第 37 卷 第 2 期
2018 年 3 月

应用声学
Journal of Applied Acoustics

Vol. 37, No.2
March, 2018

◇ 研究报告 ◇

基于手机的多通路环绕声动态双耳重放*

林慧镔[1] 谢菠荪[1,2†]

(1 华南理工大学物理与光电学院 声学研究所 广州 510641)

(2 华南理工大学 亚热带建筑科学国家重点实验室 广州 510641)

摘要 随着VR眼镜技术的发展，普通的智能手机已可以作为虚拟现实和动态声、视频重放的平台。该文提出了一种基于手机的多通路环绕声动态双耳重放技术及其信号处理的高效实现方法。利用手机内的加速度传感器、电子罗盘、陀螺仪组成头踪迹跟踪器，实时检测倾听者头部的方向，并利用手机的信号处理芯片实现动态双耳合成。采用头相关脉冲响应的最小相位近似和主成分分解的方法简化双耳合成处理，提高了信号处理的效率。文中给出了系统的结构和软、硬件设计方法，并给出了实现22.2通路空间环绕声动态双耳重放的例子。客观测量和心理声学实验验证了所提出的方法。

关键词 多通路声，动态双耳重放，手机

中图法分类号: O429 **文献标识码:** A **文章编号:** 1000-310X(2018)02-0187-09

DOI: 10.11684/j.issn.1000-310X.2018.02.001

Dynamic binaural reproduction of multichannel surround sound based on mobile phone

LIN Huibin[1] XIE Bosun[1,2]

(1 *Acoustic Lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou* 510641, *China*)

(2 *State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou* 510641, *China*)

Abstract With the recent development of VR glass technique, smart mobile phones can serve as platform for virtual reality and dynamic video and audio reproduction. In present work, a technique for dynamic binaural reproduction of multichannel surround sound based on mobile phone as well as corresponding scheme for high-efficient signal processing are developed. The orientation of listener's head is real-time detected by using the head tracker consisting of acceleration sensor, electronic compass and gyroscope in the mobile phone, and then dynamic binaural synthesis is implemented by the DSP in mobile phone. Minimum phase approximation and principal component decomposition are applied to head-related impulse responses so as to simplify binaural synthesis and then to improve the efficiency of signal processing. The system structure and design methods for software and hardware are addressed. Objective measurement and psychoacoustic experiment validate the proposed system and technique.

Key words Multichannel sound, Dynamic binaural reproduction, Mobile phone

2017-07-07 收稿; 2017-11-14 定稿

*国家自然科学基金项目 (11674105), 华南理工大学亚热带建筑科学重点实验室自主研究课题

作者简介: 林慧镔 (1990-), 男, 广东佛山人, 硕士, 研究方向: 电声与听觉。

† 通讯作者 E-mail: phbsxie@scut.edu.cn

1 引言

多通路环绕声原本是为影院或家庭环境下的扬声器重放而设计的。它将多通路信号馈给环绕倾听者的多个扬声器重放，在倾听者双耳处产生适当的叠加声压，从而产生期望的空间听觉效果。国际上已发展了多种不同的多通路环绕声系统[1]，包括传统的5.1通路、7.1通路水平面环绕声系统；后来发展的9.1通路、10.2通路、11.1通路、22.2通路三维空间环绕声系统；近年所发展的面向目标三维空间环绕声系统，如Dolby Atmos、MPEG-H 3D Audio标准等[2]。

但另一方面，各种便携与手持播放设备，如智能手机等，也逐渐成为声、视频重放的终端。这些设备需要通过耳机来重放多通路环绕声信号。但将多通路环绕声信号直接向下混合成两通路信号后用耳机重放，会破坏其中声音的空间信息，不能获得正确的空间听觉效果。为改善耳机重放多通路环绕声的效果，可采用各扬声器方向的头相关脉冲响应对信号进行卷积(或等价地，用头相关传输函数滤波)，合成双耳声信号并用耳机重放[3−4]。这相当于用信号处理的方法虚拟出多通路环绕声的多个扬声器。而目前多数这类技术是采用合成稳态双耳信号的方法，也就是假定倾听者头部是不动的，因而虚拟扬声器相对倾听者的方向是不变的。而在实际的扬声器重放中，倾听者头部是可以自由转动的。头部转动引起的双耳声压改变带来了动态定位信息。虽然稳态双耳重放比较简单，也可以改善声音空间信息的重放，但由于失去了动态信息，它还是存在一定的缺陷，例如虚拟声源前后混乱与上下混乱、感知的虚拟声源方向与期望值有偏差等。由于听觉定位信息有一定的冗余性，采用个性化或者定制的头相关脉冲响应进行信号处理可以准确保留双耳声压的高频谱信息，从而部分地减少上述缺陷[5−6]，但还未能从根本上解决问题。

