

基于偏最小二乘回归的与头相关传递函数的个人化

胡红梅^{①②} 周琳^① 马浩^① 吴镇扬^①

^①(东南大学信息科学与工程学院 南京 210096)

^②(江苏大学机械学院 镇江 212013)

摘要: 该文提出了一种基于偏最小二乘回归(Partial Least Squares Regression, PLSR)的与头相关传递函数(Head Related Transfer Function, HRTF)的个人化方法。通过对 HRTF 进行预处理和主元分析,并对人体参数进行筛选,只要相对简单的人体参数测量,就可利用 PLSR 得到特定人的 HRTF。客观误差分析和主观声音定位测试结果表明估计的 HRTF 与实际测量的 HRTF 之间不仅均方误差较小,而且感知区别不大;同时由 PLSR 估计的个人化 HRTF 在水平面上的主观测试定位准确率明显优于非个人化 HRTF,也优于由最小二乘回归(Least Squares Regression, LSR)估计的个人化 HRTF。

关键词: 偏最小二乘回归; 最小二乘回归; 与头相关传递函数; 声音定位测试

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2008)01-154-05

Head-Related Transfer Function Personalization Based on Partial Least Square Regression

Hu Hong-mei^{①②} Zhou Lin^① Ma Hao^① Wu Zhen-yang^①

^①(School of Information Science and Engineering, Southeast University, Nanjing 210096, China)

^②(School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: A method is proposed to obtain individual Head Related Transfer Function (HRTF) based on the Partial Least Square Regression (PLSR), in which anthropometric parameters are selected by correlation analysis and the original HRTF data are preprocessed before principal component analysis. Then objective simulation experiment and subjective sound localization experiment are implemented to evaluate the method. The experiments show that the estimated HRTF not only has small mean square error, but also has little perception difference from the measured one; the localization performance of the individual HRTF estimated by PLSR presented in this article is more accurate than that of the non-individual HRTF obtained from the database and also more accurate than that estimated via Least Square Regression (LSR).

Key words: Partial least square regression; Least square regression; Head related transfer function; Sound localization experiment

1 引言

基于与头相关传递函数(Head Related Transfer Function, HRTF)的虚拟听觉空间系统是当今虚拟现实技术的研究热点之一。HRTF 反映了外耳、头部、躯干对来自不同方位的声音信号的综合滤波效果,描述了声源到耳道口的传播特征^[1],其时域表示称为与头相关脉冲响应(Head Related Impulse Response, HRIR)。它既与声源相对于听者的方向有关,也因人体某些部位的形状及大小不同而存在个体差异。所以 HRTF 的个人化成了理想虚拟听觉空间合成中一个重要而急待解决的问题。

目前绝大部分空间听觉系统都基于通用 HRTF,这种系统忽略了 HRTF 的个体差异,常常会存在明显的头中效应、

前后颠倒以及仰角误判等现象^[1,2]。因此获取精确的个人化 HRTF 是得到理想 3D 声场的前提。但是 HRTF 直接测量法需要专门的设备和技术,测量过程复杂耗时^[3],所以很多学者提出了只通过简单而不昂贵的测量获取个人化 HRTF 的方法,如数据库匹配法^[4]、频谱修正法^[5]和边界元法^[6]。其中边界元法要求计算机有强大的计算能力和海量的内存以及一套三维形状测量系统。于是一些学者试图通过统计方法来建立人体参数和 HRTF 模型特征参数之间的关系^[7,8]。本文利用统计的思想,在前一阶段工作^[8]的基础上,通过对 HRIR 数据的处理和人体参数的筛选,提出了一种改进的 HRTF 个人化方法:首先对数据库中的 HRTF 数据进行预处理和主元分析,然后用相关分析法选取一些主要的人体参数,接着用偏最小二乘回归(Partial Least Squares Regression, PLSR)建立每个方向主元系数与选定人体参数之间的线性方程,最后对特定听者的人体参数进行测量,利用该方程得到的主元

系数来重构听者的 HRTF。

2 HRTF 的个人化方法

本文采用加州大学戴维斯分校图像处理和集成计算中心(CIPIC)实验室提供的数据库^[9], 通过对 CIPIC 数据库中 HRIR 数据的处理和人体参数的筛选, 利用 PLSR 建立不同测试者的 HRTF 与对应人体参数之间的关系, 并将这个关系用于估计特定听者的 HRTF, 实现 HRTF 的个人化。图 1 给出了整个算法流程, 其中点划线框中内容是 HRTF 的对数幅值响应估计部分。图中实线和虚线箭头分别表示模型的建立方向和模型的检验、应用方向。

