

# 可听化技术综述

张 琼 宋 哲 石教英  
(浙江大学 CAD& CG 国家重点实验室 杭州 310027)

**摘 要** 概述了可听化技术的研究现状及发展方向.在给出可听化技术的分类标准之后,文中重点讨论了在计算机系统中实现可听化系统的一般过程,即通常可分成室内声学仿真、双耳房间脉冲响应的计算及卷积 3 个步骤来实现.另外,文中还介绍了涉及可听化系统实现的主要软硬件产品,分析了该领域今后的研究方向,并指出了其在多个领域广泛的应用前景.

**关键词** 可听化,房间脉冲响应,与头相关的传递函数  
**中图法分类号** TP391

## AN OVERVIEW OF AURALIZATION

ZHANG Qiong SONG Zhe SHI Jiao-Ying  
(State Key Laboratory of CAD& CG, Zhejiang University, Hangzhou 310027)

**Abstract** The article reviews the current research situations on auralization. After the introduction of several classification methods, much attention is paid to the general scheme of the realization of auralization system in a computer system, that is, generally including room acoustical simulation, calculation of the binaural room impulse response and convolution. Furthermore, some principle software/hardware products in auralization domain are introduced and future directions are proposed. At last, it analyzes the possible application domains of auralization.

**Key words** auralization, room impulse response (RIR), head-related transfer function (HRTF)

## 1 概 述

可听化 (Auralization)<sup>[1]</sup>是近年来随着声学仿真技术的长足发展而出现的新名词,它的具体含义是通过对一包含单个(或者多个)声源的声场进行物理或数学建模,以达到声音绘制(audio rendering,或称声学仿真:acoustical simulation)的目的.这样,人们可以获得该声场中任意位置的双耳(bin-aural)听觉感受.

换句话说,可听化技术在客观上主要是模拟特定声场(包括声源,声传播环境以及聆听者 3 要素)中声音传播的物理过程,从而使其中的聆听者作为一个主体能够获得对整个场景声学特性的主观感知.

如果将声源作为信源,聆听者的耳鼓作为信宿,则声信号传播的一般物理过程可用简单的信源-媒介-信宿模型来表示.如图 1 所示.

由信号与系统的有关理论知道,一般的线形系统可以通过该系统在冲激信号激励下的响应来表征,该响应称为系统的脉冲响应或冲激响应.由此,声音从声源传播至人的耳鼓

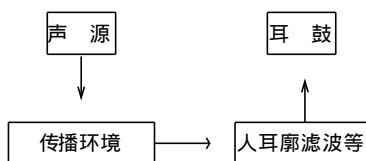


图 1 声信号传播的物理过程

处的过程也可以看作是声信号经过多个线形系统滤波的过程,其中主要包括环境滤波与人的耳廓滤波等.环境的滤波效应在室内可用房间脉冲响应(RIR: Room Impulse Response)来表征,人的耳廓滤波效应则可用与头相关的传递函数(HRTF: Head-Related Transfer Function)来表征.进一步将 RIR 与 HRTF 结合起来,就可以获得双耳房间脉冲响应(BRIR: Binaural Room Impulse Response).如果把 BRIR 作为一 FIR(Finite Impulse Response)滤波器的特征系数,则原始的声信号通过该滤波器的滤波处理即产生了所需的结果.最后可以通过耳机或音箱进行播放.

## 2 可听化系统的分类

按照不同的标准,可听化技术存在着多种分类方法。

如果按照室内声学仿真(即获取房间脉冲响应 RIR的过程)的建模方式来划分,存在着纯物理模型和数学模型的差别。物理模型使用的是真实声场按比例缩小的实物模型(scale model)<sup>[2]</sup>;而数学模型则利用室内声学的一般原理来推导声信号在场景中的传播过程,如基于几何声学的声线跟踪算法(ray-racing)和虚声源算法(image-source)<sup>[3]</sup>,基于波动声学的有限元方法<sup>[4]</sup>与边界元方法等。

如果按照 RIR或 BRIR的获取方式来划分,同样也可以分成两类:一类是通过实际的场景或按比例缩小的模型进行声学测量后直接获取的,这种方式的测量工作量很大,而且只能固定于有限的几个位置,不利于交互式的实现。另一类则利用计算机对数学模型的求解来获得,具有实现方便等显著优点,是今后的发展方向。

