

三维音频技术综述

殷福亮, 汪林, 陈喆

(大连理工大学 电子与信息工程学院, 辽宁 大连 116023)

摘 要: 论述了三维音频的基本原理, 从头相关传递函数 (HRTF) 的测量、计算、插值和模型逼近以及串声消除算法与稳健性等方面, 综述了信号处理方法在三维音频技术中的应用, 总结了当前三维音频技术中的一些热点问题。最后, 对三维音频技术的未来研究方向进行了展望。

关键词: 三维音频; 信号处理; 头传递相关函数; 串声消除

中图分类号: TN912

文献标识码: A

文章编号: 1000-436X(2011)02-0130-09

Review on 3D audio technology

YIN Fu-liang, WANG Lin, CHEN Zhe

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116023, China)

Abstract: The principle of 3D audio technology was introduced, the application of signal processing methods was reviewed in 3D audio from the measure, computation, interpolation, approximation of head-related transfer function (HRTF) and the methods for crosstalk cancellation, summarized currently hot topics in this area. Finally, the future research trends of 3D audio were discussed.

Key words: 3D audio; signal processing; head-related transfer function; crosstalk cancellation

1 引言

在日常生活中, 人们常利用双耳听觉处理信息, 比如在某个聚会中, 有很多人在同时讲话, 但人们总是能够轻松地与某个特定的人交谈而不受别人影响, 这种“鸡尾酒会效应”就是因为人的双耳能够感知声源的空间位置差异而采取不同的处理。三维音频 (3D audio), 也称为虚拟声 (virtual acoustics)、双耳音频 (binaural audio)、空间声 (spatialized sound) 等, 它根据人耳对声音信号的感知特点, 使用信号处理方法对声源到两耳之间的传递函数进行模拟, 以重建复杂三维虚拟空间声场^[1, 2]。三维音频技术具有结构简单、

易于实现、重建听觉三维空间真实自然的优点, 因此在多媒体通信、虚拟现实、人机交互、消费电子、数字娱乐、心理声学等领域有着广泛的应用前景。但是, 为了得到逼真的空间声音效果, 虚拟声场的重建通常需要巨大的计算量和存储量, 这是制约三维音频技术发展的瓶颈问题。近年来, 随着计算机技术的发展和 DSP 处理能力的提高, 三维音频系统的实现也变得容易, 因而成为音频处理领域的热门研究课题之一, 国内外一些从事声学、信号处理、计算机技术方面的学者正在进行相关研究。本文将综述三维音频的基本原理及信号处理方法在三维音频技术中的应用。

收稿日期: 2010-08-23; 修回日期: 2010-11-09

基金项目: 国家自然科学基金资助项目 (60772161, 60372082); 高等学校博士学科点专项科研基金资助项目 (200801410015)

Foundation Items: The National Natural Science Foundation of China (60772161, 60372082); The Specialized Research Fund for the Doctoral Program of Higher Education of China (200801410015)

2 三维音频的基本原理

三维音频的基本原理是通过模拟空间某点声源在两耳产生的声场，使听者感觉虚拟声源从空间对应位置发出，因此该技术又称为双耳声学技术。为了理解双耳声学技术，首先必须了解人耳定位的基本原理。空间某点声源发出的声音传递到人的两耳，会有时间差（ITD, interaural time difference）和强度差（ILD, interaural level difference），并且声音在传递过程中会与人的躯干、肩膀、头部以及耳廓发生一系列的相互作用，使声音的音质改变，人耳则根据到达两耳声音的时间、强度和音质的差别来确定声源的位置。上述声源定位信息，都可以用一个传递函数来表示，称为头相关传递函数（HRTF, head-related transfer function）。由于 HRTF 包含了有关声源位置的大部分信息，因此一组准确的 HRTF 函数对精确重现三维声场有重要意义^[3]。

在自由场情况下，左耳和右耳的 HRTF 定义为

$$\begin{cases} H_L = H_L(r, \theta, \varphi, f) = P_L(r, \theta, \varphi, f) / P_0(r, f) \\ H_R = H_R(r, \theta, \varphi, f) = P_R(r, \theta, \varphi, f) / P_0(r, f) \end{cases} \quad (1)$$

其中， P_L 、 P_R 分别为声源在倾听者左、右耳产生的声压； P_0 是将被测者移走后，声源在原头部中心位置处产生的声压。由式(1)可以看出， H_L 和 H_R 是声源的水平角 θ 、仰角 φ 、声源到头部中心距离 r 以及声波频率 f 的函数。

HRTF 有 2 个特点，即随着声源位置的不同，HRTF 也不相同，并且每个人的 HRTF 都不尽相同。如果让单路信号通过一组 HRTF，则会产生人的两耳定位所需的必要信息，再通过耳机播放后，听者会感觉声音是从 HRTF 所描述的空间位置发出的。这个处理过程称为双耳声合成，所产生的信号称为

双耳声。

三维音频的实现可以分为两步进行：双耳声合成（synthesis）和双耳声重拾（reproduction）。双耳声合成就是使用 HRTF 产生双耳声信号的过程。单路信号 S 经过 HRTF 滤波器后，产生的双耳信号 X_L 和 X_R 分别为

$$\begin{cases} X_L = H_L S \\ X_R = H_R S \end{cases} \quad (2)$$

其中， H_L 和 H_R 分别为左耳和右耳的 HRTF。

双耳声重拾是指将合成的双耳声信号传递到人耳的过程，该过程必须保证两耳只能接收到各自声道的信号。双耳声重拾有 2 种方法，一种是通过耳机，该双耳声信号直接从耳机播放到人耳，如图 1 所示。另一种方法是使用扬声器，声音从扬声器播放后传递到人耳，这会产生串声（crosstalk）干扰，因此需要额外的串声消除过程。使用立体声扬声器的双耳声重拾过程如图 2 所示。