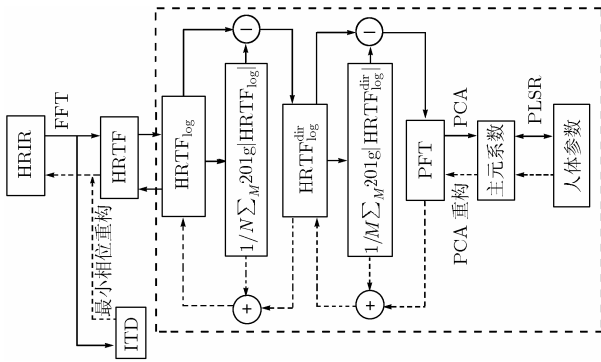


图 1 HRTF 的个人化

2.1 HRTF 的主元分析

CIPIC 提供了 200 维的 HRIR 数据, 为了便于实现, 本文采用了主元分析(Principal Components Analysis, PCA)对 HRIR 进行降维^[10,11]。由于 HRIR 包含了共性特征和个性特征, 为了强调 HRIR 的个性差异, 本文在进行 PCA 分析之前对所有的 HRIR 数据进行了如图 1 所示的预处理: 先求数据库里所有 HRIR 对应的 HRTF, 然后求出去总体均值的对数幅度谱 $HRTF_{\log}^{\text{dir}}$; 接着求出去方向均值的传递函数, 本文称之为该方向上因人而异的个人传递函数(Personal Transfer Function, PTF); 最后将所有经过去均值处理的 PTF 作为一个整体进行 K-L 变换^[11,12], 得到一个正交变换矩阵 C 以及每个 PTF 的权系数, 如 $w_{\theta,k}$ 表示方向 θ 上第 k 个 PTF 对应的权系数(本文称之为 PC 系数)。 $w_{\theta,k}$ 一方面是声源位置的函数, 体现了 PTF 的空间特征; 另一方面还体现了不同听者之间的个性差异。于是方向 θ 上第 k 个人的 PTF $p_{\theta,k}$ 可以表示成基矢量的加权和

$$p_{\theta,k} = Cw_{\theta,k} \quad (1)$$

实验表明 PTF 的第一主元的能量占到 70% 以上, 只要选用前 12 个主元就可以将重构的 PTF 和原始的 PTF 之间的逼近最小均方误差控制在 0.05 以内, 这也是本文选择主元个数的依据。

2.2 人体参数的选取

CIPIC 数据库不仅包含 HRIR 数据, 而且还提供了某些

个体的 27 个人体参数数据^[9]。由于没有文献说明这 27 个人体参数对声音定位都起着举足轻重的作用, 所以有必要根据具体的研究内容选择合适的人体参数。在进行 HRTF 的个人化研究中, 具体采用哪些人体参数不仅影响个人化 HRTF 重建的准确性, 还影响重建的复杂度。为了选择合适的人体参数, 本文采用 Pearson 相关系数作为两变量之间相关程度的判断指标, 对人体参数进行了两次筛选。

首先, 为选出与 HRTF 相关较大的人体参数, 本文对各方向所有人的 HRTF 和相应的人体参数进行了大量的互相关分析, 其 Pearson 相关系数计算公式为

$$R = \left| \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \right| \quad (2)$$

式(2)中 X 表示人体参数矩阵, Y 表示 HRTF 矩阵; \bar{X} 和 \bar{Y} 分别是由矩阵 X 和 Y 中各列均值组成的矩阵, R 是 X 和 Y 的相关系数矩阵, 其中每个元素都介于 0 与 1 之间。为了选取与 HRTF 相关较大的人体参数, 本文用 R 值的大小来判断两变量之间的相关性, R 值越大表示该人体参数对 HRTF 的影响越大。图 2 用灰度值给出了方位角 -45° 方向上两个不同仰角位置 HRTF 和人体参数之间的相关系数。为了表示清晰, 图中只给出了绝对值大于 0.45 以上的 Pearson 相关系数, 其余都置零。综合各方向的相关分析结果, 表明 $x_1, x_3, x_6, x_8 \sim x_{10}, x_{12}, x_{14} \sim x_{17}, d_1, d_3 \sim d_8$ 这 18 个人体参数和 HRTF 关系较强, 于是这些变量就作为一次筛选的结果保留下来, 具体变量定义见文献^[9]。