如果按照播放手段来划分,一般可分为3类:第1类与第2类同为双声道播放,只是第1类利用的是高保真耳机(即所谓 binaural 方式)<sup>[5]</sup>,这种方式具有实现便捷,成本低廉等特点,更为重要的是,它不受周围场景的限制与影响,尤适于计算机系统中的人机交互;第2类是第1类的变形,只是用两个立体声音箱来取代耳机进行播放,但由于这时引入了左右声道间的串声(crosstalk),所以在播放前需进行串声消除的相应处理(经过串声消除处理的双音箱播放方式称为 transaural)<sup>[6,7]</sup>。这种方式对播放环境的要求比较高,一般需要在消声室(anechoic chamber)中进行,而且对聆听者的位置也有严格的要求;第3类为多音箱的播放方式<sup>[8]</sup>,也可称为环绕声系统(surround array system)播放,通常也在消声室中进行。这种方式的最大优点在于其自然具备的方向感,如听起来后方来的声音确实是来自后方,所以不象前二种方式那样需要进行 BRIR的计算或测量。但它与 transaural方式一样,拥有音箱播放时的一系列缺点,并且代价比较昂贵。

由于用耳机播放的实现方式在计算机系统内的特殊重要性,下面就以数学方法建模,而后用耳机作为播放手段的可听化系统为例来阐述其实现的一般过程。

## 3 可听化实现的一般过程

可听化实现的过程一般可分为室内声学仿真,双耳房间脉冲响应的计算,卷积等3个过程。

### (1) 室内声学仿真——房间脉冲响应的计算

自从 Sabine 提出著名的混响计算公式以来,室内声学在本世纪得到了长足发展<sup>[3,9]</sup>。现在已比较透彻地研究了声信号在一般媒质中的传播机理,在此基础上,人们提出了一系列基于统计声学、几何声学及波动声学的室内声信号传播模型。如基于统计声学的 Sabine, Eyring 混响估算公式;基于几何声学的声线跟踪算法、虚声源模型;以及刚刚发展起来的有限元与边界元计算方法。

基于统计声学的混响计算方法一般来说只用于厅堂的简

单声学性能评估,所以对虚拟环境的声学仿真意义不大。基于几何声学的声线跟踪算法、虚声源模型由于原理简单,算法实现方便而在声学仿真领域获得了广泛应用。但是几何声学关于声信号及声信号传播的一系列前提假设<sup>[3]</sup>使得基于该理论的仿真方法存在着先天的缺陷,不能很好地模拟声音的吸收、散射、衍射、干涉等声学物理现象。基于波动声学理论的有限元方法<sup>[4]</sup>及边界元方法的引入有望较好地解决以上不足。

(2) 双耳房间脉冲响应(BRIR: Binaural Room Impulse Response)的计算。假设已给出了具体的 RIR,则每个 RIR 实际上都代表一条传播路径,其中包括声源直接传播至聆听者的路径,也包括通过一次或者多次发射后到达聆听者的路径。所以 RIR 可以表示为:  $h_i, \theta_i, T_i(t)$ 。其中  $i$  从 0 开始(代表从声源直接传播至聆听者的路径),可以直至无穷大(单纯从理论角度而言)。实际计算中,我们仅可能考虑前  $n$  阶反射或者从时间角度仅考虑前面的  $m$  毫秒;  $h_i$  和  $\theta_i$  分别表示相对于聆听者,声信号入射的偏角和仰角;  $T_i(t)$  则表示该条传播路径的脉冲响应。这里,  $T_i(t)$  包括了传播时延和传播路径上墙面的反射滤波效应,同时也可包括声源在该方向上的指向性因素以及声音在空气中传播时的衰减效应。

如果对应于入射角  $h_i$  和  $\theta_i$  的左右耳的 HRTF 分别为:  $h_{left}(t)$  和  $h_{right}(t)$ , 则该传播路径上的声信号的贡献为: (\* 代表卷积操作)  $T_i(t) * h_{left}(t)$ ,  $T_i(t) * h_{right}(t)$ 。

所以总的 BRIR 可以通过下式计算获得:  $BRIR_{left}(t) = \sum T_i(t) * h_{left}(t)$ ,  $BRIR_{right}(t) = \sum T_i(t) * h_{right}(t)$ 。其中  $h_{left}(t)$  和  $h_{right}(t)$  要随着入射角  $h_i$  和  $\theta_i$  变化而变化。

(3) 卷积。将计算得到的 BRIR 与原始的声信号作卷积操作,即可获得所需的声信号。

## 4 该领域主要的软硬件产品及相关数据

该领域主要的软硬件产品:

(1) ODEON。该软件由丹麦技术大学声学实验室设计,主要用于室内声学仿真。它采用了声线跟踪与虚声源相结合的建模方法,在降低了计算复杂度的同时,又提高了预测精度。

