◇ 研究报告 ◇

Mel频率倒谱系数平滑的耳机均衡*

李光炬1,2† 罗平展3 钱 鹏1 甘维明1 邢 锰1

(1 中国科学院声学研究所南海研究站 海口 570105)
(2 中国科学院噪声与振动重点实验室 北京 100190)
(3 哈曼科技(深圳)有限公司 深圳 518067)

摘要:适当均衡耳机到鼓膜的传递函数可有效提高耳机声重放效果。耳廓与耳道滤波效应引起的幅度峰谷有助于人耳听觉感知,以平直幅频响应为目标的幅度均衡无法保持适当的峰谷。该文提出了基于 roex 滤波器与 Mel 频率倒谱系数的耳机到鼓膜的传递函数平滑方法,用于模拟人耳听觉感知特性和平滑耳机到鼓膜的传递函数,使均衡后的幅频响应保持相应的峰谷,避免了幅度峰谷过渡均衡。实验结果表明,进行耳机到鼓膜的传递函数平滑的幅度均衡对提高耳机的音色有显著作用,基于 Mel 频率倒谱系数平滑的幅度均衡对提高耳机的音色最为显著。

关键词:耳机到鼓膜的传递函数;Mel频率倒谱系数;平滑;均衡;耳机声重放
中图法分类号:TN64
文献标识码:A
文章编号:1000-310X(2023)01-0067-09
DOI: 10.11684/j.issn.1000-310X.2023.01.009

MFCC-based smoothing for equalization of headphone-to-eardrum transfer function

LI Guangju^{1,2} LUO Pingzhan³ QIAN Peng¹ GAN Weiming¹ XING Meng¹

(1 Haikou Branch, Institute of Acoustics, Chinese Academy of Sciences, Haikou 570105, China)

(2 Key Laboratory of Noise and Vibration Research, Chinese Academy of Sciences, Beijing 100190, China)

(3 Harman Technology Shenzhen Co., Ltd., Shenzhen 518067, China)

Abstract: Binaural headphone reproduction can be improved by appropriate equalization of headphone-toeardrum transfer function (HETF). Equalization by direct inversion of HETF targeting at a flat frequency response cannot keep the beneficial peaks and valleys caused by pinna and ear canal filtering. To simulate human auditory filtering and smooth HETF, this research proposes two smoothing methods based on roex filtering and Mel-frequency cepstrum coefficients (MFCC). Thus, the peaks and valleys of the response are preserved to avoid over equalization. Subjective experiments show that perception of binaural headphone reproduction can be improved by smoothing HETF, and the performance is most significant by smoothing method based on MFCC.

Keywords: Headphone-to-eardrum transfer function; Mel-frequency cepstrum coefficients; Smoothing; Equalization; Headphone reproduction

²⁰²¹⁻¹¹⁻²² 收稿; 2022-05-18 定稿

^{*}海南省自然科学基金青年基金项目(619QN265),海南省科协青年科技英才创新计划项目(QCXM201809)

作者简介:李光炬(1989-),男,海南海口人,博士,助理研究员,研究方向:语声和声频信号处理。

[†]通信作者 E-mail: liguangju@mail.ioa.ac.cn

0 引言

耳机是一种直接与人耳耦合并在容积约为 2 cm^3 空腔内产生声压的电声换能器^[1], 它作为最 常用的声重放设备,在现代通信、虚拟现实、听力 康复、智能硬件等领域应用广泛。采用耳机进行 声重放时,双耳声信号依次经耳机重放、耳廓耦 合与耳道传输到达鼓膜,引入了耳机到鼓膜的传 递函数 (Headphone-to-eardrum transfer function, HETF)^[2]。耳机和耳道的非理想传输特性会影响重 放中鼓膜接收声压的正确性,因而需要进行耳机 均衡^[3],但由于HETF具有非常多的窄峰和谷^[4], 直接用未平滑的HETF逆滤波器^[5]进行耳机均衡 不但信号处理困难^[6-7],且容易引起误差^[8-9],效 果适得其反,因而需要对HETF做平滑处理。文 献[10]采用了广义分数倍频程对HETF进行平滑, 但这种广义分数倍频程与人耳听觉感知的频率分 辨率不匹配。文献[11-12]通过统计方法对多次测 量的HETF近似取平均值,该方法虽然避免了均衡 HETF时引入高Q值频谱峰,但并未考虑人耳的听 觉感知特性。文献[13]提出了一种用于HETF平滑 与均衡的自动正则化方法,通过测量的HETF自动 估计正则化因子,使均衡后的耳机产生更好的听觉 感知效果。

