第 62 卷　第 5 期
2026 年 9 月

南京大学学报(自然科学)
JOURNAL OF NANJING UNIVERSITY
(NATURAL SCIENCE)

Vol. 62, No. 5
Sept., 2026

DOI:10.13232/j.cnki.jnju.2026.05.003

基于多数据库的HRTF方向与个体聚类分析

吕朋博,刘嘉伟,林志斌*

(近代声学教育部重点实验室,南京大学,南京,210093)

摘　要:头相关传输函数(Head-Related Transfer Function,HRTF)数据库是空间音频渲染和个性化建模的重要数据基础.选取SONICOM,PHRTF-NJU和HUTUBS三个HRTF数据库,在统一特征和相似性定义下比较其方向聚类和个体聚类结构.方向维以等效矩形带宽(Equivalent Rectangular Bandwidth,ERB)平滑后的HRTF幅度谱为特征,采用K-means聚类分析方向频谱模式,个体维则基于双耳全方向平均归一化互相关系数(Mean Normalized Cross-Correlation Coefficient,*MNCC*)来构建受试者间相似性矩阵,并采用平均连接层次聚类来分析受试者间整体相似性.方向聚类结果表明,在提出的特征与评价指标条件下,三个数据库均呈现相近的聚类数变化规律:划分为两个方向簇时可概括出相对稳定的粗粒度结构,划分为三个方向簇时可进一步描述较细的空间分区.个体聚类结果表明,三个数据库均呈现由碎片化向逐步合并的阈值响应规律.总体上,在当前的分析条件下,三个数据库的HRTF在方向维上呈现相近的粗粒度结构,同时保留各数据库自身的数值尺度和局部结构差异.

关键词:头相关传输函数,HRTF数据库,方向聚类,个体聚类,平均归一化互相关系数

中图分类号:O429　　**文献标志码:**A

Directional and subject clustering analysis of head-related transfer functions based on multiple databases

Lü Pengbo, Liu Jiawei, Lin Zhibin*

(Key Laboratory of Modern Acoustics, The Ministry of Education, Nanjing University, Nanjing, 210093, China)

Abstract: Head-related transfer function (HRTF) databases are important data resources for spatial audio rendering and personalized HRTF modeling. This study compared directional and subject-level clustering structures in the SONICOM, PHRTF-NJU, and HUTUBS databases using consistent feature representations and similarity definitions across the three databases. For directional clustering, HRTF magnitude spectra smoothed on the equivalent rectangular bandwidth (ERB) scale were used as features, and k-means clustering was applied to analyze directional spectral patterns. For subject-level clustering, an inter-subject similarity matrix was constructed using the mean normalized cross-correlation coefficient (MNCC) computed over both ears and all directions, followed by average-linkage hierarchical clustering to analyze overall inter-subject similarity. Under the feature set and evaluation metrics used in this study, the three databases showed similar trends as the number of directional clusters varied: a two-cluster solution yielded a relatively stable coarse-grained structure, whereas a three-cluster solution provided a finer spatial partition. Subject-level clustering in all three databases exhibited a common threshold-dependent transition from fragmentation to progressive merging. Overall, HRTFs from the three databases exhibited similar coarse-grained directional structures while retaining database-specific differences in similarity scale and local structure.

Keywords: head-related transfer function, HRTF database, directional clustering, subject clustering, mean normalized cross-correlation coefficient

基金项目:国家自然科学基金(12274221)

收稿日期:2026－07－29

*通信联系人,E-mail:zblin@nju.edu.cn

随着虚拟现实[1]、增强现实[2]、三维游戏、听觉显示和助听设备[3]等应用的发展,高真实感空间音频再现的需求不断增加. 高质量空间音频不仅关系沉浸感,还会影响声源定位、距离感知与交互体验[4]. 对于以耳机为主要输出方式的空间音频系统,双耳渲染是重建空间听觉的重要技术途径,该方法通过重现与声源方向相关的双耳时间差、双耳声级差和频谱线索,使听者形成对声源方位、距离和外部化位置的感知. 头相关传输函数(Head-Related Transfer Function,HRTF)描述声波从给定空间方向传播至人耳过程中的声学传输特性,这一传输过程受到头部、耳廓和躯干等结构的影响,HRTF是双耳空间音频渲染和虚拟听觉显示的关键数据基础[5−6].