(2) CATT-Acoustic (Computer-Aided Theatre Technique)。该软件分成室内声学建模与可听化实现二个模块,其中在可听化的实现过程中需要 Lake 公司 FDP-1 卷积设备的支持。该软件由瑞典 Chalmers 技术大学应用声学系的 Chalmers 室内声学研究小组设计。

(3) EASE (Electro-Acoustic Simulator for Engineers) - EARS (Electronically Auralized Room Simulation)。该软件也分成室内声学建模与可听化实现二个模块,其中 EASE 是基于 PC 平台的室内声学仿真模块; EARS 是基于 EASE 实现的可听化模块,它需要一些硬件上的支持,比如可用 Media Vision 或 Ariel 公司生产的一些 PC 插件。该软件由德国 ADA Acoustic Design Ahnert 公司设计。

(4) 高档的 Huron 系列大规模卷积设备及低档的 CP4 系列卷积设备。由澳大利亚 Lake 公司设计生产,适用于实时要求较高的各种场合,可直接与 PC 机相连,并且有丰富的软件

支持.

公开或半公开的 HRTF数据:

(1) Shaw, E. A. G., and Vaillancourt, M. M. (1985).

Transformation of sound-pressure level from the free field to the eardrum presented in numerical form, Journal of the Acoustical Society of America, 78 1120-1122.

在该文中,作者以表格的形式给出了耳平面上的 HRTF数据,但只包含幅值信息,而无相位数据.在实现时,需要人为加入左右声道间的组延时信息.

(2) KEMAR-HRTF由 MIT媒介实验室(Media Lab.)

Machine Listening Group的 Gardner, W. G.和 Martin, K.提供<sup>[10]</sup>.该组数据利用假人头模型测得,以时域脉冲采样的形式提供,可从 Internet上获取.

(3) 美国 Wisconsin-Madison大学 F. Wightman和 D.

Kistler可提供多组采用实际个人测得的 HRTF数据.

## 5 可听化技术目前的发展与研究方向

(1)质量的验证与评估.可听化系统质量的验证与评估是一个亟待解决的课题<sup>[1]</sup>,其中包括主观指标与客观指标的确定,定性的评估与定量的分析.该问题的解决将为可听化的进一步应用奠定更为坚实的理论基础.

(2)降低计算量问题.由于可听化过程牵涉到巨大的计算量,使其在计算机系统上的交互实现带来了困难.如采用专用的硬件设备,如 Hiron等,价格又过于昂贵.所以,计算量的降低今后与将来都是可听化系统的实现所必须考虑的中心问题.

(3)播放手段的选择.在计算机系统中,采用耳机作为播放手段无疑是最明智的选择.但用耳机播放时,一般存在着头内定位(in-head localization)和前后听觉感知的二义性(front-back ambiguity)等固有问题,在一定程度上会影响人机交互的效果.采用个人的 HRTF数据或采用头部跟踪设备可以较好地解决以上问题<sup>[11]</sup>.

但是随着 DSP技术日新月异的发展及硬件芯片价格的不断降低,双通道及多通道音箱播放的方式会越来越受到人们的青睐.

(4) RIR与 BRIR的计算问题.基于几何声学的室内声学仿真方法经过多年的研究,现在已经逐步走向成熟.但其理论上的一系列前提假设使其存在着先天缺陷.更精细的模型还要考虑声音的散射、衍射以及空气衰减等因素<sup>[12,13]</sup>.在计算 RIR时的另一个问题是后期反射声(即高阶反射声,又可称混响)的处理.由于后期反射声一般与方向的关系不是很大,所以如果对高阶的反射声不象低阶反射声那样逐个进行 HRTF处理,则整个算法的计算复杂度将会有一定程度的降低.

基于理论声学的波动方程,由于很难确定其复杂的边界条件,所以一般不可能获得解析解;但基于有限元及边界元方法的数值解法近几年引起了不少学者的关注<sup>[4]</sup>.但其庞大的计算量仍然限制了在实际中的应用.

在计算 BRIR时,如果假设接收点是球形的,而不是使用真正的 HRTF数据,则其计算过程可以得到进一步的简化.

(5) RIR与 BRIR计算的接口的标准化问题.由于 RIR的计算方法有许多种,所以 RIR与后续的 BRIR计算之间存在着接口的标准化问题.<sup>[14]</sup>中作者提出了一种文件交换格式.

## 6 可听化技术的应用

可听化技术在许多领域有着广泛的应用.(1)作为声学设计的辅助工具与评价手段.室内声学的计算机仿真技术已经成为建筑声学设计及厅堂音质评价的有力手段<sup>[9]</sup>.如果在厅堂建设之前就能聆听到其大致的音效,无疑将进一步提高设计效率,缩短设计周期,并且切实保证设计的成功.