基于上述背景,本文考虑到人耳听觉感知特性 与耳机均衡实践经验,采用传统的1/3倍频程幅度 平滑方法^[10]和提出的两种幅度平滑方法对HETF 进行平滑再做均衡, 使均衡后的 HETF 保持一些 有助于人耳听觉感知的重要峰谷,以期更真实地 仿真人耳对高频区域 HETF 峰谷的感知特性,并避 免HETF 过渡均衡引入高Q值频谱峰,最终达到优 化耳机的听觉感知效果。一种幅度平滑方法是基 于roex滤波器平滑,该方法基于听觉滤波器的等 效矩形带宽来模拟人耳基底膜对频率的选择性与 分辨率。另一种幅度平滑方法是基于 Mel 频率倒谱 系数 (Mel-frequency cepstrum coefficients, MFCC) 平滑,该方法根据 Mel 尺度的频率仿真人耳对频率 感知的非线性特征,能够充分模拟人耳的听觉感知 特性。研究结果表明,基于 MFCC 平滑的幅度均衡 对提高耳机的音色最为显著。

1 耳机均衡方法

耳机重放时引入了HETF,实际到达聆听者鼓

膜处的声信号与馈给耳机的声信号存在以下频域 关系:

$$E(k) = E_p(k)M(k), \tag{1}$$

其中, E(k) 是鼓膜处实际声信号的第k个频谱分量, $E_p(k)$ 表示馈给耳机声信号的第k个频谱分量, M(k) 表示HETF的第k个频谱分量。为了消除耳机传递特性的影响,可将实际的双耳信号声信号 $E_d(k)$ 经均衡滤波器滤波后,再馈给耳机重放,表示如下:

$$E_p(k) = H(k)E_d(k), \tag{2}$$

其中,H(k)表示耳机均衡滤波器的频域特性。 为了在鼓膜处准确地重放双耳声信号,要求 $E(k) = E_d(k)$,根据式(1)和式(2)有

$$H(k) = 1/M(k). \tag{3}$$

可见,理想的耳机均衡滤波器是HETF的逆 滤波器。为了保证均衡滤波器满足因果性和稳定 性^[14],均衡目标函数需要加上一定的时延。耳机均 衡时若无约束抬升HETF的低频与高频响应,可能 引起耳机系统响应的畸变、失真甚至损坏耳机。因 此,需要根据耳机的HETF来设计均衡滤波器的均 衡目标函数,以最大化的减小均衡误差。

本文采用有限冲激响应(Finite impulse response, FIR)均衡方法^[15]对实验耳机的HETF进 行均衡。该均衡方法的设计基于最小二乘准则与正 则化滤波器使均衡误差最小化。已有研究^[15-16]表 明,正则化滤波器有助于减小系统响应的时域混叠。 该均衡方法的频域表达式为

$$H(k) = \frac{D^*(k)M(k)}{M(k)M^*(k) + \beta \cdot B(k)B^*(k)},$$
 (4)

其中, β表示正则化滤波器加权标量, B(k)表示正则化滤波器响应的傅里叶变换, D(k)表示理想带通滤波器响应的傅里叶变换。

2 HETF 平滑方法

2.1 1/3 Octave 平滑

1/3倍频程(1/3 Octave)^[10]作为声学频率的 一种相对尺度,广泛应用于声频与声学频谱分析 等领域。它基于人耳的听觉相对声音的大小和 频率具有对数关系的原则,对可听声的频率范围 20 Hz~20 kHz进行了划分,能够很好地体现信号带 宽的能量分布。假设上下截止频率各为 f_u 与 f_l 的 频带 $(f_u > f_l)$, 中心频率为 f_c , 满足以下关系式:

$$\begin{cases} f_u = 2^n \cdot f_l, \\ f_c = \sqrt{f_u f_l}, \end{cases}$$
(5)

其中, n 表示倍频程, 当n = 1/3时, 上下截止频率 的关系为 1/3 倍频程, f_c 为上下截止频率的几何平 均值。采用 1/3 Octave 平滑方法对 HETF 做平滑处 理: 首先对耳机的瞬态响应进行傅里叶变换并求出 幅度谱, 然后根据式 (5) 计算每个中心频率 f_c 的带 宽内幅度谱的平均值, 最后合成 1/3 倍频程幅度谱 为平滑后的 HETF。