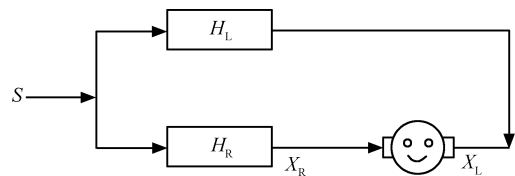


图 1 使用耳机的双耳声重拾

使用耳机播放双耳声信号时容易保证左右声道信号的分离性，但使用扬声器播放时，在扬声器与双耳之间存在着严重的串声现象，即左扬声器的声音有很大一部分被右耳听到，右扬声器的声音有很大一部分被左耳听到，这就使得两耳接收到的空间定位信息严重失真，不能重现所预想的虚拟声场。在图 2 中，扬声器信号经过电声传递网络 G 到达两耳，其中路径 G_{12} 和 G_{21} 产生的信号为串声干

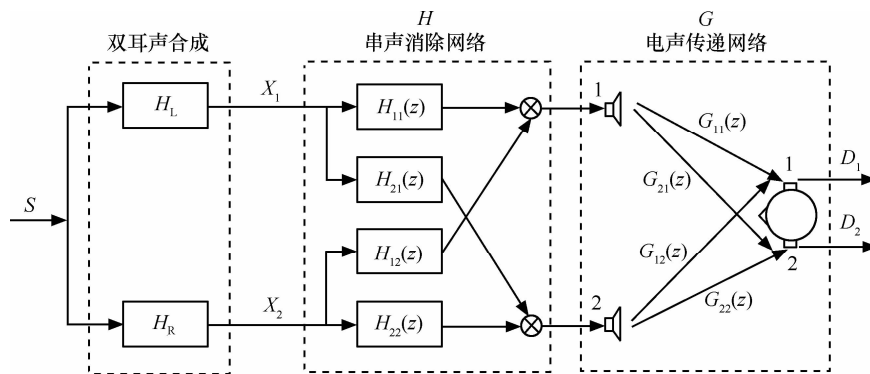


图 2 使用扬声器的双耳声重拾

扰。串声干扰严重降低了双耳声系统的空间定位性能,因此需要设计串声消除网络 H ,使得串声在人耳处能自动消除。在图 2 中,由输入信号 S 通过 HRTF 后,得到的双耳声信号为 $X=[X_1, X_2]^T$,经过扬声器播放后,人耳最终接收到的信号为 $D=[D_1, D_2]^T$,电声传递网络 G 和串声消除网络 H 分别为

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \quad (3)$$

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \quad (4)$$

串声消除就是使人耳接收到的信号与双耳声信号相同,即

$$D = GHX = X$$

于是有:

$$GH = I \quad (5)$$

其中, I 为单位矩阵。对式(5)进行求解,即可得到串声消除网络 H 。

通过对三维音频工作原理的分析可以发现,精确的 HRTF 函数和串声消除网络,对三维音频系统的实现有着非常重要的意义。实际上,近些年,对三维音频技术的研究也基本围绕着这 2 个问题展开。

3 头相关传递函数

3.1 HRTF 的测量与计算

HRTF 通常由实验测量方法得到。测量在消声室中进行,测量对象可以是人工头或真人,在测量时,将微型麦克风放置在被测者耳道中,在相对头部特定的水平角、仰角和距离的位置用扬声器播放已知的冲激信号,由麦克风接收到达人耳的声音,然后使用系统辨识方法获得声音传递系统的传递函数 HRTF^[4]。冲激信号可以是简单的冲激序列、伪随机二进制序列或格雷码。由于测量到的数据通常会包含测量系统或耳道的其他额外响应,所以要使用自由场均衡方法或散射场均衡方法对测量数据进行均衡处理,以去除这些影响^[5]。有些机构对 HRTF 进行了实际测量,并将部分数据公布在互联网上。例如 MIT 的 Gardner 等于 1994 年对 KEMAR 人工头的 HRTF 进行测量,得到了包含仰角 $-40^\circ \sim 90^\circ$ 共 710 个位置的 HRTF 数据库^[5];加州大学 Davis

分校 CIPIC 实验室的 Algazi 和 Duda 等于 2000 年在 25 个不同水平角和 50 个不同仰角的 1250 个位置对 45 个真人的 HRTF 进行测量,得到了 CIPIC 数据库,同时还提供每位被测者的 27 种人体测量参数(头部和肩膀 17 个参数,耳廓 10 个参数)^[6]。此外,国内华南理工大学的谢波荪等于 2006 年对中国人的 HRTF 数据进行了测量^[7]。

实验测量方法获得的 HRTF 数据虽然精确,但实验过程非常繁琐,成本很高,为此,一些学者提出通过声学模型计算的方法来得到 HRTF。从物理学角度讲,计算 HRTF 就是求解耳廓、头部、躯干等组成的系统对声波的散射问题,即在一定的边界条件下求解波动方程^[8-10]。虽然采用理论计算的方法可以避免繁琐的实验过程,但是其计算涉及人头、躯干及耳廓的几何描述,因此采用何种指标进行描述以及如何正确描述,仍是难点。通常,通过理论计算只能得到 HRTF 的近似值。