(2)在虚拟环境中的应用.我们知道,视觉和听觉是人类感知外部世界的两种最重要的信息渠道,其信息载体分别为图象和声音.为了创造出符合人类感知习惯的虚拟环境,一方面需要提供多种媒体的信息表达方式;另一方面,更重要的是要提高各种媒体的表达效果.计算机图形学为人类在虚拟环境中的视觉感知与交互提供了有力的技术支持.对于应用较晚的听觉感知与交互手段,则需要可听化技术的相应支持<sup>[11,15]</sup>.运用可听化技术,在虚拟现实系统中能构造出非常逼真的自然声场效果,从而增强用户的沉浸感并改善人机交互的效果.

(3)在语音通讯系统中的应用.在语音通讯系统中适当地运用可听化技术,将增强语言的易懂度<sup>[16]</sup>.比如在远程会议系统(teleconference)中采用该项技术,可以增强与会者的临场感,改善语音通讯的效果.

(4)在可视化系统中的应用.可视化是多维数据表示的一种有效方式,显而易见,可听化技术同样也是一种表示数据的手段,可作为可视化系统的辅助与补充<sup>[17]</sup>.

(5)在游戏、教育、娱乐中的应用.现在的许多游戏软件给用户提供了非常逼真的视听环境,其中声环境的构造就需要可听化技术的支持.如果在教育等软件中也加上可听化技术的支持,无疑会提高教学质量.

另外,可听化技术在电声设计、厅堂扩声等方面都有着重要的应用.

## 7 结 论

可听化技术是由于计算机技术与声学仿真技术的长足进步而得以发展的,但是其实现过程中所牵涉的巨大计算量,不仅需要高性能计算机的支持,更需要算法效率的进一步改进.此外,可听化技术的有效性问题的客观点量的标准一直没有得到很好的解决,这也在一定程度上制约了它在各个领域中的广泛应用.

## 参 考 文 献

- 1 Kleiner M., et al. Auralization—An overview. Journal of the Audio Engineering Society, 1993, 41(11): 861-875
- 2 Xiang N., Blauert J. Binaural scale modeling for auralization prediction of acoustics in auditoria. Journal of Applied Acoustics, 1993, 38 267-290

- 3 Cremer L, Mülle H. The Design Principles and Its Application of Room Acoustics. Translated by Wang Ji-Qing *et al*, Shanghai Tongji University Press, 1995  
( Cremer L, Mülle H. 王季卿等译. 室内声学设计原理及其应用. 上海: 同济大学出版社, 1995)
- 4 Wright J R. An exact model of acoustic radiation in enclosed spaces. Journal of the Audio Engineering Society, 1995, 43 (10): 813– 820
- 5 Begault D R. 3-D Sound for Virtual Reality and Multimedia. New York: Academic Press, 1994
- 6 Gardner W G. Transaural 3-D audio. Technical Report# 342, MA: MIT Media Lab, 1995
- 7 Gardner W G. 3-D audio using loudspeakers [Dissertation]. MA: MIT Media Lab, 1997
- 8 Tabata A. Auralization of computer simulated field and measured field by a multi- loudspeaker system. Journal of the Audio Engineering Society, 1993, 40 1045
- 9 Bradley J, Soulodre G. The acoustics of concert halls. Physics World, 1997, 10(5): 33– 37
- 10 Gardner W G, Martin K D. HRTF measurements of a KEM AR. Journal of the Acoustical Society of America, 1995, 97(6): 3907 – 3908
- 11 Burgess D A, Verlinden J C. A first experience with spatial audio in a virtual environment. Technical Report, Georgia: GIT GVU center, 1993
- 12 Nelisse H, Nicolas J. Characterization of a diffuse field in a reverberant room. Journal of the Acoustical Society of America, 1997, 101(6): 3517– 3524
- 13 Dalenback D I L. Room acoustic prediction based on a unified treatment of diffuse and specular reflection. Journal of the Acoustical Society America, 1996, 100(2): 899– 909
- 14 Moller H. Interfacing room simulation programs and auralization systems. Journal of Applied Acoustics, 1993, 38 333– 348
- 15 Zhang Q, Shi J Y. Audio rendering in virtual reality. In Proceedings of CAD/Graphics 97, Beijing: International Academic Publishers, 1997, 123– 126
- 16 Zurek P M. Binaural Advantages and Directional Effects in Speech Intelligibility. In: Acoustical Factors Affecting Hearing Aid Performance. MA Allyn and Bacon, 1993
- 17 Rossiter D, Wai-Yin Ng. A system for the complementary visualization of 3D volume images using 2D and 3D binaurally processed sonification representations. In Proceedings Visualization 96, 1996, 351– 354