2.2 基于 roex 滤波器平滑

耳机听觉感知效果与人耳听觉生理有关。因 人耳基底膜的频率选择特性,人耳的频率分辨率随 频率增大而递减,而听觉滤波器可仿真人耳基底膜 的频率选择特性和分辨率。由于 roex 滤波器符合 耳蜗非对称与强度依赖的听觉滤波特性^[17],采用 roex 滤波器平滑 HETF 可以很好地仿真人耳基底 膜对频率的选择特性和分辨率,达到减小 HETF 的 尖锐峰谷被过度均衡的目的。Roex 滤波器^[18] 的表 达式为

$$W(g) = (1+r)(1+pg) \cdot e^{-pg} + r, \qquad (6)$$

其中,W(g)为听觉滤波器的形状,g表示与听觉滤 波器的频率 f_i 相对于中心频率 f_c 的偏差,p和r是 与滤波器形状有关的参数,p决定滤波器的尖锐程 度,r用于限制滤波器的动态响应范围。对正常听 力人群,听觉滤波器的等效矩形带宽(Equivalent rectangular bandwidth, ERB)与p的关系为

$$ERB = 4f_c/p.$$
 (7)

ERB与中心频率 f_c 之间的关系为

$$\text{ERB} = 24.7 \times (0.00437 f_c + 1). \tag{8}$$

根据式(6)、式(7)与式(8)得到roex滤波器的形状 与中心频率的关系为

$$\times \exp\left(-\frac{4f_c |f_i - f_c|}{24.7(0.00437f_c + 1)f_c}\right).$$
(9)

对W(f_i)进行能量归一化处理为

$$W_n(f_i) = W(f_i) / \sum_{i=1}^N W(f_i),$$
 (10)

其中,N表示采样频率的一半对应的离散频率序列。 设平滑前的频域幅度为 $H(f_i)$,则平滑后的频域幅 度 $H_s(f_i)$ 表示为

$$H_s(f_i) = \sqrt{\sum_{i=1}^{N} H(f_i)^2 \cdot W_n(f_i)}.$$
 (11)

2.3 基于 MFCC 平滑

人耳对频率的分辨率不是线性的,对低频的分 辨率高于高频的分辨率,线性刻度下的HETF平滑 不符合人耳听觉特性。MFCC^[19]将频率转换为Mel 频率,在1000 Hz以下人耳对声信号的感知能力与 频率呈近似线性关系,在1000 Hz以上则与频率呈 对数关系,能够充分模拟人耳听觉感知特性。Mel频 率 f_{Mel} 与频率之间的转换关系表示如下^[20]:

$$f_{\rm Mel} = 2595 \times \lg(1 + f_{\rm Hz}/700),$$
 (12)

其中, f_{Hz}表示线性频率,单位为Hz。采用基于MFCC平滑方法对HETF做平滑处理:首先将HETF的频谱转化为Mel频率域上的非线性频谱,其次将非线性频谱转化到倒谱域上得到Mel频率倒谱系数,最终使HETF的频谱转换到可以被人耳感知的频域中,实现更好的耳机听觉感知效果。平滑HETF流程如图1所示,平滑过程如下所述3步骤:

步骤 1:由x(n)表示 HETF,为了减小频谱泄漏, 采用 hamming 窗对x(n)进行加窗处理得到 $x_w(n)$, 然后对 $x_w(n)$ 进行快速傅里叶变换 (Fast Fourier transform, FFT) 得到的线性频谱为

$$X_w(k) = \sum_{n=0}^{N-1} x_w(n) e^{-j2\pi nk/N}, 0 \le n, \ k \le N-1,$$
(13)



步骤2: 计算 $X_w(k)$ 的能量谱E(k),并将E(k)通过Mel频率滤波器组得到S(k),相当于E(k)与Mel频率滤波器组的频域响应 $H_m(k)$ 相乘,并通过式(12)将实际线性频率尺度转换为Mel频率尺度,即Mel频谱为

$$S(k) = E(k) \cdot H_m(k). \tag{14}$$

步骤 3: 对 S(k) 进行对数能量处理得到对数频 谱 S(m) 如式 (15) 所示,并将 S(m) 经过离散余弦变 换 (Discrete Cosine transform, DCT) 转化到倒频 谱域,得到 Mel 频率倒谱系数 mfcc_M 如式 (16) 所示, 再利用 mfcc_M 重构 x(n) 得到平滑后的 HETF_{mfcc}。

$$S(m) = \ln[S^{2}(k) \cdot H_{m}(k)], \quad 0 \leq m \leq M, \quad (15)$$
$$\operatorname{mfcc}_{M} = \sum_{m=0}^{M-1} S(m) \cdot \cos\left[\frac{\pi(m-1/2)}{M}\right],$$
$$0 \leq m \leq M, \quad (16)$$

