实感包括时间真实感 (Temporal Reality) 和空间真实感 (Spatial Reality) 2 个方面, 所谓时间真实感是指信息到达人类感知系统时间上的准确性; 而空间真实感是指从信源到信宿的空间方位上的准确性. 作为多媒体的 2 个主要方面: 图象和声音, 在以前研究得较多的是图象在空间上的真实感和声音在时间上的真实感. 随着计算机各方面技术的发展, 更多的研究逐渐转向图象的时间真实问题和声音的空间真实问题. 所谓空间真实的声音, 就是把声音不仅仅作为时间的函数, 而是作为时空四维的函数.

三维空间声合成技术是虚拟环境的一个重要组成部分<sup>[1]</sup>, 在难度上比真实感图象合成稍易, 但其地位是同样重要的. 表现真实感空间声有 2 类方法, 一类是基于双耳脉冲响应函数合成各个特定方位角的声音, 这在声学界研究得较为广泛; 另一类是基于空间向量分析合成空间声, 这就是本文提出研究的问题. 空间声不仅可用于虚拟环境, 而且还可以应用到各

\* 本文研究得到国家自然科学基金项目基金资助. 作者张爱东, 1969 年生, 博士生, 主要研究领域为计算机多媒体技术, 虚拟环境. 石教英, 1937 年生, 教授, 博士导师, 主要研究领域为多媒体计算, 计算机图形和 CAD, 可视化, 虚拟环境. 潘志庚, 1965 年生, 副研究员, 主要研究领域为多媒体计算, 分布式图形, 虚拟环境.

本文通讯联系人: 张爱东, 杭州 310027, 浙江大学 CAD&CG 国家重点实验室

本文 1995-07-31 收到修改稿

种警示系统中,以空间声的语义信息和方位信息表达警示源的类型和位置,可以使警示系统达到一个更高的层次.

本文要讨论的是建造听觉的真实感环境,也就是用计算机模拟生成真实感的三维空间声.在第1节将讨论真实感空间声的研究现状,第2节讨论模拟真实感空间声的算法,最后是实验结果和结论部分.

### 1 真实感空间声

目前,典型的虚拟环境按其图象显示设备可以分为这样3类:头盔式(HMD),吊杆式(BOOM),洞穴式(CAVE).人的负荷越来越小,设备费用依次提高,临场感、舒适感增强.按照3类不同的显示设备,可以配置不同的声音播放设备:与HMD相配的是耳机;在BOOM上可以装配小型的音箱,随之一起运动;在CAVE中则可于各个墙角固定放置音箱.当然,声音播放设备的选择是与图象显示设备无关的,但是它将影响到算法的效率.下面分别讨论基于耳机和音箱的声音系统.

假定人的听觉系统是一个线性系统,那么从声源到听觉中枢经过了这样一些传输环节:空气、外耳、耳道、鼓膜、听小骨、耳蜗和听神经.只要在任何一个环节测定从声源到该处的传递函数,即该处的脉冲响应函数,那么通过与原始声源信号卷积,可以得到次生声源信号.在该处播放次生声源信号的听觉效果与在声源位置播放原始信号的效果完全一致.通常可以用耳机或音箱作为模拟声音的设备,耳机是处于外耳这一环节,音箱是处于空气环节.声学界对模拟空间声的研究主要是表现水平面上的不同方位角<sup>[2,3]</sup>和中垂面的不同仰角.<sup>[4]</sup>目前基于耳机模拟空间声的方法研究得较多<sup>[5]</sup>,其方法是在左右耳的内耳道外端安装探测麦克风,记录下很多典型位置声音脉冲的双耳响应函数(HRTF).在模拟空间声时,根据其方位角选定对应的双耳脉冲响应函数,进行卷积运算,产生双耳声音效果,从而达到空间声效果.<sup>[6,7]</sup>这种方法主要由环境声模拟和空间声模拟2部分组成<sup>[8]</sup>,图1是用双耳相关函数合成空间声方法的流程.