在合成双耳信号中引入动态信息可以明显改善听觉效果，包括降低虚拟声源前后混乱率与上下混乱率、提高虚拟声源定位准确性等[7]。这是通过各种方法检测倾听者头部运动信息，然后动态地刷新合成双耳信号处理而实现的。各种科学研究用的虚拟听觉环境实时绘制系统是采用专业的头踪迹跟踪器检测倾听者头部运动信息的，可以达到很高的技术指标。但其体积大并且价格较高，不适用于手机等消费类电子产品应用。对于消费电子类应用，我们实验室曾采用廉价的磁传感器和加速度传感器获取头部在水平面运动引起的一个自由度的物理信息变化[8]。这种头踪迹跟踪器比较简单，可以近似实现水平面动态双耳重放，但不能实现整个三维空间的动态双耳重放。近年，VR 眼镜技术的发展，普通的智能手机已可以作为虚拟现实和动态声、视频重放的简单平台。2016年5月谷歌公司发布了Google VR SDK，成功地在手机上使用虚拟Ambisonics技术实现动态虚拟听觉重放(见网页https://developers.google.com/vr/unity/)。但其虚拟声源的数目是有限的。

事实上，检测倾听者头部运动信息的技术是多样的，其技术性能、复杂程度与成本、信号处理和实现方法也不同。对于不同的应用，可以根据实际情况采用不同的技术检测倾听者头部运动信息，实现多通路环绕声的动态双耳重放。

为了克服现有技术存在的缺点与不足，本文提出了一种基于手机的多通路环绕声动态双耳重放系统的实现方法，通过手机获取头部三个转动自由度的运动信息，并采用高效的信号处理方法实现动态双耳信号合成，最后以22.2通路空间环绕声的动态双耳重放为例，从实验上验证所提出方法的有效性。

2 多通路环绕声的动态双耳重放原理

声源相对于倾听者的方向用方位角$0^\circ \leqslant \theta < 360^\circ$和仰角$-90^\circ \leqslant \phi \leqslant 90^\circ$表示。而$\phi = -90^\circ$、$0^\circ$和$90^\circ$分别表示正下方、水平面与正上方；水平面$\theta = 0^\circ$、$90^\circ$、$180^\circ$和$270^\circ$分别表示正前方、正右方、正后方、正左方。

在采用扬声器的多通路声重放中，假定有M个布置在不同空间方向的扬声器，第i个扬声器的方向为(θ_i, ϕ_i)，频域信号为$E_i(f)$，f代表频率。则M个扬声器所产生的双耳频域叠加声压可以由头相关传输函数(HRTF)计算得到[3]

$$P_\alpha(f) = \sum_{i=1}^{M} H_\alpha(\theta_i, \phi_i, f) E_i(f), \tag{1}$$

其中，$H_\alpha(\theta_i,\phi_i,f)$ 为第 i 个扬声器到左耳或右耳的 HRTF，$\alpha=\mathrm{L}$ 或 R 表示左耳或右耳。

如果将多通路环绕声信号按式 (2) 用 M 个扬声器方向的 HRTF 滤波处理并叠加，则可以合成双耳信号：

$$E_\alpha(f)=\sum_{i=1}^{M}H_\alpha(\theta_i,\phi_i,f)E_i(f), \tag{2}$$

将 $E_\alpha(f)$ 用耳机重放，即可以产生和扬声器重放相同的双耳声压，也就是 M 个虚拟扬声器，从而产生和扬声器重放类似的空间听觉感知。考虑到耳机与耳道的耦合和非理想的耳机传输特性，可以在式 (2) 中补充耳机–耳道传输特性的逆滤波器对双耳信号进行均衡处理。

式 (2) 的信号处理也可以在时域实施，并写成：

$$e_\alpha(t)=hp_\alpha^{-1}(t)\otimes\sum_{i=1}^{M}h_\alpha(\theta_i,\phi_i,t)\otimes e_i(t), \tag{3}$$

式 (3) 中，各小写的函数是式 (2) 所对应的各大写频域函数的时域形式，它们之间由傅里叶变换相联系，其中 $h_\alpha(\theta_i,\ \phi_i,\ t)$ 为头相关脉冲响应 (HRIR)，符号“$\otimes$”表示卷积。另外，式 (3) 已补上了耳机–耳道传输特性的逆滤波器均衡处理，$hp_\alpha^{-1}(t)$ 表示逆滤波器的脉冲响应。

由于在实际的多通路环绕声重放中，倾听者头部的转动会引起各扬声器相对于倾听者方向的改变，从而引起双耳声压的变化。这种变化带来的动态双耳因素对于前后和垂直方向的定位是非常重要的。为了模拟这种动态因素，可以采用头踪迹跟踪器实时检测倾听者头部的位置，并动态改变式 (3) 的 HRIR 数据，这就是多通路环绕声动态双耳重放的基本原理。过去的实验表明，引入动态因素可以减少对个性化 HRTF 的依赖，减少重放的前后和上下定位混乱 [7]。