其中,*M*表示平滑参数,Mel频率滤波器组为设置的 *M*个带通滤波器 *H_m(k)*,每个滤波器具有三角形滤 波特性,而三角形滤波器是对人耳前端听觉处理中 沿基底膜的冲激响应的建模。

3 HETF 实验测量

按照耳机的佩戴方式不同分类,将耳机分为 單耳式耳机、贴耳式耳机、耳塞式耳机与入耳式耳 机。由于耳机的佩戴方式不同影响了HETF的传递 特性,因此需要选择测量可重复性较好的耳机以提 高均衡的有效性。实验采用一款罩耳式耳机与一 款入耳式耳机进行HETF测量,如图2(a)与图2(b) 所示编号分别为HP1、HP2,耳机的参数及耳机价 格如表1所示。HETF的测量可取耳道入口至鼓膜 之间的任一参考点,这是因为声波在耳道内可以 近似认为是一维传递的^[2],采用耳机到鼓膜之间任 意一点的传递函数进行耳机均衡,不影响耳机的均 衡效果^[7]。对于真人受试者,采用探针传声器在鼓膜前 1~2 mm 位置处测量其 HETF 具有一定的困 难性与危险性,而且探针传声器在鼓膜处的测量位 置不易控制,实际测量的 HETF 会引入一定的测量 误差。为了减小这些因素的影响,通常采用一个通用 HETF 来表征真人受试者的 HETF。本文采用 B&K 人工头测量 HETF,传声器内置于 B&K 人工 头鼓膜处作为测量参考点,利用 B&K 人工头内置 的传声器捡拾声信号测量 HETF。

图 2(c) 与图 2(d) 是 HETF 测量的 B&K 人工头 及其系统框图。首先,在消声室内采用最大长度 序列 (Maximum length sequences, MLS) 信号作为 激励信号,将由计算机产生的MLS信号(长度为 8191,8次平均,44.1 kHz采样,16 bits 量化) 经B&K PULSE的D/A(Digital/Analog数字/模拟)馈给戴 在B&K人工头(Type 4128C)上的耳机进行重放; 其次,由B&K人工头鼓膜处的传声器B&K 4192 捡 拾耳机重放的声信号,实验过程中采用B&K 4192 自由场响应的逆函数进行滤波以消除传声器的影 响;最后,再经B&K PULSE的A/D变换器输入到 计算机进行记录,将记录到的信号进行解卷积以及 傅里叶变换,即得到HETF。每款耳机进行4次重复 测量,每次测量后将耳机取下并重新戴上,然后进行 新一次测量。为了减小实验过程中的耳机声泄漏, 尽量使耳机与B&K人工头的耳廓、耳道紧密耦合。

表1 实验耳机的参数与价格

Table 1The prices and descriptions of theheadphones

编号	名称	阻抗/ Ω	频响范围	类型	单价
HP1	Sennheiser HD650	300	$10~{\rm Hz}{\sim}41~{\rm kHz}$	罩耳式	¥2659
HP2	Sennheiser IE800	16	$8~{\rm Hz}{\sim}41~{\rm kHz}$	入耳式	¥3999



Fig. 2 The experimental setup of HETF measurements and its block diagram

图3为上述两款实验耳机在B&K人工头上的 HETF四次可重复性测量结果,以幅频响应和标 准偏差的形式给出,其频率范围为20 Hz~20 kHz。 因耳机左右耳的HETF具有对称性,只给出左耳 的HETF测量结果。图3(a)与图3(c)表明HP1的 HETF测量可重复性较好,其幅频响应的标准偏差 在20 Hz~8 kHz范围内均在3 dB以下。罩耳式耳 机的测量可重复性不仅受耳廓耦合作用与耳道共 振作用的影响^[21],还与耳机的耳罩大小有关,重