由于每个人的 HRTF 都不尽相同,当听者使用本人的 HRTF 时,称为个性(individualized) HRTF;如果使用他人或人工头的 HRTF,则称为非个性(non-individualized) HRTF 或通用(generalized) HRTF。由于实验条件的限制,不可能为每个听者都准备个性 HRTF,通常都是以通用 HRTF 作为替代,但通用 HRTF 会产生较大的空间定位误差,并以前后混淆和仰角误差最为显著,这是三维音频技术的瓶颈问题之一^[11]。为此,通用 HRTF 的个性化方法受到广泛关注。心理声学认为, HRTF 中某些频段对声音定位起关键作用,据此 Tan 和 Park 分别提出了通用 HRTF 频谱的修改方法来减少定位误差,但由于听者个体之间的差异较大,该类方法效果有限^[12,13]。一些学者提出主观选择方法,即通过听者的主观听觉效果比较,从包含多位测量者 HRTF 的数据库中,为听者选择一套最适合本人的 HRTF,并作为近似的个性 HRTF。在个性 HRTF 难以获得的情况下,近似个性 HRTF 则是一种较好的替代方案^[14-16]。此外,一些学者还研究了 HRTF 与外耳(包括头、肩、耳廓)尺寸之间的关系,希望能利用外耳尺寸参数来选择 HRTF^[17-19]。

3.2 HRTF 的低阶模型逼近

通常,测量获得的 HRTF 所对应的 FIR 滤波器阶数在 200~512 之间,甚至更高,这样高的阶数使得实时合成复杂的声学环境很困难,因此有必要使

用低阶的全零点模型(FIR)或极零点模型(IIR)对其逼近^[20]。在低阶模型逼近之前,通常将HRTF表示为最小相位函数,这样做的好处如下:

- 1) 可以直接由幅度谱求得最小相位滤波器的相位;
- 2) 最小相位滤波器具有最小群时延特性和最小能量时延特性。

Kulkarni 等对 HRTF 函数作最小相位逼近,分析了人耳对声音相位的敏感度,发现人耳对原始 HRTF 和最小相位重建 HRTF 的区别不敏感^[21]。

3.2.1 HRTF 的全零点模型逼近

使用低阶 FIR 滤波器逼近测量 HRTF 的最直接方法是窗函数法^[20],该方法对原始冲激响应加矩形窗,以获得所需阶数的滤波器。Sandvad 等对不同的加窗方法进行了讨论^[22],发现使用矩形窗会在幅度响应不连续的地方产生“吉布斯”现象。Kulkarni 等基于对数幅度误差最小化准则,使用加权最小二乘(WLS)方法设计 FIR 滤波器^[23],但由于阶数限制,其效果不太理想。

有些学者提出将 HRTF 函数表示为若干 FIR 函数的加权和的形式,称之为级数分解方法^[24]。Kistler 等使用主元分析方法(PCA, principal components analysis)逼近最小相位 HRTF,将 HRTF 幅度谱分解为若干主频谱成分的加权和,该方法重建的 HRTF 失真较小^[25]。Chen 等提出了特征提取(SFER, spatial feature extraction and regularization)的方法,该方法对 HRTF 频谱进行 Karhunen-Loeve 变换后,得到其特征传递函数,这样可将 HRTF 频谱分解为若干个特征传递函数的加权和^[26, 27]。上述几种级数分解方法充分利用了 HRTF 数据的相似性,仅用若干个加权值来表示 HRTF 函数,因此与原始数据相比,其所需存储量大幅度减少。文献[28]提出了 HRTF 数据的混合压缩方法,该方法联合使用主元分析、矢量量化和曲面拟合方法对 HRTF 数据进行压缩。

3.2.2 HRTF 的极零点模型逼近

由于 HRTF 的频谱变化比较剧烈,仅用少量的零点很难将其完全表示出来,因此使用极点和零点共同表示频谱变化是最有效的方法,这也是近些年来三维音频研究的一个热点。HRTF 的极零点模型逼近大致可分为如下 4 类。

- 1) 传统的极零点建模方法。如 Prony 算法^[29]和 Yule-Walker 算法^[22]等是极零点建模的典型方法,但这些通用算法在建模过程中没有考虑 HRTF 数据

本身的特性。Haneda 等使用共极点/零点模型(CAPZ, common acoustical pole/zero)来逼近 HRTF, CAPZ 模型由不随声源位置变化的极点和随声源位置变化的零点共同表示 HRTF^[30],与传统的极零点模型相比, CAPZ 模型需要更少的参数,且拥有更高的计算效率。

- 2) 状态方程建模方法。该类方法将 HRTF 表示为状态差分方程的矩阵形式,然后对矩阵的 Hankel 范式作奇异值分解,仅保留较大奇异值所对应的成分,从而得到 HRTF 的逼近形式。该类方法将 HRTF 频谱中不重要的成分丢弃,仅保留最重要的成分,具有较好的建模效果。Mackenzie 等提出的基于平衡模型截断(BMT, balanced model truncation)的方法^[31], Georgiou 等提出的状态空间模型方法^[32],均属此类方法,它们都是对单个 HRTF 分别建模。Grantham 等则将整个 HRTF 数据集组合成一个大的 Hankel 矩阵,对其作奇异值分解,最终得到具有共极点的状态空间模型^[33]。