3　信号处理的简化设计

实际中直接应用式 (3) 的信号处理合成双耳信号存在以下问题：

(1) 式 (3) 涉及 M 个虚拟扬声器方向的 HRIR 卷积处理，运算量与虚拟扬声器的数目 M 成正比。近年所发展的三维空间环绕声 (包括面向通路和面向目标的系统) 的扬声器数目都很大，因而式 (3) 的信号处理将需要很大的计算资源。对于手机应用，由于计算资源有限，这将成为一个问题。

(2) 动态双耳重放需要根据倾听者头部的位置实时刷新式 (3) 的 HRIR 数据。但直接切换 HRIR 数据会引起合成双耳信号的不连续，产生可听的缺陷。

为了解决上面的问题，本文设计的系统将采用与式 (3) 等价但简化的信号处理方法，即基于 HRIR 最小相位近似与主成分分解的信号处理方法，在我们以前的虚拟听觉环境实时绘制系统中已用过类似的方法 [7]。过去的研究指出 [3,9]，任意虚拟扬声器方向的 HRTF 或 HRIR 可以近似用其最小相位函数和一个线性延时表示，对 HRIR 可以写为

$$h_\alpha(\theta_i,\phi_i,t)\approx h_{\min,\alpha}(\theta_i,\phi_i,t-\tau_\alpha), \tag{4}$$

其中，线性延时 τ_α 与虚拟扬声器方向以及左耳、右耳有关，即 $\tau_\alpha=\tau_\alpha(\theta_i,\phi_i)$。在 HRIR 的最小相位近似下，式 (3) 的信号处理成为

$$\begin{aligned}e_\alpha(t)=hp_\alpha^{-1}(t)\otimes\sum_{i=1}^{M}h_{\min,\alpha}(\theta_i,\phi_i,t)\\ \otimes e_i[t-\tau_\alpha(\theta_i,\phi_i)],\end{aligned} \tag{5}$$

采用最小相位近似可以减少卷积的 HRIR 脉冲应的有效长度，初步实现信号处理的简化。

另一方面，进一步对 HRIR 进行主成分分解 [10]，将任意虚拟扬声器方向的最小相位 HRIR 分解为 Q 个时域基函数的权重和，加上一个平均时域函数：

$$\begin{aligned}&h_{\min,\alpha}(\theta_i,\phi_i,t)\\&=\sum_{q=1}^{Q}w_{q,\alpha}(\theta_i,\phi_i)d_q(t)+\bar{h}_{\min,\alpha}(t),\end{aligned} \tag{6}$$

其中，时域基函数 $d_q(t)$ 和平均时域函数 $\bar{h}_{\min,\alpha}(t)$ 只和时间有关而与虚拟扬声器无关。而 Q 个权重系数 $w_{q,\alpha}(\theta_i,\phi_i)$ 只和虚拟扬声器的方向和左耳、右耳有关，与时间无关。至于时域基函数、平均时域函数和权重系数的具体计算方法以及阶数 Q 的选择见文献 [3]，在此不再详述。

将式 (6) 代入式 (5)，并令

$$d_0(t)=\bar{h}_{\min,\alpha}(t),\quad w_{0,\alpha}(\theta_i,\phi_i)=1, \tag{7}$$

可以得到

$$e_\alpha(t) = hp_\alpha^{-1}(t) \otimes \sum_{q=0}^{Q} d_q(t) \otimes \{w_{q,\alpha}(\theta_i, \phi_i) \times e_i[t - \tau_\alpha(\theta_i, \phi_i)]\}, \tag{8}$$

由式(8)可以看出，将多通路环绕声每一通路信号作延时并以权重 $w_{q,\alpha}(\theta_i, \phi_i)$ 叠加后，和 $(Q+1)$ 个脉冲响应 $d_q(t)$ 卷积，或等价于 $(Q+1)$ 个滤波处理。将卷积输出叠加后再进行耳机–耳道传输特性的逆滤波器均衡处理，即可得到双耳信号。

与式(3)或式(5)的HRIR直接卷积比较，经过HRIR的主成分分解后，式(8)的信号处理将各通路的信号重新叠加组合后再和 $(Q+1)$ 个共用的脉冲响应 $d_q(t)$ 卷积，或等价地，用 $(Q+1)$ 个共用的滤波器滤波。这时，滤波器的数目是固定的，不随虚拟扬声器的数目 M 增加。只要 $Q < M$，即可减少信号处理中卷积运算的次数。另外，对于动态双耳合成，只需要根据倾听者头部相对虚拟扬声器的位置实时刷新式(8)的权重 $w_{q,\alpha}(\theta_i, \phi_i)$，同时刷新延时 $\tau_\alpha(\theta_i, \phi_i)$，从而避免了直接刷新脉冲响应带来的缺陷。当然，手机中只能存储有限个离散方向的权重 $w_{q,\alpha}(\theta_i, \phi_i)$ 和延时 $\tau_\alpha(\theta_i, \phi_i)$。其他未存储方向的权重可通过简单的线性插值得到[3]。至于延时 $\tau_\alpha(\theta_i, \phi_i)$，可仿照文献[11]对双耳时间差的处理的方法，用低阶的球谐函数分解而得到连续的延时函数。因此式(8)的双耳合成信号处理可以解决式(3)的传统信号处理的问题。