可用于耳机均衡。 1510 10 50 幅度/dB 幅度/dB 0 -5-10-10200.020 0.020 0.10010200.10010 1 20频率/kHz 频率/kHz (a) HP1的幅频响应 (b) HP2的幅频响应 10 10 8 8 标准偏差/dB 示准偏差/dB 6 6 4 4 $\mathbf{2}$ 0-0.0200.020 0.100 20 0.100 10 20 10 1 频率/kHz 频率/kHz (c) HP1的标准偏差 (d) HP2的标准偏差



4 客观平滑与均衡

耳机左右耳的HETF具有对称性,只给出上述 两款实验耳机HP1、HP2左耳的客观平滑与均衡结 果,其中客观平滑结果由4.1节给出,客观均衡结果 由4.2节给出。

4.1 客观平滑结果与分析

采用本文1/3 Octave 平滑、基于 roex 滤波器平 滑与基于 MFCC 平滑对 HP1、HP2 四次重复性测量 的 HETF 进行平滑处理, 平滑结果如图4所示, 通 过基于 roex 滤波器平滑的 HETF 均增加了 30 dB 的偏移量, 通过1/3 Octave 平滑的 HETF 均增加了 60 dB 的偏移量。由图4可知, 1/3 Octave 平滑方法、 基于 roex 滤波器平滑方法以及基于 MFCC 平滑方 法能够很好平滑 HETF 高频区域的峰谷结构,可以 明显减小4次重复性测量的 HETF 在高频窄带内的 差异。

复佩戴时耳罩较小的耳机对耳廓有压迫作用,在

9~20 kHz 范围内 HP1 的幅频响应出现不同 Q 值的

峰谷结构。图3(b)与图3(d)表明HP2具有很好的

HETF测量可重复性,在20 Hz~14 kHz范围内幅

频响应的标准偏差小于1 dB。这是由于入耳式耳机

直接与耳道耦合,避免了耳廓因素的影响,重复佩戴

时耳机位置的变化对 HETF 影响很小。因此,上述

实验验证了HP1和HP2的HETF测量可重复性好,

4.2 客观均衡结果与分析

图 3(a) 与图 3(b) 反映了耳机输出信号的幅度 随频率变化而变化的规律, 影响了耳机声信号的保 真度与染色度, 可用 HETF 的均方误差来衡量。均 方误差函数的表达式^[22] 为

$$e_{\omega} = \left[\frac{1}{N_f - N_i + 1} \sum_{k=N_i}^{N_f} \left(20 \lg |H_{\rm EQ}(k) - \bar{H}_{\rm EQ}|\right)^2\right]^{\frac{1}{2}},\tag{17}$$

$$\bar{H}_{\rm EQ} = \left[\frac{1}{N_f - N_i + 1} \sum_{k=N_i}^{N_f} \left(20 \lg |H_{\rm EQ}(k)|\right)\right]^{\frac{1}{2}},\tag{18}$$

其中, N_i 与 N_f 分别表示均衡起始频率与均衡终止 频率的下标,均衡误差函数 e_{ω} 只在均衡的频段内 考量, $H_{EQ}(k)$ 表示均衡后系统响应的傅里叶变换, \bar{H}_{EQ} 表示均衡后的系统响应傅里叶变换的模的均 值。根据图3(a)与图3(b)设计耳机均衡滤波器的均 衡目标函数,选择的均衡起始频率 f_i 、均衡终止频率 f_f 以及均衡频段 $f_i \sim f_f$ 如表2所示。

表2 耳机均衡参数







采用本文第1节的FIR均衡方法分别对原始 4次测量HETF的均值与平滑4次测量HETF的均 值进行均衡,并利用式(17)与式(18)计算均衡前 后HETF的均方误差,其均衡结果如图5与表3所 示。从图5与表3中可知,未经过平滑的FIR均衡, 其HETF未保持与耳廓耳道滤波相关的峰谷,均 衡频段内的 HETF 趋于平直,其均方误差最小分别 为0.61 dB 与0.73 dB; 而1/3 Octave 平滑的 FIR 均 衡、基于 roex 滤波器平滑的 FIR 均衡与基于 MFCC 平滑的 FIR 均衡,其HETF 保持了可能与耳廓耳 道滤波相关的低 Q 值峰谷,高频区域的低 Q 值 (约 3 dB)频谱峰不会引起时域振铃效应 (ringing artifacts)^[21-22],且在低频区域保留了一个鼓起约 3 dB 包络使耳机得到高质量的低声,其均方误差大小相 当。与未经过平滑的 FIR 均衡比较,1/3 Octave 平 滑的 FIR 均衡、基于 roex 滤波器平滑与基于 MFCC 平滑的 FIR 均衡使 HETF 保持了适当的峰谷,既避 免了 HETF 过度均衡又避免了引入高 Q 值频谱峰。