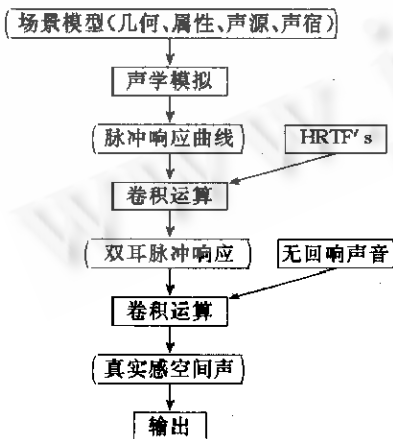


图1 双耳相关函数方法

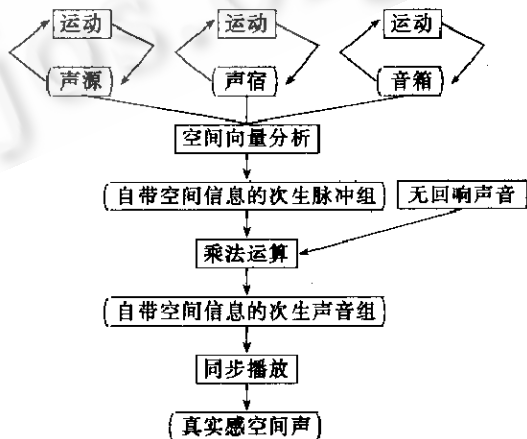


图2 空间向量分析方法

我们认为基于音箱的声音表达方式更容易表现空间声效果,这是因为:音箱可以不只2个,多个声道的表现力更强;音箱距离耳朵远,更接近实际声源,模拟误差相对较小;由于音

箱远离耳朵,本身固有其空间感,而多个音箱结合更可以表现各方向的空间感,从而可以避免十分耗时的双耳脉冲响应函数的卷积运算. 音箱作为声音输出设备与耳机相比,虽然较重,对环境要求高,但对于建造虚拟环境来说,反而比耳机更加舒适,真实感效果更佳,因此基于音箱建造虚拟环境的声音系统是一个更合适的选择.

基于音箱的模拟方法使人同时处于 2 个环境,一个是需要模拟的虚拟环境,另一个是人实际所处的环境. 为了表示空间声的真实感,只有将各方面因素同时放入虚拟仿真环境和实际播放环境加以分析,才能从本质上了解空间声的物理特性,才能严格地描述空间声,从而得到一个模拟产生真实感空间声的方法. 为了全面表达声音的空间感,就必须考虑虚拟环境中声音的反射以及实际环境中声音的反射. 前者是虚拟环境必须表达的,而后者通过增大实际环境中各表面的吸声系数即可忽略. 同时,必须建立一个动态模型,全面地表达真实感空间声需要模拟的这样 5 种情形:运动声源、运动声宿、运动音箱、多个声源、多个反射虚声源. 图 2 是我们提出用空间向量分析方法合成空间声的流程,其中没有包括环境声合成. 基于几何声学的环境声模拟有这样 3 类方法:适合于早期镜面反射的虚声源算法<sup>[9]</sup>;适合于后期规则反射的声线跟踪算法<sup>[10]</sup>;适合于漫反射的改进的辐射度算法.<sup>[11]</sup>

## 2 真实感空间声合成算法

为了描述真实感空间声合成算法,先给出以下定义.

定义 1. 数字声音脉冲是一个二元组:  $\langle A, T \rangle$

其中  $A$  表示脉冲幅度,  $T$  表示脉冲发生时刻.

定义 2. 数字声音是数字声音脉冲的集合:  $\{ \langle A_i, T_i \rangle, i=1, 2, 3, \dots \}$

定义 3. 三维空间声脉冲是一个五元组:  $\langle A, T, X, Y, Z \rangle$

其中  $X, Y, Z$  表示声脉冲发生点的三维坐标.