4 系统的结构与硬件、软件设计

图1是基于手机的多通路环绕声动态双耳重放系统原理的方块图。用VR眼镜将智能手机变成虚拟现实的3D显示设备。设定虚拟扬声器的位置和信号处理用的HRIR数据，并输入多通路环绕声信号。利用智能手机上的陀螺仪、加速度传感器等可以实现头踪迹跟踪，并由手机的DSP芯片按第3节的方法实现动态双耳信号合成。合成的双耳信号由耳机重放。

图2是系统信号处理流程的方块图，它包括信号处理参数计算模块、头踪迹跟踪模块以及信号处理模块。头相关脉冲响应数据包含其最小相位函数主成分分解的公共时域基函数、主成分分解权重系数、HRIR线性延时的球谐函数分解系数等。声音播放是指调用手机操作系统的函数将合成后的双耳信号复制至底层系统播放。三个主要功能模块作用为

(A) 信号处理参数计算模块

信号处理参数计算模块用于虚拟扬声器方向的HRIR线性延时和最小相位头相关脉冲响应主成分分解系数的插值计算。根据头踪迹跟踪模块传递的信息，通过球谐函数插值和双线性插值得到上述两个参数。

(B) 头踪迹跟踪模块

头踪迹跟踪模块用于实时检测倾听者的头部运动和计算虚拟扬声器相对于头部的方向。采用手机的电子罗盘、加速度传感器、陀螺仪检测倾听者运动信息，通过互补滤波器实现数据融合，继而求得对应的旋转矩阵。然后通过空间坐标转换得到虚拟扬声起相对倾听者的方位角和仰角并传递给其他模块，为动态处理提供依据。附录给出了实现头踪迹跟踪的方法。

(C) 信号处理模块

信号处理模块是系统的核心部分，接收头踪迹跟踪模块传递的虚拟扬声器方向信息，并用于上面信号处理参数的计算和实现动态双耳信号的合成，它包括了信号的延时处理和卷积处理。

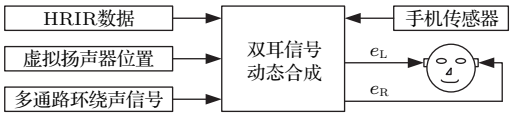


图1 基于手机的多通路环绕声动态双耳重放系统原理的方块图

Fig. 1 Block diagram of dynamic binaural reproduction of multichannel surround sound based on mobile phone

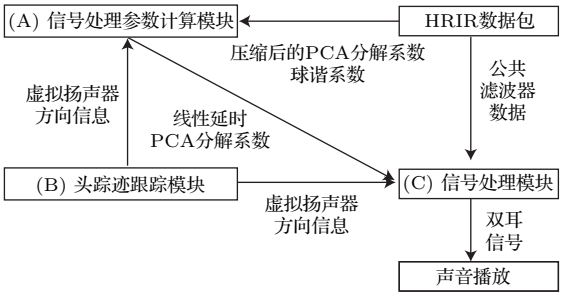


图2 系统信号处理流程的方块图

Fig. 2 Block diagram of signal processing scheme

5 设计实例与性能参数

我们选取某品牌的手机作为实验用手机。其主要参数是：Exynos 7420 处理器 (ARM® Cortex®-A57™2.1 GHz×4 + ARM® Cortex®-A53™1.5 GHz ×4)，Mali T760 图形处理器，3 GB LPDDR4 内存，包含重力感应、陀螺仪、电子罗盘等一系列传感器，运行的系统是安卓 5.1。

本实例是在谷歌公司提供的安卓集成开发环境 Android Studio 上使用 Java 编程实现。其中还使用了 JTransforms 开源库，它提供了在 Java 环境下的多线程 FFT 计算功能，以及 Commons Math 组件库，它是一个轻量级包含数学与统计计算的库。

表 1 列出了测量得到的系统实例的动态性能参数，这是采用与文献 [7] 相同的方法测量得到的。其中系统刷新率指的是单位时间内的信号处理 (场景) 刷新次数。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。Sandvad 的心理声学实验结果表明 [12]，20 Hz 的场景刷新率是基本足够的。至于听觉上可察觉的滞后时间，不同实验得到结果差异较大，从几十 ms 到 500 ms[12−13]，这可能取决于重放的信号。但对于手机类的消费电子应用，略为降低的动态性能也是可接受的。因此本设计实例的动态性能是基本上可满足应用要求的。