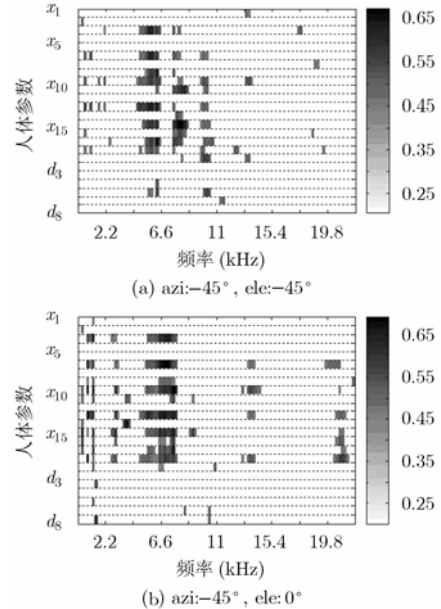


图 2 HRTF 和人体参数之间的相关系数

其次, 为了进一步减少人体参数数量, 降低实际应用中人体参数测量的难度和成本, 本文又对上面 18 个人体参数进行了相关性分析, 结果如图 3 所示, 图中只显示了相关系

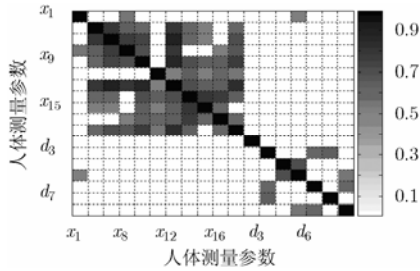


图3 18个人体参数之间的相关系数

数绝对值大于0.5的值。

由图可见有些人体参数间存在较大相关性，利用这些相关信息可以进行变量筛选。例如，图3中 d_8 和 d_6 的相关系数为0.61，因为 d_8 的测量比较困难，所以本文保留了 d_6 而舍弃了 d_8 ；再如 x_3 和 x_6 的相关系数为0.79，但是头部尺寸对HRTF的谱特征影响比较明显，所以 x_6 可以删除。

综上，综合考虑了实现的简单性和可行性，本文最终选择了8个人体参数，其变量和物理意义见表1。

表1 选定的人体参数

变量	x_1	x_3	x_{12}	d_1	d_3	d_4	d_5	d_6
人体参数	头宽	头深度	肩宽	耳甲腔高	耳甲腔宽	耳窝高度	外耳高度	外耳宽度

2.3 偏最小二乘回归

PLSR 主要是在自变量空间里寻找线性组合来解释因变量的变异信息，是经典LSR的拓展。为了建立HRTF与人体参数之间的关系，实现HRTF的个人化，文献[8]利用LSR建立了每个方向上左右耳HRTF的PC系数与相应人体参数之间的线性关系，因为文献[8]中每个方向都使用了相同的自变量，所以在自变量的筛选中尽量选择那些对各方向HRTF都有相当影响程度的8个人体参数。而实际上这8个人体参数对各个方向上的HRTF影响程度并不完全相同，图2也表明了这一点。为了在不丢失重要信息的前提下减少每个方向的回归变量和提高回归模型精度，本文对每个方向分别进行了PLSR分析，而各方向上最终选择的主元个数则由交叉验证法确定。基本思想和实现步骤如下：

(1) 设由2.2节选定的 $K=8$ 个人体参数的 M 个样本组成自变量矩阵 $\mathbf{X}=[x_1 \ x_2 \ \dots \ x_K]_{M \times K}$ ，由2.1节得到的方向 θ 上的PC系数矩阵 \mathbf{W}_θ 中 L 个PC系数的 M 个样本组成因变量矩阵 $\mathbf{Y}=[w_1 \ w_2 \ \dots \ w_L]_{M \times L}$ 。

(2) 对 \mathbf{X} 和 \mathbf{Y} 进行标准化处理，得到标准化的自变量矩阵 \mathbf{E}_0 和因变量矩阵 \mathbf{F}_0 。