Fig. 5 Equalization results before and after HETF smoothing

表3 均衡前后 HETF 的均方误差

Table 3 The mean square error of HETFbefore and after equalization

(单位:dB)	(自	自位	: d	B)
---------	----	----	-----	----

		未经过平	1/3 Octave	基于 roex 滤	基于 MFCC	
编号	原始	滑的 FIR	平滑的FIR	波器平滑的	平滑的 FIR	
		均衡	均衡	FIR 均衡	均衡	
HP1	3.81	0.61	2.52	2.48	2.46	
HP2	3.99	0.73	2.35	2.32	2.33	

5 主观测听

5.1 测试准备

安排了主观测听实验来验证不同的耳机处理 方式对耳机声重放造成的影响。实验声频由录制 的两类声音素材(弦乐、对白)与耳机HP1、HP2的 HETF的逆函数卷积得到。弦乐与对白涵盖不同的 频率范围,有利于进行比较全面的主观比较。实验 声频经耳机HP1与HP2播放给受试者聆听。测试 在中国科学院声学研究所测听室内进行,共有30名 受试者参与主观测听实验,其听力正常、年龄介于 25~35岁之间,其中15人的听声经验丰富,其余15 人有多次参与听声实验的经历。

主观测听实验采用 MUSHRA 方法^[23] (主观测 听实验不设置参考实验声频,所有实验声频均进 行盲听),要求所有受试者对实验声频的音色进行 评价。由于ITU-RBS.1534-1是一种针对声频质量 评价的主观测试标准,偏重于测试带宽较高的声 频信号,多用于数字声频存储与声频流媒体等领 域,因此评分标准参考ITU-RBS.1534-1 的主观测 试标准^[24-25]。评分等级为100~80非常好,80~60 好,60~40一般,40~20差,20~0非常差,即实验声 频音色越好评分越高。受试者通过 MUSHRA 音色 测试界面进行操作,点击测试界面的播放按钮聆 听实验声频,可以反复播放多次,最后拖动滑动按 钮给出评分。在主观测听实验之前对实验声频进 行编号,具体做法为每名受试者随机听取两类实验 声频,每类实验声频包含5种耳机处理方式,分别 为原始(TS1)、未经过平滑的FIR均衡(TS2)、1/3 Octave 平滑的 FIR 均衡 (TS3)、基于 roex 滤波器平 滑的FIR均衡(TS4)、基于MFCC平滑的FIR均衡

(TS5)。在正式听声实验前,每名受试者需要完成至 少一次完整试听比较练习,以便熟悉实验评分规则。

5.2 测试结果与分析

测试结果采用单因素方差(one-way ANOV A)^[26]分析TS1、TS2、TS3、TS4、TS5对耳机声重放 影响的差异性。P表示统计学差异,取P<0.05具 有显著统计学差异。30名受试者的主观测听实验平 均得分与标准偏差,如图6所示,表示弦乐、对白的 主观测听实验平均得分与标准偏差。图6(a)表示罩 耳式耳机HP1的弦乐、对白的主观测听实验的平均 得分与标准偏差,TS5的弦乐与对白平均得分72.8 与68.7为最高,TS3、TS4的弦乐与对白平均得分相 当为次高分。图6(b)表示入耳式耳机HP2的弦乐、 对白的主观测听实验的平均得分与标准偏差,TS5 的弦乐与对白平均得分75.5 与73.9为最高, TS3、 TS4的弦乐与对白平均得分相当为次高分。弦乐与 对白的平均得分与标准偏差相似,而且平均得分在 40与80之间分布。这说明了在不同的耳机处理方 式下弦乐与对白的听声感知(音色)表现出相同的 趋势。通过one-way ANOVA进行fisher LSD 事后 检验,检验分别在弦乐与对白中的5种耳机处理方 式之间的统计学差异,结果如表4与表5所示,其中 粗体标记的罗马数字表示不具有显著统计学差异, 其余表示具有显著统计学差异。HP1的弦乐与对 白组间比较统计结果(见表4)显示,TS3、TS4、TS5 显著比TS1与TS2高 (P < 0.05), TS3、TS4与TS5 均对HETF进行平滑,而TS1与TS2没有对HETF 做平滑处理,这说明了进行幅度平滑的耳机均衡有 利于提高耳机主观音色。表4还显示, TS5显著比 其余耳机处理方法高(P<0.05),TS5提高耳机主观