3) 基于人耳听觉特性的建模方法。HRTF 与人耳特性密切相关,因此一些学者提出了基于人耳听觉特性的建模方法。通常人耳对对数幅度误差更为敏感,据此 Blommer 等提出了基于对数幅度误差最小化的拟牛顿逼近算法^[34]; Kulkarni 等提出了基于对数幅度误差最小化的加权最小二乘(WLS)方法^[35]; Durant 等使用遗传算法来设计基于对数幅度误差最小化的极零点模型^[36]。人耳的另一听觉特性是使用非均匀的频率尺度来处理听觉信息,据此, Huopaniemi 等提出了在非线性的频域(warped frequency)内设计 FIR 和 IIR 滤波器,以逼近 HRTF^[37]。

4) 物理建模方法。物理模型逼近是指将 HRTF 函数与声波传播到耳道时发生的一系列物理现象联系起来,通过使用与这些物理现象相对应的数学模型来逼近 HRTF。Genuit 等提出了表征人体肩、头、耳廓形状的 34 个测量参数,并使用该模型来逼近 HRTF^[38]。Brown 等提出了结构化模型方法,该方法使用一个简化的时域物理模型来表示 HRTF 函数,将 HRTF 表示为 IIR 头部阴影模型、FIR 耳廓回声模型和 FIR 肩部回声模型的组合^[39]。

关于 HRTF 的极零点建模方法的研究持续了很多年,也提出了很多方法。通常,基于人耳特性的建模方法由于考虑了心理声学因素,其主观听觉效果较好;而物理建模方法则能够使用较少的参数表征出与 HRTF 相关的物理现象。状态方程方法由于其

结构规则,有利于集成电路实现,在计算量很大的复杂虚拟声场模拟场合具有更大优势^[40]。

3.3 HRTF 的插值

由 HRTF 测量过程可知, HRTF 只能在有限离散位置获得。当声源或听者头部移动时,需要对已知的 HRTF 进行插值,以获得任意位置的 HRTF,从而使不同位置 HRTF 之间的切换能快速平滑地实现^[41]。根据插值对象的不同,可以将插值方法分为全零点模型插值和极零点模型插值。

3.3.1 全零点模型插值

通常,测量得到的 HRTF 为全零点模型,因此 HRTF 插值研究通常是针对全零点 HRTF 模型。常用的插值方法可以分为 2 类:直接插值方法和间接插值方法。

1) 直接插值方法是对目标点周围 2 个或多个最近的相邻点进行插值得到目标 HRTF,典型的插值方法有使用 2 个相邻点的线性插值方法和使用 4 个相邻点的双线性插值方法^[1]。文献[42]提出了使用 3 个相邻点的相对位置传递函数来实现插值。文献[43]提出了分段线性插值方法,该方法将整个频域分为若干个子频段,在每个子频段上分别进行线性插值。直接插值方法比较简单,但由于其只利用周围相邻点的信息,因此插值精度有限。

2) 间接插值方法是用全部已知的 HRTF 数据来估计出 HRTF 的空间变化函数,然后根据目标位置的空间坐标计算目标 HRTF。间接插值方法有多种,如样条插值方法^[44]、有理数多项式插值方法^[45]、空间频率响应曲面插值方法^[46]、Plenacoustic 函数插值方法^[47]、基于主元分析插值方法^[48]等。间接插值方法可利用所有已知信息,因此其效果通常要优于直接插值方法,但其计算量和存储量消耗较大。

3.3.2 极零点模型插值

在双耳声合成过程中,通常采用极零点 HRTF 模型,以提高计算效率,此时 HRTF 的插值对象是极零点模型,但由于极零点模型建模方法的多样性,目前还没有比较成熟的插值方法。常用的极零点模型插值方法是对相邻 HRTF 的零点和极点分别线性组合来得到目标 HRTF 的零点和极点,该方法的插值精度较差^[49,50]。文献[51]给出了针对 CAPZ 模型的线性插值方法,由于 CAPZ 模型共享同一组极点,该方法仅需对零点进行插值。文献[52]给出基于全零点模型插值的间接插值方法。

4 串声消除算法

使用耳机很容易将双耳声传输到人耳,但是长时间戴耳机会使人不适,而且会产生颅内定位现象,导致听觉疲劳,因此人们更倾向于使用扬声器重拾双耳声。但每个扬声器发出的声音都会被两耳听到,这会严重影响声源定位效果,所以必须要做串声消除处理。串声消除的概念首先由 Bauer 在 1961 年提出^[53],然后由 Schroeder 和 Atal 应用到实际中^[54],此后又出现了许多方法。串声消除系统的计算方法可以分为 2 类:直接计算方法和自适应计算方法。

4.1 串声消除网络的直接计算方法

直接计算方法是在已知扬声器到人耳的电声传递函数 H 的情况下,利用 H 直接设计串声消除滤波器,通常有时域设计方法和频域设计方法。时域设计方法的优点是精度高,但计算量较大;而频域设计方法则相反,计算量较小,但精度较低。文献[55]提出了基于正则化方法的频域快速卷积算法,该方法分别在每个频率点设计串声消除器,因而简单有效,具有实用价值,但当串声消除滤波器的阶数与电声传递函数的阶数相近时,串声消除器的性能会由于 FFT 的圆周卷积效应而急剧下降。因此,当串声消除滤波器的阶数较低时,人们更倾向于使用时域设计方法。常用的时域设计方法有最小二乘(LS)方法,该方法简单直接,但计算量较大^[56-58]。此外,还有维纳滤波方法^[59],最大误差最小化方法^[60]等。

直接计算方法的优点是简单直接,但是它需要事先已知电声传递函数,以更新串声消除滤波器,这就需要确定头部与扬声器之间的相对位置,从而引入了头部跟踪问题。对移动的倾听者,通常使用传感器连续监控倾听者头部的位置,并根据头部位置实时更新电声传递函数和串声消除滤波器。此处传感器可以使用声学的、机械的、光学的或者磁电的。文献[61, 62]实现了基于摄像机的头部跟踪系统,该系统使用视频摄像机监控倾听者的位置,再使用图像处理方法来进一步提高跟踪精度。文献[63]利用倾听者发出的声音,由波达方向估计方法计算出倾听者位置,并实时更新。