最高支持虚拟扬声器数量是指按第 2 节所述的传统动态合成双耳信号处理方法，系统能同时最多合成多个不同方向虚拟扬声器的数目，表 1 中给出的结果是 32。当采用第 3 节所述的基于 HRIR 的最小相位近似与主成分分解的信号处理方法，只要用于卷积的公共脉冲响应数目或等价的滤波器数目 $(Q+1)$ 不大于 32，即可在本实施平台上实现。

需要指出的是，系统的动态性能参数很大程度是取决于所用手机平台的软、硬件性能。今后随着手机性能的提高，系统的动态性能参数会进一步改善。

表 1　系统性能参数列表

Table 1　Parameters for system performance

系统刷新率/Hz	系统滞后时间/ms	最高支持虚拟扬声器数量/个
60	260	32

6 22.2 通路空间环绕声的动态双耳重放及实验验证

作为本文设计系统的应用，本节讨论在上面的手机平台上实现 22.2 通路声的动态双耳重放。

22.2 通路声是近年发展的多通路空间环绕声的典型代表，ISO/IEC 所制定 MPEG-H 3D Audio 空间声标准也支持 22.2 通路声的重放。不同的研究和标准中，22.2 通路系统的扬声器布置有一些差异，且一些标准只给出了一定的布置范围，本文采用 NHK 在听音区域评估实验中使用的布置方案 [14]。它包括顶层、中层和底层共 22 个全频带扬声器布置及信号通路，另加两个次低频扬声器及信号通路。图 3 给出了三层扬声器布置，其中顶层除了正上方扬声器，其他 8 个扬声器仰角是 $\phi = 45°$，中层 10 个扬声器布置在水平面 $\phi = 0°$，底层 3 个扬声器布置的仰角 $\phi = -30°$。为了清晰描述系统中每一个扬声器，本文使用国际电信联盟中的英文简写记号描述该系统的扬声器，图 3 中同时标出了各扬声器的记号和方向 (θ, ϕ)。

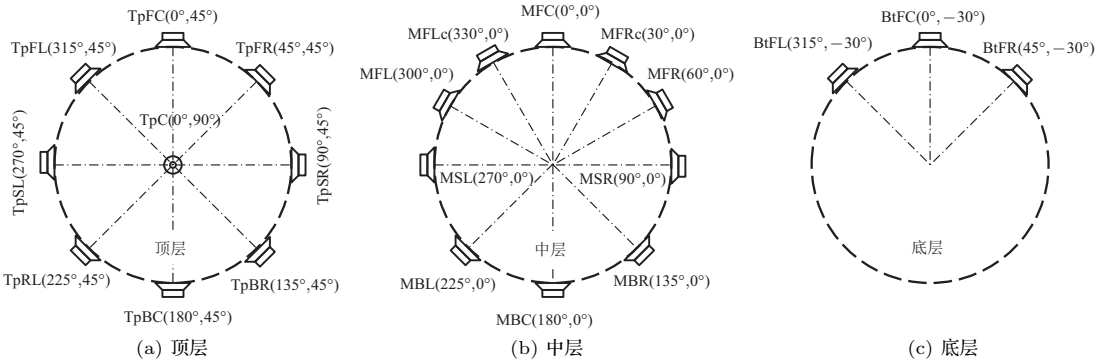


图 3　22.2 通路环绕声系统布置

Fig. 3　Loudspeaker configuration of 22.2 channel surround sound

根据上述22.2通路扬声器布置，用第3节给出的信号处理方法和第4节给出的基于手机的系统平台，实现22.2通路空间环绕声的动态双耳重放。信号处理采用边界元方法计算得到的KEMAR人工头远场HRIR数据，最小相位HRIR的长度是128点。在最小相位HRIR主成分分解中，采用$Q=9$个时域基函数组成公共滤波器。按文献[3]的方法分析，$Q=9$个时域基函数可以代表最小相位HRIR累积能量变化的99.2%，而HRIR的线性延时则用前6阶球谐函数分解与重构，前6阶球谐函数贡献的累计百分比达到97.8%。上述参数的精度足以满足听觉的要求。

虚拟源定位的心理声学实验验证了22.2通路环绕声双耳重放的效果。对输入信号分别作动态和稳态双耳合成处理并用耳机重放，然后对比虚拟扬声器(虚拟声源)方向的定位结果。实验在本底噪声不大于30 dB的听音室内进行，实验使用的是AKG702头戴式耳机。实验采用全可听频带粉红噪声作为原始信号，采样频率44.1 kHz、量化精度16 bit，信号长度是10 s。

用于实验的手机被放置在普通的VR眼镜里，所以在两种重放条件下，在实验过程中受试者佩戴VR眼镜，无法观察到外界环境，同时手机屏幕不会显示与声源定位有关的信息。实验过程中，受试者右手会手持着Fastrack电磁定位接收器，受试者将通过该接收器指示感知到的虚拟扬声器方向。