Fig. 6 The mean score and standard deviation of subjective evaluation test for thirty subjects

表4 弦乐与对白的组间比较统计结果(HP1)

Table 4 Statistical between-group comparison of string and dialogue(HP1)

从田子计	Р					
处理力法	TS1	TS2	TS3	TS4	TS5	-
	(45.5 ± 6.12)	(55.1 ± 4.52)	(62.5 ± 3.36)	(63.0 ± 3.55)	(68.7 ± 4.79)	
$TS1(47.7\pm6.29)$		0.000^{*}	0.000^{*}	0.000^{*}	0.000^{*}	
$TS2(58.7\pm5.07)$	0.000*		0.000^{*}	0.000^{*}	0.000^{*}	オ白
$TS3(66.2 \pm 3.86)$	0.000*	0.000*		1.000^{*}	0.000^{*}	N1 [1
$TS4(66.7 \pm 4.49)$	0.000*	0.000*	1.000*		0.000^{*}	
$TS5(72.8\pm 5.06)$	0.000*	0.000*	0.000*	0.000*	_	
弦乐					-	

注:*P < 0.05,括号内表示(均值±标准偏差)。

表5 弦乐与对白的组间比较统计结果 (HP2)

Table 5	Statistical	between-group	comparison	of string and	dialogue(1	HP2)	ļ
---------	-------------	---------------	------------	---------------	------------	------	---

は四子が	Р					
处理力法	TS1	TS2	TS3	TS4	TS5	-
	(50.0 ± 6.45)	(58.6 ± 4.84)	(66.1 ± 4.52)	(66.4 ± 4.59)	(73.9 ± 6.05)	
$TS1(50.7\pm5.58)$		0.001^{*}	0.000^{*}	0.000^{*}	0.000^{*}	
$TS2(58.9 \pm 4.78)$	0.001^{*}		0.000^{*}	0.000^{*}	0.000^{*}	オ白
$TS3(67.1\pm 5.46)$	0.000*	0.000*		1.000^{*}	0.000^{*}	V1 LI
$TS4(67.4 \pm 5.51)$	0.000*	0.000*	1.000*		0.000^{*}	
$TS5(75.5\pm5.74)$	0.000*	0.000*	0.000*	0.000*		
弦乐						

注:*P < 0.05,括号内表示(均值±标准偏差)。

音色显著,TS3与TS4之间无显著性差异(P > 0.05),TS2实现以平直的HETF为目标的幅度均衡,这也说明以平直HETF为目标的幅度均衡对耳机声重放时的音色提升有限,市面高端耳机的频响设计中往往经验性的保留一些峰谷,也是基于类似考虑。HP1的弦乐与对白组间比较统计结果(见表5),表明对白具有与弦乐类似的特征,这也说明了以平直HETF为目标的幅度均衡不利于耳机重放时音色的提升,而耳机均衡时需要考虑人耳听觉特性与根据耳机均衡实践经验适当保持一些有助于耳机听觉感知效果的幅度峰谷。

6 结论

本文采用了 1/3 Octave 平滑、基于 roex 滤波器 平滑、基于 MFCC 平滑的幅度均衡方法对罩耳式耳 机与入耳式耳机进行均衡, 验证了这些幅度均衡方 法对耳机声重放的影响。客观实验结果表明, 与以 实现平直幅频响应为目标的幅度均衡相比, 1/3 Octave 平滑、基于 roex 滤波器平滑和基于 MFCC 平 滑的幅度均衡使耳机幅频响应保持了某些与耳廓 耳道滤波相关的峰谷,同时避免了幅度过度均衡所 引入影响人耳听觉感知的高Q值频谱峰。主观测听 实验结果表明,进行 HETF 平滑的幅度均衡有利于 提高耳机的主观音色,基于 MFCC 平滑的幅度均衡 对提高耳机的主观音色最为显著,而以平直幅频响 应为目标的幅度均衡对耳机声重放的音色提升有 限。因此,耳机均衡需结合人耳听觉特性考虑与听 觉滤波相关的峰谷,并非优先考虑平直的耳机目标 频响曲线。

参考文献

- [1] 马大猷, 沈蠔. 声学手册 [M]. 修订版. 北京: 科学出版社, 2004.
- [2] Møller H. Fundamentals of binaural technology[J]. Applied Acoustics, 1992, 36(3/4): 171–218.