定义 4. 三维空间声是三维空间声脉冲的集合:

$$\{ \langle A_i, T_i, X_i, Y_i, Z_i \rangle, i=1, 2, 3, \dots \}$$

真实感空间声合成就可以描述成这样一个问题:对每一个三维空间声脉冲  $S \langle A, T, X_s, Y_s, Z_s \rangle$  计算出相应于接收器的三维空间声脉冲  $R \langle A_R, T_R, X_R, Y_R, Z_R \rangle$ , 并且用若干个位置固定的音箱表达这一具有方向、时刻、幅度的空间声脉冲. 我们知道每个音箱也只能表达一个空间声脉冲  $S_i \langle A_{S_i}, T_{S_i}, X_{S_i}, Y_{S_i}, Z_{S_i} \rangle, i=1, 2, 3, \dots$ , 现在就是要用  $S_i$  表达从  $S$  到  $R$  的声音效果. 在以上各量中, 已知各点三维坐标及声源数字声音脉冲, 要求各音箱的数字声音脉冲, 使其合成效果与声源到接收者的效果一致.

三维空间任何一个向量都可以用 3 个线性无关的向量的线性组合表示, 这 3 个线性无关的向量就是三维空间的一组基. 同一个向量在不同的基下有不同的坐标, 而这样的坐标可以用过渡矩阵进行转换. 现在我们给出一个表达空间向量的方法.

笛卡尔坐标系的 3 个坐标轴  $\vec{OX}, \vec{OY}, \vec{OZ}$  构成一组基, 从声源到接收者的向量  $\vec{SR}$  可以用这组基的线性组合表达, 其坐标为  $(X_R - X_S, Y_R - Y_S, Z_R - Z_S)$

选定合适的 3 点  $S_1, S_2, S_3$ , 使得  $R$  不在三角形  $\triangle S_1 S_2 S_3$  内, 这样,  $\vec{RS}_1, \vec{RS}_2, \vec{RS}_3$  线性无关, 构成三维空间的另一组基, 向量  $\vec{SR}$  在这组基下的坐标设为  $(\alpha_1, \alpha_2, \alpha_3)$

从基 $\vec{RS}_1, \vec{RS}_2, \vec{RS}_3$ 到基 $\vec{OX}, \vec{OY}, \vec{OZ}$ 的过渡矩阵 $H$ 为

$$H = \begin{pmatrix} X_1 & X_2 & X_3 \\ Y_1 & Y_2 & Y_3 \\ Z_1 & Z_2 & Z_3 \end{pmatrix}$$

其中对于 $i=1, 2, 3$ , 令 $|RS_i|$ 表示从 $R$ 到 $S_i$ 的距离, 有

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \frac{1}{|RS_i|} \cdot \begin{pmatrix} X_R - X_{S_i} \\ Y_R - Y_{S_i} \\ Z_R - Z_{S_i} \end{pmatrix}$$

向量 $\vec{SR}$ 在2组基下的坐标有如下关系

$$\begin{pmatrix} X_R - X_S \\ Y_R - Y_S \\ Z_R - Z_S \end{pmatrix} = H \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \quad (1)$$

由于 $\vec{RS}_1, \vec{RS}_2, \vec{RS}_3$ 线性无关, 过渡矩阵 $H$ 存在逆阵, 因此

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = H^{-1} \cdot \begin{pmatrix} X_R - X_S \\ Y_R - Y_S \\ Z_R - Z_S \end{pmatrix} \quad (2)$$

这样对于给定各点坐标, 存在一组值 $(\alpha_1, \alpha_2, \alpha_3)$ , 它表示向量 $\vec{SR}$ 在 $\vec{RS}_1, \vec{RS}_2, \vec{RS}_3$ 方向的3个投影分量.