HRTF具有明显的个体差异. 不同听者的头部尺寸、耳廓形态和躯干结构等方面存在差别,同一空间方向上的HRTF频谱也会随个体发生变化. 实际系统中若采用与目标听者不匹配的通用HRTF,可能影响声源定位准确性、外部化感知和整体空间感知质量[7−8],因此,个性化HRTF的获取、建模和选择一直是空间音频研究中的重要问题. 然而,直接测量个体HRTF通常需要消声环境、扬声器阵列或旋转测量系统以及入耳麦克风等专业设备,测量流程复杂且成本较高[9],公共HRTF数据库由此成为个性化建模、典型HRTF筛选、跨个体相似性分析和数据驱动方法评价的重要基础.

目前已有多个公开或专用HRTF数据库用于空间音频和个性化建模研究,例如,SONICOM[10],HUTUBS[11]和南京大学PHRTF-NJU数据库[12]等,但它们在受试者数量、测量设备、方向采样网格、坐标定义、采样率、频率范围和后处理流程等方面存在差异,反映了不同数据库的采集条件和应用目标,也增加了多数据库联合分析的复杂性. 近年来,面向多数据库协同使用的数据协调与融合问题受到关注,相关工作的目标是在保留感知相关信息的同时,减弱数据来源差异对后续分析和建模的影响[13]. 从单一数据库得到的结论可能受其测量设置、受试者构成和方向采样方式的影响,直接合并不同数据库又可能引入系统性偏差,因此,在开展数据融合或联合建模前,有必要分析不同数据库中是否存在共同结构,并明确数据库差异的主要表现.

HRTF的频谱特征既随声源方向变化,也因受试者解剖结构的差异而不同,基于这一特点,可分别从空间方向和受试者两个层面分析HRTF的数据结构. 聚类分析无需预设具体声学模型,可根据样本间的相似性描述数据内部的组织关系. 方向聚类可分析不同空间方向的HRTF频谱能否归纳为若干方向簇,并考察这些方向簇在不同个体和数据库中的分布. 个体聚类用于刻画不同受试者整体HRTF之间的相似性,为典型HRTF选择和数据库组织提供依据.

已有研究从不同角度讨论了HRTF聚类及其在空间听觉中的应用. Fahn and Lo[14]针对HRTF数据量和计算复杂度问题,采用聚类方法进行数据压缩和代表性选择. So et al[15]通过聚类构建非个性化HRTF组,并结合听音实验考察其在前后方向声像中的作用. Tame et al[16]利用聚类改善耳机虚拟化中非个性化HRTF的定位和外部化表现. Xie et al[17]基于中国受试者数据开展典型HRTF和聚类分析,验证了利用个体间相似性选择代表性HRTF的可行性. 除了直接对HRTF进行聚类外,Guo et al[18]从耳廓人体测量特征出发进行聚类,并将其用于HRTF个性化. Wang et al[19]将深度神经网络与亲和传播聚类结合,用于HRTF失配条件下的双耳声定位.

上述研究主要关注单一数据库内部的聚类结构、典型HRTF选择或特定应用条件下的聚类策略,不同数据库的方向聚类是否形成相近的空间分区,受试者间相似性是否呈现一致的层次变化规律,仍缺少统一方法下的系统分析. 因此,有必要在相同的特征提取、相似性定义和聚类流程下,同时考察方向聚类结构和个体聚类结构,以区分跨数据库的共同规律和数据库差异.