由于22.2通路扬声器布置和人类头部的近似左右对称性，实验只选取22.2通路环绕声系统中的14个扬声器方向进行定位。其中，顶层有6个扬声器(正上方扬声器，仰角$\phi=90°$)，仰角为$\phi=45°$，方位角为$\theta=0°,45°,90°,135°,180°$；中层有6个扬声器，仰角为$\phi=0°$，方位角为$\theta=0°,30°,60°,90°,135°,180°$；底层有2个扬声器，仰角为$\phi=-30°$，方位角为$\theta=0°,45°$。

本实验共有8名受试者，男女各半，年龄在23至26岁之间，均无听力障碍。在两种重放条件下，每名受试者都需要对实验的每个目标虚拟扬声器方向进行4次定位。因此，每种重放条件下每一个目标虚拟扬声器方向一共有32个定位数据。

考虑到实验时受试者无法观察外界环境以及要佩戴数种设备，因此将每一组实验控制在20 min左右，并将两种重放条件下的实验个划分成两组，每组有28个信号，包含了2×14个方位。为了让受试者能够熟悉实验流程，在开始实验之前会有短暂的听音训练。

实验将使用以下三个统计指标来评估虚拟扬声器的定位效果。

(1) 上下混乱率：受试者感知的虚拟扬声器方向出现上下混乱的次数与总的虚拟扬声器定位次数之比，其中不考虑水平面虚拟扬声器的情况(水平面是上下的分界)。

(2) 前后混乱率：受试者感知的虚拟扬声器方向出现前后混乱的次数与总的虚拟扬声器定位次数之比，其中不考虑侧垂面的情况(侧垂面是前后分界的情况)。

(3) 平均空间角度偏差：$\boldsymbol{r}_s(i)$表示实验中的第i个虚拟扬声器目标方向的单位矢量，共有M个虚拟扬声器；$\boldsymbol{r}_I(n,i)$表示对第i个虚拟扬声器方向第n次判断的单位矢量，共有N次判断。所以平均空间角度偏差可由式(9)求得

$$\Delta=\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{n=1}^{N}|\arccos[\boldsymbol{r}_I(n,i)\cdot\boldsymbol{r}_s(i)]|. \quad (9)$$

表2给出了所有受试者的平均实验结果，与稳态重放比较，动态重放时上下混乱率、前后混乱率、平均空间角度偏差都有明显的改善。进一步地统计t检验表明，在显著性水平$\alpha=0.05$的条件下，上述结论是成立的。另外，过去在虚拟听觉环境实时绘制系统平台(具有相对足够的计算资源)的虚拟源定位实验表明，远场虚拟源的平均前后混乱率为2.5%，与本文的结果为同一量级。事实上，手机平台的计算资源有限，主要是对系统的动态特性，如刷新率、系统滞后时间有影响。

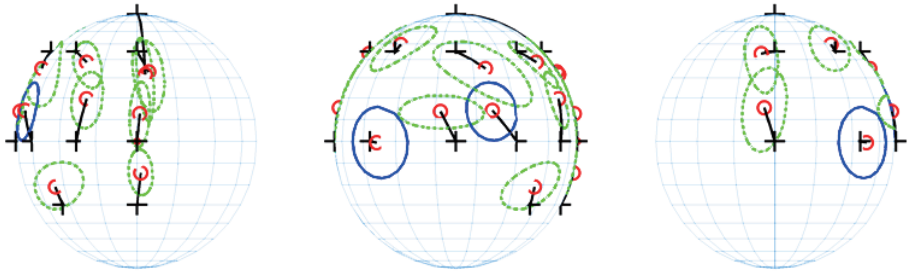
表2 定位实验数据的总体统计结果

Table 2 Statistics of experimental results

	上下混乱率/%	前后混乱率/%	Δ/(°)
稳态重放	13.3	27.0	24.5
动态重放	5.9	2.8	15.3

为了直观清晰展示稳态重放和动态重放的虚拟扬声器定位实验统计结果，图4和图5使用Leong

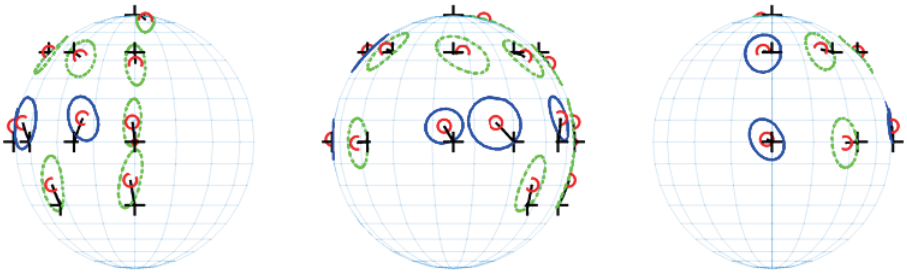
等[15]提出的图示化方法，将统计结果用球表面分布上展示。图中符号“+”表示目标虚拟扬声器的方向，红圈“◯”表示样本均值。椭圆线表示样本的标准差，其中蓝色椭圆实线表示样本离散特性服从 Kent 分布[16]，而绿色椭圆虚线则表示服从 Fisher 分布[17]。显著水平 $\alpha = 0.05$ 的条件下，Kent 分布表示样本在空间的任意方向上的方差无显著性差异，Fisher 分布表示样本在空间某个方向的方差更大。图 4 和图 5 表明，动态重放方差比稳态重放小，所以动态重放能够改善定位精度。