考虑到声音传播的衰减和延时, 有

$$A_R = A_S \cdot \frac{k}{|RS|^2} \cdot \exp(m \cdot |RS|)$$

$$T_R = T_S + \frac{|RS|}{c}$$

其中 $|RS|$ 表示从 $R$ 到 $S$ 的距离,  $k$ 是球面声波传播衰减因子,  $m$ 是空气吸收声能因子,  $c$ 是声速. 现在用3个音箱 $S_1, S_2, S_3$ 表达衰减和延时效果, 有

$$A_R = \sum_i A_{S_i} \cdot \frac{k}{|RS_i|^2} \cdot \exp(m \cdot |RS_i|)$$

$$T_R = T_{S_i} + \frac{|RS_i|}{c}$$

记 $\varphi_R(S_i) = \frac{k}{|RS_i|^2} \cdot \exp(m \cdot |RS_i|)$ , 则有

$$A_S \cdot \varphi_R(S) = \sum_i A_{S_i} \cdot \varphi_R(S_i)$$

$$T_S + \frac{|RS|}{c} = T_{S_i} + \frac{|RS_i|}{c} \quad (3)$$

为了在声宿 $R$ 处产生具有方向 $\vec{SR}$ 的声音, 3个音箱 $S_i$ 在声宿 $R$ 处起作用的声音 $A_{S_i} \cdot \varphi_R(S_i)$ 必须与向量 $\vec{SR}$ 在 $\vec{RS}_1, \vec{RS}_2, \vec{RS}_3$ 方向的3个投影分量 $\alpha_i$ 成正比, 即满足以下关系

$$\frac{A_{S_i} \cdot \varphi_R(S_i)}{\alpha_i} = \text{Constant} \quad (4)$$

由(3)和(4),对于  $i=1,2,3$ ,有  $A_{S_i} = A_S \cdot \frac{\varphi_R(S)}{\varphi_R(S_i)} \cdot \frac{\alpha_i}{\sum_j \alpha_j}$

$$T_{S_i} = T_S + \frac{|RS| - |RS_i|}{c} \tag{5}$$

下面我们讨论如何选择合适的 3 点  $S_1, S_2, S_3$  放置音箱. 由于音箱无法表达负值, 因此, 基  $\vec{RS}_1, \vec{RS}_2, \vec{RS}_3$  的每个分量方向上需要位于接收者两侧的 2 个音箱, 这样在三维空间要 6 个音箱, 且任一时刻每一方向上至少有 1 个音箱是处于关闭状态. 事实上, 分析音箱的表现能力可以知道, 1 只音箱仅可表现从声源到音箱的射线上的声音效果, 2 只音箱可以表现二维扇区内的声音效果, 3 只音箱可以表现三棱锥区域内的声音效果(见图 3), 而 4 只音箱可以将三维空间分为 4 个三棱锥区域, 空间任一点都是其中一个区域中的一点, 因此可以将声源所属的那个三棱锥区域的 3 只音箱激活, 作为表现空间声的 3 个基点, 这样就可以保证  $\alpha_1, \alpha_2, \alpha_3$  均为非负值, 而剩余的 1 只音箱的控制分量为 0. 通过对人的听觉感知敏感程度的分析, 将 2 只音箱分别置于水平面上的左前方和右前方, 另 2 只音箱分别置于中垂面上的后上方和后下方, 各音箱到接收者的距离相等(见图 4).