4.2 串声消除网络的自适应计算方法

直接计算方法需要准确地知道系统的电声传递函数 H ,但声音的传播常受一些不确定因素影响,

因此 H 很难获得。自适应方法可以在给定的听觉空间内计算串声消除滤波器,并根据 H 的变化自适应地更新串声消除滤波器的系数。LMS 算法由于其计算简单和良好的稳健性,在串声消除滤波器设计中经常使用^[64-66],并出现了多种改进算法,如多误差 Filtered-x LMS 算法^[67,68]、频域 LMS 算法^[69]、频域多误差 Filtered-x LMS 算法^[70]等。针对 LMS 算法收敛较慢的问题,一些学者使用了收敛快但计算量较大的 RLS 算法。文献[71]使用频域快速 RLS 算法,以提高收敛速度。自适应方法的主要缺点是需要在人耳处放置 2 个微型麦克风,用来采集到达人耳的声音,以作为自适应滤波器的反馈信号,这样就会对采集设备的精度要求高,增加成本,同时也会令听者感到不适。

4.3 串声消除系统的稳健性

串声消除滤波器通常是在听者的头部固定不动的情况下计算得到,该头部位置即为串声消除系统的“甜点”(sweet spot),即此处的听觉效果最好。但实际中,即使头部偏离“甜点”几厘米,也会产生较大失真,因此需要实时跟踪头部位置,同时有必要增大甜点范围,以增强串声消除系统的稳健性。目前增强串声消除系统稳健性的方法主要有 2 种,一种是通过调整扬声器的位置来保证电声传递路径矩阵的良好性能,以增强系统的稳健性。通常,传统的扬声器是 30° 放置的,而 Kirkeby 等提出了立体声极子对(stereo dipole)概念,将 2 个扬声器相对人头 $\pm 5^\circ$ 放置,以增强串声消除系统的稳健性^[72]。Ward 和 Takeuchi 也分别证明了扬声器的摆放位置对串声消除系统的稳健性有很大影响^[73,74]。另一种增强串声消除系统的稳健性的方法是使用多个扬声器传递双耳声(扬声器个数 >2),例如,文献[75]和文献[76]分别实现了使用多个扬声器的稳健串声消除系统。

5 总结与展望

双耳声信号包含了大部分有关声音的空间信息。三维音频技术使用信号处理的方法产生双耳声,并通过耳机或者扬声器重放,为听者再现真实自然的空间声场。近些年来,有关三维音频技术的研究已经取得了很大的进展,但在算法性能和计算复杂度等方面仍有一些问题有待于深入研究。

1) 通用 HRTF 的个性化问题。

由于 HRTF 包含了有关声源位置的大部分信

息,因此一组准确的个性 HRTF 函数对精确重现三维声场有重要意义。但在实际应用中,不可能为每个听者准备个性 HRTF,而使用通用 HRTF 则会产生较大的定位误差,严重影响三维声场重建效果,这是三维音频技术的瓶颈问题之一。目前关于通用 HRTF 个性化问题的研究虽然取得了一定的进展,但由于人耳听觉系统的复杂性,其效果不是非常理想。

2) 多声源和混响情况下的双耳声合成。

为了保证逼真的空间声音效果,在多声源和混响情况下重建虚拟声场需要巨大的计算量和存储量,这对三维音频系统的处理能力提出了很高的要求。如何根据人耳特性,在不影响三维声场重建效果的前提下,对双耳声合成算法进行简化,以减小计算复杂度,这是未来研究的一个重要方向。

3) HRTF 数据处理算法需要符合人耳听觉的客观评价标准,以衡量其有效性。

目前的 HRTF 处理效果主要依赖于主观评价,存在成本较高、灵活性不够、可重复性和稳定性较差、受人的主观影响较大的缺点。为了克服主观评价的缺点,有必要根据人耳听觉特性,制定出方便、快捷的客观评价方法。

4) 头部跟踪问题。

在三维音频系统中,通常需要精确检测听者头部位置,以实时更新 HRTF 和串声消除器。如何提高头部跟踪系统的舒适性和头部定位的准确度,降低系统的成本,也是三维音频实用化的一个重要课题。

5) 实时稳健的串声消除算法。

目前的串声消除算法主要处于理论研究阶段,其计算复杂度高,且需要头部跟踪器等辅助设备,其效果受听者头部移动和混响的影响较大。如何开发出简单易行、实时稳健的串声消除算法,是三维音频技术中的重要研究领域。

总之,三维音频技术在虚拟现实、心理声学研究和消费电子等领域有着广泛的应用前景,因此成为近年来音频信号处理领域中颇为引人注目的研究课题之一。国内对三维音频技术的研究尽管起步较晚,但发展较快,国内一些科研机构在 HRTF 的测量、个性化处理以及三维音频技术的应用等领域也取得了一定的研究成果^[7,16,77,78]。相信随着研究的不断深入,将会有越来越多的三维音频技术问题得到解决,并在实际中得到广泛应用。

参考文献:

- [1] BEGAULT D R. 3D Sound for Virtual Reality and Multimedia[M]. London: Academic Press, 1994.
- [2] BLAUERT J P. Spatial Hearing[M]. MIT, Cambridge, MA, 1997.
- [3] ZHANG W, ABHAYAPALA T D, KENNEDY R A, *et al.* Insights into head-related transfer function: spatial dimensionality and continuous representation[J]. *Journal of Acoustical Society of America*, 2010, 127(4): 2347-2357.
- [4] CHENG C I, WAKEFIELD G H. Introduction to head-related transfer functions (HRTFs): representations of HRTFs in time, frequency, and space[J]. *Journal of the Audio Engineering Society*, 2001, 49(4): 231-249.
- [5] GARDNER W G, MARTIN K D. HRTF measurements of a KEMAR[J]. *Journal of the Acoustical Society of America*, 1995, 97(6): 3907.
- [6] ALGAZI V R, DUDA R O, THOMPSON D M, *et al.* The CIPIC HRTF database[A]. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics[C]*. New Paltz, NY, 2001.
- [7] 谢菠菝, 钟小丽, 饶丹等. 头传递相关函数数据库及其特性分析[J]. *中国科学 G 辑: 物理学 力学 天文学*, 2006, 36(5): 464-479.
- XIE B S, ZHONG X L, RAO D, *et al.* Head-related transfer function database and its analyses[J]. *Science in China (Ser G Physics, Mechanics & Astronomy)*, 2006, 36(5): 464-479.
- [8] GUMEROV N A, O'DONOVAN A E, DURAISWAMI R, *et al.* Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation[J]. *Journal of the Acoustical Society of America*, 2010, 127(1): 370-386.
- [9] KATZ B F G. Boundary element method calculation of individual head-related transfer function[J]. *Journal of Acoustical Society of America*, 2001, 110(5): 2440-2455.
- [10] XIAO T, LIU Q H. Finite difference computation of head-related transfer function for human hearing[J]. *Journal of the Acoustical Society of America*, 2003, 113(5): 2434-2441.
- [11] WENZEL E M, ARRUDA M. Localization using nonindividualized head-related transfer functions[J]. *Journal of Acoustical Society of America*, 1993, 94(1): 111-123.
- [12] TAN C J, GAN W S. User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems[J]. *Electronics Letters*, 1998, 34(25): 2387-2389.
- [13] PARK M H, CHOI S I, KIM S H. Improvement of front-back sound localization characteristics in headphone-based 3D sound generation[A]. *The 7th International Conference on Advanced Communication Technology[C]*. Phoenix Park, South Korea, 2005.
- [14] IWAYA Y. Individualization of head-related transfer functions with tournament-style listening test: listening with other's ear[J]. *Acoustic Science and Technology*, 2006, 27(6): 340-343.
- [15] SEEBER B, FASTL H. Subjective selection of non-individual head-related transfer functions[A]. *Proceedings of the 2003 International Conference on Auditory Display[C]*. Boston, MA, USA, 2003.
- [16] 汪林, 殷福亮, 陈喆. 3D 声场合成中近似个性相关传递函数的主观选择方法[J]. *信号处理*, 2009, 25(7): 1097-1102.
- WANG L, YIN F L, CHEN Z. Subjective selection of near individualized head-related transfer functions for 3D sound[J]. *Signal Processing*, 2009, 25(7): 1097-1102.
- [17] ALGAZI V R, DUDA R O, DURAISWAMI R, *et al.* Approximating the head-related transfer function using simple geometric models of the head and torso[J]. *Journal of the Acoustical Society of America*, 2002, 112(5): 2053-2064.
- [18] HU H, ZHOU L, MA H, *et al.* HRTF personalization based on artificial neural network in individual virtual auditory space[J]. *Applied Acoustics*, 2008, 69(2): 163-172.
- [19] ZENG X, WANG S, GAO L. A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures[J]. *Journal of Sound and Vibration*, 2010, 329(19): 4093-4106.
- [20] HUOPANIEMI J, KARJALAINEN M. Review of digital filter design and implementation methods for 3-D sound[A]. *Proceedings of the 102 nd Convention of the Audio Engineering Society[C]*. Munich, Germany, 1997.
- [21] KULKARNI A, ISABELLE S K, COLBURN H S. On the minimum phase approximation of head-related transfer functions[A]. *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics[C]*. New Paltz, NY, USA, 1995.
- [22] SANDVAD J, HAMMERSHOI D. Binaural auralization: comparison of FIR and IIR representation of HRIRs[A]. *The 94th AES Convention[C]*. Amsterdam, Netherlands, 1994.
- [23] KULKARNI A, COLBURN H S. Efficient finite-impulse-response filter models of the head-related transfer function[J]. *Journal of Acoustical Society of America*, 1995, 97: 3278.
- [24] HWANG S, PARK Y. Modeling and customization of head-related impulse responses based on general basis functions in time domain[J]. *Acta Acustica united with Acustica*, 2008, 94(6): 965-980.
- [25] KISTLER D J, WIGHTMAN F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction[J]. *Journal of Acoustical Society of America*, 1992, 91(3): 1637-1647.
- [26] CHEN J, VAN VEEN B D, HECOX K E. A spatial feature extraction and regularization model for the head related transfer function[J]. *Journal of Acoustical Society of America*, 1995, 97(1): 439-452.
- [27] WU Z, CHAN F, LAM F K, *et al.* A time domain binaural model based on spatial feature extraction for the head-related transfer function[J]. *Journal of Acoustical Society of America*, 1997, 102(4): 2211-2218.
- [28] WANG L, YIN F, CHEN Z. A hybrid compression method for