(3) 建立PLS的外关系和内关系，并实施 \mathbf{E}_0 和 \mathbf{F}_0 在 \mathbf{t}_1 上的回归：

$$\mathbf{E}_0 = \mathbf{t}_1 \mathbf{p}'_1 + \mathbf{E}_1, \quad \mathbf{F}_0 = \mathbf{t}_1 \mathbf{r}'_1 + \mathbf{F}_1 \quad (3)$$

式(3)中 \mathbf{t}_1 是从 \mathbf{E}_0 中提取的第一主元， \mathbf{p}_1 和 \mathbf{r}_1 是回归系数， \mathbf{E}_1 和 \mathbf{F}_1 为回归残差矩阵，其中参数的求取可以参考文献[13]。

(4) 用残差矩阵 \mathbf{E}_1 和 \mathbf{F}_1 代替 \mathbf{E}_0 和 \mathbf{F}_0 ，重复步骤(3)，直至完成 a 次运算。本文中PLSR利用交叉验证作为运算停止准则，从而确定 a 的值，其具体步骤可参考文献[13]。

(5) 假设经过 a 次主元提取与回归后，回归方程满足精度要求，于是由 a 个主元 $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_a$ 可得 \mathbf{F}_0 的估计值 $\hat{\mathbf{F}}_0$ ：

$$\hat{\mathbf{F}}_0 = \sum_{i=1}^a \mathbf{t}_i \mathbf{r}'_i \quad (4)$$

(6) 按照标准化的逆过程，将式(4)的回归方程还原为 \mathbf{Y} 对 \mathbf{X} 的PLSR方程：

$$\hat{\mathbf{Y}} = f(\mathbf{X}) \quad (5)$$

这样就建立了HRTF与人体参数之间的关系(5)，如果已经测量得到某特定人的 K 个人体参数，只要将这些人体参数代入PLSR方程(5)就可得到PTF的PC系数，然后由这些PC系数重构PTF，最后经过如图1虚线所指方向上的一系列处理得到该特定人最终的个人化HRTF。

3 实验及结果

由于CIPIC数据库中有8个人的上述8个人体参数数据不全，所以这里只用了其余37个人左右耳所有方向共92500($37 \times 2 \times 1250$)个PTF的PC系数和相应的人体参数作为回归分析的实验数据。其中30个人的数据用于建立回归模型，7个人的数据用于检验模型性能。为了检验本文提出的HRTF个人化方法的有效性，下面分别进行了客观误差分析和主观定位测试实验。

3.1 客观误差分析及其结果

为了评价本文提出的估计方法的有效性，以及便于与其它个人化算法进行比较，本文采用谱失真度(SD)^[7]来度量HRTF估计值和测量值之间的逼近误差：

$$SD = \sqrt{\frac{1}{G} \sum_{g=1}^G \left(20 \lg \frac{|H(f_g)|}{|\hat{H}(f_g)|} \right)^2} \quad (\text{dB}) \quad (6)$$

式(6)中 $H(f_g)$ 是测量的HRTF， $\hat{H}(f_g)$ 是估计的HRTF， f 是频率， G 是HRTF数据点数， $G=100$ 。

图4给出了数据库中编号为003和165这两个测试者在0到22.05 kHz频带内左右耳前半平面上的SD值。在该频带内，所有水平方位角的SD平均值小于3.5dB。其整体SD值和平均SD值都普遍小于文献[7]，虽然本文和文献[7]采用的数据库不同，但还是可以推断用本文估计的HRTF能够较好地逼近测量的HRTF。

图5给出了数据库中测试者165左耳在水平 -80° ， -45° ， 45° 和 80° 四个方位角上对数幅度谱的估计值和测量值。由图5可见，除了测量数据的一些突变点外，由人体参数估计的HRTF可以在很大程度上逼近测量值。