- [3] Møller H, Sorensen M F, Jensen C B, et al. Binaural technique: do we need individual recordings?[J]. The Journal of the Audio Engineering Society, 1996, 44(6): 451–469.
- [4] Moller H, Hammershoi D, Jensen C B, et al. Transfer characteristics of headphones measured on human ears[J]. The Journal of the Audio Engineering Society, 1995, 43(4): 203–217.
- [5] Wightman F L, Kistler D J. Headphone simulation of freefield listening. I: stimulus synthesis[J]. The Journal of the Acoustical Society of America, 1989, 85(2): 858–867.
- [6] Lorho G. Subjective evaluation of headphone target frequency responses[C]//Audio Engineering Society Convention 126. Audio Engineering Society, 2009.
- [7] Fleischmann F, Silzle A, Plogsties J. Identification and evaluation of target curves for headphones[C]//Audio Engineering Society Convention 133. Audio Engineering Society, 2012.
- [8] Hiekkanen T, Mäkivirta A, Karjalainen M. Virtualized listening tests for loudspeakers[J]. Journal of the Audio Engineering Society, 2009, 57(4): 237–251.
- [9] McAnally K I, Martin R L. Variability in the headphoneto-ear-canal transfer function[J]. Journal of the Audio Engineering Society, 2002, 50(4): 263–266.
- [10] Hatziantoniou P D, Mourjopoulos J N. Generalized fractional-octave smoothing of audio and acoustic responses[J]. The Journal of the Audio Engineering Society, 2000, 48(4): 259–280.
- [11] Masiero B, Fels J. Perceptually robust headphone equalization for binaural reproduction[C]//Audio Engineering Society Convention 130. Audio Engineering Society, 2011.
- [12] Lorho G. Subjective evaluation of headphone target frequency responses[C]//Audio Engineering Society Convention 126. Audio Engineering Society, 2009.
- [13] Bolanos J G, Makivirta A, Pulkki V. Automatic regularization parameter for headphone transfer function inversion[J]. Journal of the Audio Engineering Society, 2016, 64(10): 752–761.
- [14] Ryan C, Furlong D. Effects of headphone placement on headphone equalization for binaural reproduction[C]//Audio Engineering Society Convention 98. Au-

dio Engineering Society, 1995.

- [15] Scharer Z, Lindau A. Evaluation of equalization methods for binaural signals[C]//Audio Engineering Society Convention 126. Audio Engineering Society, 2009.
- [16] Lavoie M C, Norcross S G, Soulodre G A. Distortion audibility in inverse filtering[C]//Audio Engineering Society Convention 117. Audio Engineering Society, 2004.
- [17] Patterson R D. Auditory filters and excitation patterns as representations of frequency resolution[J]. Frequency Selectivity in Hearing, 1986: 123–177.
- [18] Moore B C, Peters R W, Glasberg B R. Auditory filter shapes at low center frequencies[J]. The Journal of the Acoustical Society of America, 1990, 88(1): 132–140.
- [19] Davis S B, Mermelstein P. Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980, 28(4): 357–366.
- [20] Muda L, Begam M, Elamvazuthi I. Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques[J]. Ttps, 2010, 2.
- [21] Paquier M, Koehl V. Discriminability of the placement of supra-aural and circumaural headphones[J]. Applied Acoustics, 2015, 93: 130–139.
- [22] Li G J, Zheng C S, Li X D, et al. Evaluation of headphone phase equalization on sound reproduction[J]. Applied Acoustics, 2019, 156: 208–216.
- [23] Vincent E. MUSHRAM-A Matlab interface for MUSHRA listening tests (version 1.0)[EB/OL]. [2015-12-08]. http://c4dm.eecs.qmul.ac.uk/downloads/#mushram.
- [24] ITU Recommendation BS.1534-1. Method of the subjective assessment of intermediate quality level of coding systems[EB/OL]. [2015-12-08]. http://www.itu.int/dms_ pubrec/itu-r/rec/bs/R-REC-BS.1534-2-201406-S!!PDF-E.pdf.
- [25] Liebetrau J, Nagel F, Zacharov N, et al. Revision of Rec. ITU-R BS. 1534[C]//Audio Engineering Society Convention 137. Audio Engineering Society, 2014.
- [26] 朱星宇, 陈勇强. SPSS 多元统计分析方法及应用 [M]. 北京: 清华大学出版社, 2011.