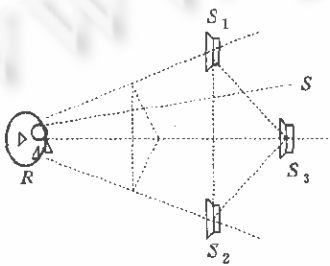


图3 3个音箱的空间组合表现范围

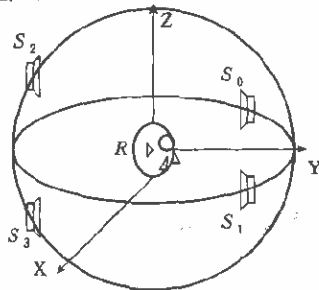


图4 4个音箱的放置

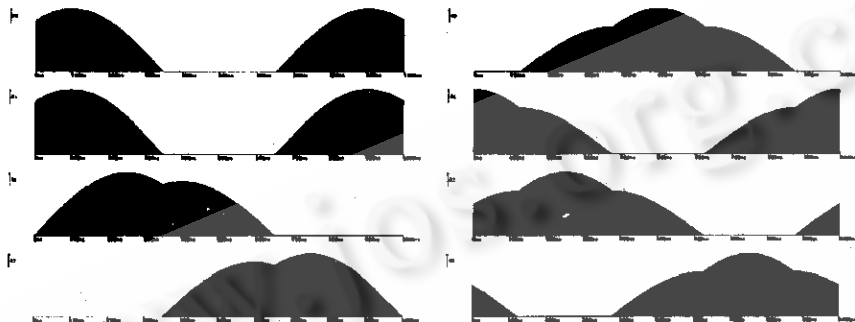
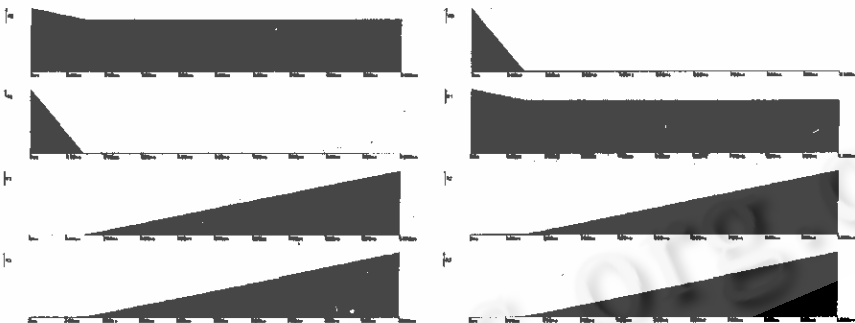
选定合适 3 点放置音箱的标准就是: 若音箱  $S_1, S_2, S_3$  构成的三角形  $\triangle S_1S_2S_3$  与从接收者  $R$  向声源  $S$  方向的射线  $RS$  相交, 则存在  $\alpha_1, \alpha_2, \alpha_3$  均为非负值, 按(5)式控制 3 个音箱, 就可以表现出从  $S$  到  $R$  的空间声方位感.

以上解决了一个声源的空间声合成, 最终的空间声合成必然要包括多个不同位置的声源, 即可以听到来自不同方向、不同内容的声音, 这被称为“鸡尾酒会效果”. 基于对虚拟环境是线性系统的假设, 可以分别按空间向量分析方法计算各个声源的效果, 进行迭加就得到多声源效果. 另外, 在需要模拟虚拟环境中的环境反射声时, 最好的方法是采用声学仿真中的虚声源算法. 该算法是计算声源对各个墙面的镜像, 将镜像位置作为虚声源位置, 虚声源信号的强度为原始声源信号与相应墙面的声音反射系数的乘积, 这样就把环境声问题归结为多声源问题, 从而完整地描述了虚拟环境中真实感空间声的主要因素.

### 3 实验结果

实现这样一个声音系统的关键问题在于实时计算, 这里我们采用限时计算技术<sup>[12]</sup>, 所谓限时计算是指把时间作为评价算法的第 1 因素, 若算法不能在给定时间得出结果, 就认为是错误的. 限时计算的主要思想是以计算结果的精度作为代价, 采用多种形式, 在给定时间

内得到某种精度的结果. 限时计算主要有 3 种技术, 即不同场景层次、不同算法层次以及逐步求精算法. 在空间声系统中, 我们同时用到以上 3 种技术实现了几个演示实例. 图 5 和图 6 分别是声源在  $YOZ$  平面和  $XOZ$  平面上环绕接收者飞行, 图 7 和图 8 是声源在  $XOY$  平面上分别从左右沿一条直线掠过接收者, 图中表示了 4 个音箱的声音幅度控制参数. 根据声音幅度和延时控制参数计算生成 4 个声道的声音数据, 其播放效果十分逼真, 可以用来建造飞行警示系统的声音环境.