(a) 面向 $\theta = 0°$　(b) 面向 $\theta = 90°$　(c) 面向 $\theta = 180°$

图 4　稳态重放统计结果，不同观察方向

Fig. 4　Statistical results for static reproduction, plotted at different observed directions



(a) 面向 $\theta = 0°$　(b) 面向 $\theta = 90°$　(c) 面向 $\theta = 180°$

图 5　动态重放统计结果，不同观察方向

Fig. 5　Statistical results for dynamic reproduction, plotted at different observed directions

值得注意的是，从简单和实用考虑，本文的系统只是采用了 KEMAR 人工头的计算 HRIR(HRTF) 数据进行双耳合成，并没有用个性化 HRIR 数据。但从表 2 给出的定位实验统计结果可以看出，动态重放已经得到了很好的定位效果，特别是给出了非常低的前后和上下混乱率。这是因为动态因素和个性化 HRTF 所包含的谱因素同时对前后和上下定位起作用，但它们所提供的信息是有冗余的[3]。动态双耳重放提供了正确的动态定位因素，明显减少了对个性化谱因素的依赖。

7　结论

手机是实现虚拟现实和声、视频重放的一个简单平台。本文提出了一种基于手机的多通路环绕声动态双耳重放技术，并实现了 22.2 通路空间环绕声的动态双耳重放。利用手机自带的传感器系统可以实现头踪迹跟踪，采用头相关脉冲响应的最小相位近似和主成分分解的方法可以提高合成双耳信号处理的效率。客观测量和心理声学实验验证了所提出方法的有效性。本文所提出的方法是通用的，不但适用于 22.2 通路声重放，也适用于其他各种面向通路和面向目标的多通路空间声重放。今后随着手机软硬件技术的发展，可以在此基础上发展出更多和更好的重放功能。

参　考　文　献

[1] ITU-R REPORT BS.2159-7, Multichannel sound technology in home and broadcasting application[S]. International Telecommunication Union, Geneva, Switzerland, 2015.

[2] Herre J, Hilpert J, Kuntzuntz A, et al. MPEG-H audio—The new standard for universal spatial/3D audio coding[J]. J. Audio Eng. Soc., 2015, 62(12): 821–830.

[3] Xie B S. Head-related transfer function and virtual auditory display[M]. 2nd Edition. USA: J Ross Publishing, 2013.

[4] 谢菠荪, 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 Bosun, Wang Jie, Guan Shanqun, et al. Virtual reproduction of 5.1 channel surround sound by headphone[J]. Acta Acustica, 2005, 30(4): 329–336.

[5] Wenzel E M, Arruda M, Kistler D J, et al. Localization using nonindividualized head-related transfer functions[J]. J. Acoust. Soc. Am., 1993, 94(1): 111–123.

[6] Xie B S, Zhong X L, He N N. Typical data and cluster analysis on head-related transfer functions from Chinese subjects[J]. Applied Acoustics, 2015, 94(1): 1–13.

[7] Zhang C Y, Xie B S. Platform for dynamic virtual auditory environment real-time rendering system[J]. Chinese Science Bulletin, 2013, 58(3): 316–327.

[8] 张承云, 谢菠荪. 一种 5.1 通路环绕声的动态双耳重放方法[J]. 应用声学, 2016, 35(4): 283–287.
Zhang Chengyun, Xie Bosun. A scheme for dynamic binaural reproduction of 5.1 channel surround sound[J]. Journal of Applied Acoustics, 2016, 35(4): 283–287.

[9] Kulkarni A, Isabelle S K, Colburn H S. Sensitivity of human subjects to head-related transfer-function phase spectra[J]. J. Acoust. Soc. Am., 1999, 105(5): 2821–2840.

[10] Kistler D J, Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction[J]. J. Acoust. Soc. Am., 1992, 91(3): 1637–1647.

[11] 钟小丽, 谢菠荪. 球谐函数展开的个性化双耳时间差模型[J]. 声学学报, 2013, 38(4): 477–485.
Zhong Xiaoli, Xie Bosun. Individualized interaural time difference model based on spherical harmonics function expansion[J]. Acta Acustica, 2013, 38(4): 477–485.

[12] Sandvad J. Dynamic aspects of auditory virtual environments[C]. AES 100th International Convention, 1996.