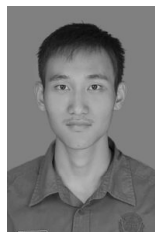
- head-related transfer functions[J]. *Applied Acoustics*, 2009, 70(9): 1212-1218.
- [29] ASANO F, SUZUKI Y. Role of spectral cues in median plane localization[J]. *Journal of Acoustical Society of America*, 1990, 88: 159-168.
- [30] HANEDA Y, MAKINO S, KANEDA Y, *et al.* Common-acoustical-pole and zero modeling of head-related transfer functions[J]. *IEEE Transactions on Speech and Audio Processing*, 1999, 7(2): 188-196.
- [31] MACKENZIE J, HUOPANIEMI J, VALIMAKI V, *et al.* Low-order modeling of head-related transfer functions using balanced model truncation[J]. *IEEE Signal Processing Letters*, 1997, 4(2): 39-41.
- [32] GEORGIU P, KYRIAKAKIS C. Modeling of head related transfer functions for immersive audio using a state-space approach[A]. *Conference Record of the Asilomar Conference on Signals, Systems and Computers*[C]. Piscataway, NJ, USA, 1999.
- [33] GRANTHAM D W, WILLHITE J A. Reduced order modeling of head related impulse responses for virtual acoustic displays[J]. *Journal of Acoustical Society of America*, 2005, 117(5): 3116-3125.
- [34] BLOMMER M A, WAKEFIELD G H. Pole-zero approximations for head-related transfer functions using a logarithmic error criterion[J]. *IEEE Transactions on Speech and Audio Processing*, 1997, 5(3): 278-287.
- [35] KULKARNI A, COLBURN H S. Infinite-impulse-response models of the head-related transfer function[J]. *Journal of the Acoustical Society of America*, 2004, 115(4): 1714-1728.
- [36] DURANT E A, WAKEFIELD G H. Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs[J]. *IEEE Transactions on Speech and Audio Processing*, 2002, 10(1): 18-27.
- [37] HUOPANIEMI J, KARJALAINEN M. Comparison of digital filter design methods for 3-D sound[A]. *Proceedings of IEEE Nordic Signal Processing Symposium (NORSIG96)*[C]. Helsinki, Finland, 1996.
- [38] GENUIT K. *Method and Apparatus for Simulating Outer Ear Free Field Transfer Function: United States*, 4672569[P]. 1987.
- [39] BROWN C P, DUDA R O. A structural model for binaural sound synthesis[J]. *IEEE Transactions on Speech and Audio Processing*, 1998, 6(5): 476-488.
- [40] ADAMS N H, WAKEFIELD G H. State-space synthesis of virtual auditory space[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, 16(5): 881-890.
- [41] KEYROUZ F, DIEPOLD K. A new HRTF interpolation approach for fast synthesis of dynamic environmental interaction[J]. *Journal of Audio Engineering Society*, 2008, 56(1/2): 28-35.
- [42] FREELAND F P, BISCAINHO L W P, DINIZ P S R. Efficient HRTF interpolation in 3D moving sound[A]. *AES 22nd International Conference in Virtual, Synthetic and Entertainment Audio*[C]. 2002.
- [43] ZHANG J, WU Z Y. A piecewise interpolation method based on log-least square error criterion for HRTF[A]. *IEEE 7th Workshop on Multimedia Signal Processing*[C]. Shanghai, China, 2005.
- [44] NISHINO T, KAJITA S, TAKEDA K, *et al.* Interpolating head related transfer functions in the median plane[A]. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*[C]. New Paltz, NY, USA, 1999.
- [45] KEYROUZ F, DIEPOLD K. A rational HRTF interpolation approach for fast synthesis of moving sound[A]. *12th Digital Signal Processing Workshop and 4th Signal Processing Education Workshop*[C]. Wyoming, USA, 2006.
- [46] CHENG C I, WAKEFIELD G H. Spatial frequency response surfaces: an alternative visualization tool for head-related transfer functions (HRTFs)[A]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*[C]. Phoenix, AZ, USA, 1999.
- [47] AJDLER T, SBAIZ L, VETTERLI M. Plenacoustic function on the circle with application to HRTF interpolation[A]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*[C]. Philadelphia, USA, 2005.
- [48] WANG L, YIN F, CHEN Z. Head-related transfer function interpolation through multivariate polynomial fitting of principal component weights[J]. *Acoustic Science and Technology*, 2009, 30(6): 395-403.
- [49] RUNKLE P R, BLOMMER M A, WAKEFIELD G H. A comparison of head related transfer function interpolation methods[A]. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*[C]. New Paltz, NY, USA, 1995.
- [50] TSUJINO K, KOBAYASHI W, ONOYE T, *et al.* Efficient 3-d sound movement with time-varying IIR filters[J]. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2007, E90-A(3): 618-625.
- [51] WATANABE K, TAKANE S, SUZUKI Y. Interpolation of head-related transfer functions based on the common-acoustical-pole and residue model[J]. *Acoustical Science and Technology*, 2003, 24(5): 335-337.
- [52] WANG L, YIN F, CHEN Z. An indirect interpolation method for HRTF pole/zero models[J]. *Acoustic Science and Technology*, 2008, 29(5): 1-3.
- [53] BAUER B B. Stereophonic earphones and binaural loudspeakers[J]. *Journal of Audio Engineering Society*, 1961, 9(1): 148-151.
- [54] SCHROEDER M R. Natural sounding artificial reverberation[J]. *Journal of the Audio Engineering Society*, 1962, 10(3): 219-223.
- [55] KIRKEBY O, NELSON P A, HAMADA H, *et al.* Fast deconvolution of multichannel systems using regularization[J]. *IEEE Transactions on Speech and Audio Processing*, 1998, 6(2): 189-195.
- [56] KALLINGER M, MERTINS A. A spatially robust least squares crosstalk canceller[A]. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing*[C]. Honolulu, 2007.
- [57] WANG L, YIN F, CHEN Z. A stereo crosstalk cancellation system based on the common-acoustical pole/zero model[J]. *EURASIP Journal on Advances in Signal Processing*, 2010 (1): 1-11.
- [58] WARD D B. Joint least squares optimization for robust acoustic crosstalk cancellation[J]. *IEEE Transactions on Speech and Audio*