综上，本文提出的HRTF个人化方法在SD值和谱逼近

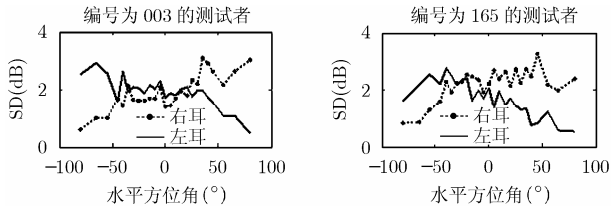


图 4 前半平面左耳和右耳的 SD 值

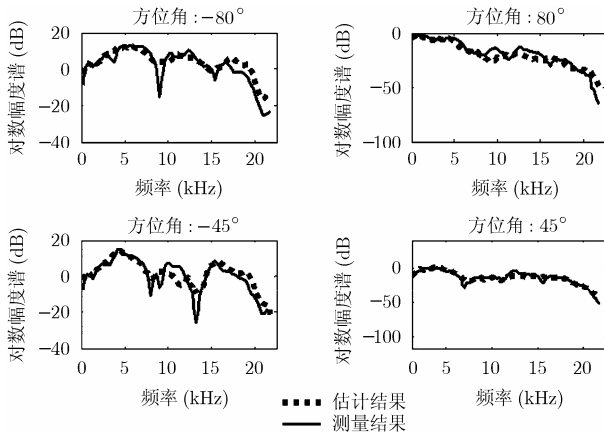


图 5 HRTF 对数幅度谱的估计值和测量值

两方面的性能都优于文献[7], 而且因为 PLSR 综合考虑了自变量之间的冗余信息以及自变量和因变量之间的相关信息, 所以本文结果也优于笔者在文献[8]中提出的基于 LSR 的结果。

3.2 主观定位实验及其结果

为了进一步评价方法的性能, 本文对估计的 HRTF 进行了主观定位实验。主观评价测试系统主要包括一台笔记本电脑, 一副耳塞式耳机以及一套实验室自创的基于 HRTF 虚拟声产生原理的测试软件系统 VasAudio。

实验中激励信号为 8 个持续时间为 0.25s、间隔为 0.3s 的白噪声。听者是来自本实验室 5 位听力正常的自愿者(2 名男性, 3 名女性)。测试时, 听者要求从目标方位中选择一个: 目标的角度范围为水平 360°, 间隔 15°, 其中 0° 表示正前方, 顺时针为正, 逆时针为负, 不包括 ±90°, 而 ±60°, ±75° 等角度则采用线性插值得到。测试时先播放分布在前、后、左、右 4 个方向的声音信号, 并告知实际方向; 然后分别在这 4 个声源的附近位置选择一个方位, 要求听者对声源进行判断; 重复上述过程, 直到完成所有 22 个方位上的一轮定位实验。

文献[8]表明利用 LSR 得到的个性化 HRTF 的定位结果优于非个性化 HRTF。为了比较本文提出的 PLSR 法和文献[8]中 LSR 模型的效果, 本文利用 5 个听者的人体参数, 分别通过这两种模型得到相应的个性化 HRTF, 在相同实验条件下, 对这两套个性化 HRTF 进行了 5 轮上面的主观定位实验。图 6 给出了这 5 轮实验的统计结果, 图中圆圈的大小表示出现在那个坐标上角度的次数, 最大的圆圈表示出现 25

次, 最小的圆圈表示出现 1。其中图 6(a)是基于 LSR 的个性化 HRTF 的声音定位的测听结果, 而图 6(b)是基于本文提出的 PLSR 的个性化 HRTF 的声音定位的测听结果。本文规定, 只有目标方位和测试方位完全一致时才算正确, 而落在两组斜率为 -1 的斜线内及其斜线上的点都判为前后混淆。