图 5 在  $YOZ$  平面上环绕飞行的 4 路声音参数图 6 在  $XOZ$  平面上环绕飞行的 4 路声音参数图 7  $XOY$  平面上从左侧直线飞行的 4 路声音参数图 8  $XOY$  平面上从右侧直线飞行的 4 路声音参数

## 4 结 论

图象和声音是虚拟环境的 2 个最主要的组成部分, 时间真实感和空间真实感构成了真实感的全部内容. 本文对多音箱声音系统表达声音真实感的问题作出了严格的分析, 提出了基于空间向量分析的完整的实现方法, 并在局域网上予以实现. 这一方法自然地解决了运动声源、运动声宿的实际问题, 由于不需要费时的卷积运算, 因此有时间计算多声源效果以及虚拟环境中的环境声效果, 从而获得更强的真实感.

基于音箱实现的声音系统还需要解决这样 2 个问题: 在单机上实现 4 通道声音驱动卡, 以降低系统硬件开销及同步控制开销; 降低实际播放环境中的环境声影响, 以增强播放效果.

## 参 考 文 献

- 1 Durand R. Begaault, 3-D sound for virtual reality and multimedia. Academic Press, 1994.
- 2 Shaw E A G. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. J. Acoust. Soc. Am., 1974, **56**(6):1848~1861.
- 3 Shaw E A G *et al.* Transformation of sound pressure level from the free field to the eardrum presented in numerical form. J. Acoust. Soc. Am., 1985, **78**(3):1120~1123.
- 4 Asano F *et al.* Role of spectral cues in median plane localization. J. Acoust. Soc. Am., 1990, **88**:159~168.
- 5 Wenzel E M. Localization in virtual acoustic displays. PRESENCE, 1992, **1**(1):80~107.
- 6 Wightman F L, Kistler D J. Headphone simulation of free-field listening I; stimulus synthesis. J. Acoust. Soc. Am., 1989, **85**:858~867.
- 7 Wightman F L, Kistler D J. Headphone simulation of free-field listening II; psychophysical validation. J. Acoust. Soc. Am., 1989, **85**:868~878.
- 8 Astheimer P. What you see is what you hear—acoustics applied in virtual worlds. IEEE Symposium on Virtual Reality, San Jose, USA, 1993.
- 9 Borish J. Extension of the image model to arbitrary polyhedra. J. Acoust. Soc. Am., 1984, **75**:1827~1836.
- 10 Vorlander M. Simulation of the transient and steady-state soundpropagation in rooms using a new combined ray-tracing/image-source algorithm. J. Acoust. Soc. Am., 1989, **86**:172~178.
- 11 Shi Jiaoying, Zhang Aidong *et al.* A modified radiosity algorithm for integrated visual and auditory rendering. Computers & Graphics, 1993, **17**(6):633~642.
- 12 Pan Zhigeng *et al.* Time-critical computing in virtual environment. Proceedings of CAD/Graphics'95, Wuhan, 1995. 1078~1083.

## REALISTIC SPATIAL SOUND RENDERING IN VIRTUAL ENVIRONMENT

Zhang Aidong Shi Jiaoying Pan Zhigeng

(State Key Laboratory of CAD&CG Zhejiang University Hangzhou 310027)

**Abstract** Virtual environment is an advanced human-computer interface which use multiple media to generate human's visual and auditory effects. A concept of overall representing temporal and spatial information of sound was provided in this paper. Infinite loudspeakers were addressed in 3D space to represent temporal reality and spatial reality of a sound from any spatial direction. With discussing the essence of spatial sound, a realistic spatial sound rendering algorithm based on 3D vector analysis and a system implementation based on time-critical computation were provided. Comparing with traditional three-dimension sound synthesis method based on HRTF, this method required less computation to generate more realistic sounds.

**Key words** Virtual environment, spatial sound, temporal reality, spatial reality, time-critical computation.