- Processing, 2000, 8(2): 211-215.
- [59] KIM S M, WANG S. A Wiener filter approach to the binaural reproduction of stereo sound[J]. Journal of Acoustical Society of America, 2003, 114(6): 3179-3188.
- [60] RAO H I K, MATHEWS V J, PARK Y C. A minimax approach for the joint design of acoustic crosstalk cancellation filters[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(8): 2287-2298.
- [61] GARDNER W G. Head-tracked 3D audio using loudspeakers[A]. Proceedings of WASPAA[C]. New Paltz, NY, 1997.
- [62] KYRIAKAKIS C, HOLMAN T. Video based head tracking for improvement in multichannel loudspeaker audio[A]. The 105th Meeting of the Audio Engineering Society[C]. San Francisco, CA, 1998.
- [63] LEE K S, LEE S P. A real-time audio system for adjusting the sweet spot to the listener's position[J]. IEEE Transactions on Consumer Electronics, 2009, 56(2): 835-843.
- [64] BOURGET C, ABOULNASR T. Inverse filtering of room impulse response for binaural recording playback through loudspeakers[A]. 1994 IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. Adelaide, SA, Australia, 1994.
- [65] GONZALEZ A, LOPEZ J J. Time domain recursive deconvolution sound reproduction[A]. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. Istanbul, Turkey, 2000.
- [66] 马浩, 周琳, 胡红梅等. 串音消除系统中的自适应级联逆滤波器设计[J]. 东南大学学报, 2008, 38(1): 6-10.
- MA H, ZHOU L, HU H M, *et al.* Adaptive cascade inverse filter design for crosstalk cancellation system[J]. Journal of Southeast University, 2008, 38(1): 6-10.
- [67] KUO S M, CANFIELD G H. Dual-channel audio equalization and cross-talk cancellation for 3-D sound reproduction[J]. IEEE Transactions on Consumer Electronics, 1997, 43(4): 1189-1196.
- [68] NELSON P A, HAMADA H, ELLIOTT S J. Adaptive inverse filters for stereophonic sound reproduction[J]. IEEE Transactions on Signal Processing, 1992, 40(7): 1621-1632.
- [69] LIM J S, KYRIAKAKIS C. Multirate adaptive filtering for immersive audio[A]. 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001[C]. Salt Lake City, UT, USA, 2001.
- [70] GARAS J. Adaptive 3D Sound Systems[M]. Dordrecht, Netherlands: Kluwer Academic, 2000.
- [71] MOUCHTARIS A, REVELIOTIS P, KYRIAKAKIS C. Inverse filter design for immersive audio rendering over loudspeakers[J]. IEEE Transactions on Multimedia, 2000, 2(2): 77-87.
- [72] KIRKEBY O, NELSON P A, HAMADA H. The stereo dipole- A virtual source imaging system using two closely spaced loudspeakers[J]. Journal of Audio Engineering Society, 1998, 46(4): 387-395.
- [73] TAKEUCHI T, NELSON P A. Extension of the optimal source distribution for binaural sound reproduction[J]. Acta Acustica United with Acustica, 2008, 94(6): 981-987.
- [74] WARD D B, ELKO G W. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation[J]. IEEE Signal Processing Letters, 1999, 6(5): 106-108.
- [75] YANG J, GAN W S, TANG S E. Development of virtual sound imaging system using triple elevated speakers[J]. IEEE Transactions on Consumer Electronics, 2004, 50(3): 916-922.
- [76] HUANG Y, BENESTY J, CHEN J. Generalized crosstalk cancellation and equalization using multiple loudspeakers for 3D sound reproduction at the ears of multiple listeners[A]. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Las Vegas, US, 2008.
- [77] 胡红梅, 周琳, 马浩等. 基于偏最小二乘回归的与头相关传递函数的个人化[J]. 电子与信息学报, 2008, 30(1): 154-158.
- HU H M, ZHOU L, WU H, *et al.* Head-related transfer function personalization based on partial least square regression[J]. Journal of Electronics and Information Technology, 2008, 30(1): 154-158.
- [78] 谢波菽, 王杰, 管善群等. 5.1 通路环绕声的耳机虚拟重放[J]. 声学学报, 2005, 30(4): 329-336.
- XIE B S, WANG J, GUAN S Q, *et al.* Virtual reproduction of 5.1 channel surround sound by headphone[J]. Journal of Acta Acustica, 2005, 30(4): 329-336.

作者简介:



殷福亮 (1962-), 男, 辽宁抚顺人, 辽宁大连理工大学教授、博士生导师, 主要研究方向为数字信号处理、语音处理、图像处理、宽带无线通信技术和集成电路设计。



汪林 (1981-), 男, 安徽砀山人, 大连理工大学博士生, 主要研究方向为数字信号处理、语音与音频信号处理、视频信号处理。



陈喆 (1975-), 男, 黑龙江泰来人, 博士, 大连理工大学副教授, 主要研究方向为数字信号处理、语音处理、宽带无线通信技术和集成电路设计。