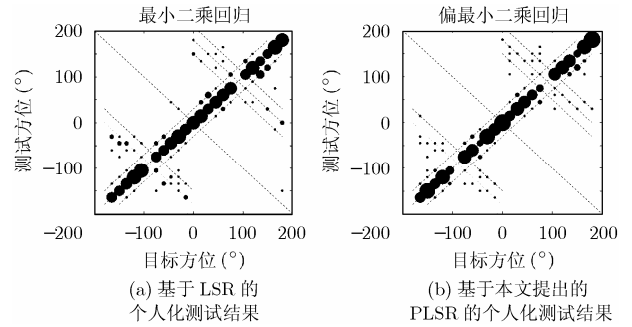


图 6 声音定位测试结果

对 5 轮测听数据进行统计, 结果表明在本文主观测试条件下, 由 LSR 得到的个性化 HRTF 在水平面内声音定位的正确率为 69.7%, 前后混淆率为 11.5%; 而由 PLSR 得到的个性化 HRTF 在水平面内声音定位的正确率为 72%, 前后混淆率为 10.2%。综合参考文献[8]的实验结果, 可以得到以下结论: 由 CIPIC 数据库提供的人体参数通过本文方法估计得到的 HRTF 和 CIPIC 数据库提供的原始测量 HRTF 在感知方面差异不大, 定位的正确率和前后混淆相近; 由听者本人人体参数通过本文方法估计的个性化 HRTF 的声音定位正确率明显高于 CIPIC 数据库中非个性化的 HRTF, 同时前后混淆也有一定的改善; PLSR 得到的个性化 HRTF 的定位正确率和前后混淆比文献[8]中 LSR 得到的个性化 HRTF 有一定的改善。由此可见, 本文提出的基于 PLSR 的个性化 HRTF 建模方法是可行的, 这不仅为 HRTF 的个性化提供了一条捷径, 在完善前一阶段工作的同时也为后一步研究工作奠定了基础。

4 结束语

随着 HRTF 在虚拟听觉空间合成中日益广泛地应用, 寻求一种简单而可行的个性化 HRTF 估计法成为空间听觉研究的一个学术热点。本文通过对 CIPIC 数据库中 HRIR 数据处理和人体参数筛选, 利用 PLSR 分析建立了 HRTF 和人体参数之间的对应关系, 实现了一种快速便捷的个性化 HRTF 估计法, 并用谱失真度和声音定位实验对估计的 HRTF 进行主客观的性能评价。实验结果表明估计的 HRTF 不仅谱失真度小, 而且与测量的 HRTF 在感知方面也没有太大的区别; 同时由测试者人体参数估计得到的个性化 HRTF 定位准确率和前后混淆现象都有所改善。这说明通过本文选定的人体参数来估计特定个人的 HRTF 是可行的, 为以后求个性化的 HRTF 提供了一种简单有效的方法。今后的工作一方面可以考虑采取不同的方法提取 HRTF 的特征和选取合

适的人体参数;另一方面也可以采用其他的非线性回归方法来建立人体参数和 HRTF 特征参数之间的关系,希望得到更好的个性化 HRTF。

参 考 文 献

- [1] Blauert J P. Spatial Hearing. Revised edition. Cambridge, MA: MIT, 1997: 37-176.
- [2] Wenzel E M and Doris J K, *et al.*. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 1993, 94(1): 111-123.
- [3] Wu Z Y, Chen J S, and Richard A R, *et al.*. Measurement and estimation of free field to eardrum transfer function for virtual acoustics space research. *ACTA Acustica*, 1996, 21(1): 645-652.
- [4] Zotkin D N and Hwang J, *et al.*. HRTF personalization using anthropometric measurements. Proc. IEEE WASPAA, New Paltz, NY, USA, 2003: 157-160.
- [5] Middlebrooks J C. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.*, 1999, 106(3): 1493-1510.
- [6] Otani M and Ise S. Numerical calculation of the head-related transfer functions by using the boundary element method. Proc. WESTPRAC' 2000, Kumamoto, Japan, 2000, VII: 305-308.
- [7] Nishino T and Nakai Y, *et al.*. Estimating head related transfer function using multiple regression analysis. IEICE Trans. A, 2001, J84-A: 260-268.
- [8] Hu H M, Zhou L, and Wu Z Y, *et al.*. Head related transfer function personalization based on multiple regression analysis. CIS'06, Guangzhou, China, 2006: 1829-1832.
- [9] Algazi V R and Duda R O, *et al.*. The CIPIC HRTF database. IEEE ASSP WASPAA, 2001: 99-102.
- [10] Kistler D J and Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 1992, 91(3): 1637-1647.
- [11] Wu Z Y and Chan H Y, *et al.*. A time domain binaural model based on spatial feature extraction for the head-related transfer function. *J. Acoust. Soc. Am.*, 1997, 102(4): 2211-2218.
- [12] Wu Z Y and Wang W B. A spatial hearing model based on spatial feature extraction and artificial neural network. *Acat Acustica*, 1999, 24(6): 645-652.
- [13] Wang H W. Partial Least Square Regression——Method and Application. Beijing: National defense Industry Press, 1999: 150-171.

胡红梅: 女, 1977年生, 博士生, 研究方向为音频、听觉信号处理及其在虚拟听觉中的应用。

吴镇扬: 男, 1949年生, 教授, 博士生导师, 主要研究方向为音频、听觉、视觉及智能信号处理等。