[13] Wenzel E M. Effect of increasing system latency on localization of virtual sounds[C]. AES 16th International Conference, 1999.

[14] Sawaya I, Oode S, Ando A, et al. Size and shape of listening area reproduced by three-dimensional multichannel sound system with various numbers of loudspeakers[C]. AES 131st International Convention, 2011.

[15] Leong P, Carkile S. Methods for spherical data analysis and visualization[J]. J. Neuroscience Methods, 1998, 80(2): 191–200.

[16] Kent J T. The Fisher-Bingham distribution on the sphere[J]. J. Royal Statistical Society. Series B (Methodological), 1982, 44(1): 71–80.

[17] Fisher N I, Lewis T, Embleton B J J. Statistical analysis of spherical data[M]. UK: Cambridge University Press, 1987.

[18] Bil C, Zegers J, Hazeleger L, et al. Experimental flight test for autonomous station-keeping of a lighter-thanair vehicle[M]//Damiani E, Howlett R J, Jain L C, et al. Intelligent interactive multimedia systems and services. Switzerland: Springer International Publishing, 2015: 299–311.

附录 A 手机实现头踪迹跟踪的基本方法

头部在空间可以有六个运动自由度，包括三个旋转自由度和三个平移自由度。受到系统实现的限制，本文只通过智能手机检测并模拟头部旋转自由度运动引起的双耳声信号变化。

这里存在两种坐标系统，分别是手机的本地坐标系统和世界坐标系统。手机的本地坐标系统固定在手机上，如图 A-1 所示。其中 O 是坐标原点，与手机的几何中心重合。手机的传感系统可以检测手机绕三个坐标轴的转动角度，即图 A-1 中的 a、b、c。世界坐标系统固定在地面，x 轴指向正东方，y 轴指向正北方，z 轴指向上方，三个坐标轴两两垂直。当手机水平放置，且手机本地坐标系的 y' 轴指向正北面时，两种坐标系统重合。

由于本系统通过 VR 眼镜将手机固定在双眼前面，当头部发生转动时，手机随头部一起转动，但手机几何中心相对头部中心的位置是固定的。当头部转动后，利用手机的传感系统，可以检测到手机在本地坐标系统绕三个坐标轴的转动角度 a、b、c。根据适当的几何关系，可从中计算出倾听者头部相对世界坐标的旋转角度 (矩阵)。对于非运动的稳态虚拟声源 (如 22.2 通路环绕声系统的各虚拟扬声器)，其在世界坐标系统的位置是固定不变的。由头部相对世界坐标的旋转角度，即可计算出各虚拟声源相对头部的位置，从而实施动态双耳声信号的合成处理。

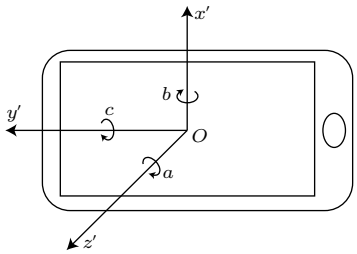


图 A-1　手机的本地坐标系统

Fig. A-1　Local coordinate system of mobile phone

手机一般是通过加速度传感器、电子罗盘和陀螺仪来获取手机运动信息的。传统的陀螺仪是基于角动量守恒定律测量空间中物体角度变化的机械装置，而手机的陀螺仪是基于微机电系统(Micro electro mechanical systems, MEMS)，利用科里奥利力原理测量出手机转动的角速度，通过积分得到手机在一段时间后转动的角度。但是MEMS陀螺仪自身存在偏移，积分运算的性质使得误差累积，因此随着时间的推移结果将变得不可靠。

安卓手机系统中通过加速度传感器和电子罗盘的数据融合也能够获得手机自身的转动信息，但是该两个传感器的输出也包含噪声，其中电子罗盘的噪声比较明显。因此本文使用互补滤波器的方法，融合手机中三个传感器的数据，从而获得较为稳定的手机转动数据。互补滤波器使得从陀螺仪获取的数据相当于经过了高通滤波器处理，而从加速度传感器和电子罗盘获得的数据相当于经过了低通滤波器处理，能一定程度上减少 3 个传感器的噪声对跟踪效果的影响 (见 Min H.G. 等，http://goo.gl/F2PeSX)。

$$R = kR_1 + (1-k)R_2, \tag{A-1}$$

其中，R 表示融合后的方向信息，R_1 表示从陀螺仪获得的方向信息，R_2 表示从加速度传感器和电子罗盘处得到的方向信息，k 是互补滤波器的系数。

在本文搭建的系统中，获取手机三个传感器的数据后，使用互补滤波器处理数据得到手机的旋转矩阵。k 可以依照互补滤波器的低通滤波器的截至频率算出，这需要传感器的采样率，但安卓系统并未提供这方面的数据。Cees Bil 为轻型飞行器设计的控制系统中 k 使用经验值 0.98[18]，本文采用此值。