本文选取SONICOM,PHRTF-NJU和HUTUBS三个HRTF数据库,在统一的预处理和特征表示框架下开展分析. 方向维以单个受试者不同空间方向的HRTF幅度谱为样本,采用K-means[20]聚类,聚类结果通过轮廓系数[21]、簇内频谱相似性、方向标签熵和空间连续性进行评价. 个体维基于双耳、全方向HRTF幅度谱计算*MNCC*

相似性和距离矩阵,并采用层次聚类、阈值扫描和多维尺度分析受试者间的整体结构.在此基础上,比较三个数据库的群体方向结构、示例受试者方向聚类结果和个体层次结构.

和已有的单数据库研究相比,本文的主要工作如下.首先,在统一特征和评价方法下比较三个数据库的方向聚类结构,分析其粗粒度空间分区及局部差异.其次,将群体方向结构与示例受试者结果结合,考察方向聚类在群体和单个受试者层面的对应关系.最后,从个体相似性、切树阈值和簇统计等方面比较三个数据库的受试者组织结构,分析其共同趋势和数值尺度差异.相关结果可为方向分组建模、典型HRTF筛选以及多数据库融合前的结构诊断提供参考.

1　数据库与分析方法

选取SONICOM,PHRTF-NJU和HUTUBS三个HRTF数据库作为分析对象,分别从空间方向维和个体维来考察HRTF聚类结构.方向聚类以同一受试者不同空间方向的HRTF幅度谱为样本,分析方向标签在不同数据库中的空间分布及其跨个体稳定性;个体聚类以不同受试者之间的HRTF相似性为基础,分析受试者在全空间方向上的相似关系和层次聚类结构.前者关注HRTF随声源方向变化形成的空间分区,后者关注不同受试者整体HRTF之间的相似性关系,二者共同描述HRTF数据的方向结构和个体结构.

1.1　HRTF数据库　三个数据库在数据来源和采样设置上存在差异.SONICOM和HUTUBS均为公开HRTF数据库,其中,HUTUBS仅选取实测SOFA文件[22];PHRTF-NJU数据库为南京大学建立的HRTF数据库.三个数据库在受试者数量、方向采样数量和原始采样率上存在差异.为了保证个体相似性计算在统一采样率下进行,SONICOM的头相关冲激响应(Head-Related Impulse Response,HRIR)在个体聚类前重采样至44.1 kHz,PHRTF-NJU和HUTUBS保持原始44.1 kHz数据.为了减小方向采样、受试者构成和原始谱值尺度差异的影响,采用统一的特征提取、相似性计算和聚类方法,比较三个数据库在方向聚类和个体聚类中呈现的结构特征.

方向聚类的K值评价、群体方向图和示例受试者方向图均以左耳HRTF幅度谱为主要分析对象.方向聚类特征均取0～14 kHz,并在500 Hz以上进行ERB[23]平滑,以降低窄带起伏对聚类结果的影响.为了考察双耳之间方向结构的对应关系,另对右耳数据采用相同方法得到方向标签,并将左右耳方向标签一致性作为补充指标.

个体聚类使用双耳和全空间方向信息.对每名受试者,在统一采样率下提取双耳、全方向HRTF幅度谱,并将每个方向上的频谱表示为128维线性频率采样特征.三个数据库的方向采样数量和方向网格存在差异,本文未对空间方向进行统一插值或方向子集对齐,而是在各数据库内部完成方向聚类和个体相似性计算,再比较其聚类结构和指标变化规律.三个数据库的基本信息如表1所示.

表1　HRTF数据库的信息

Table 1　The information of HRTF database

数据库	受试者数	测量点数	测量环境
PHRTF-NJU	60	432	全消声室
HUTUBS	96	440	全消声室
SONICOM	278	793	半消声室

1.2　方向聚类特征与方法　方向聚类以同一受试者的空间方向为样本,分析HRTF频谱能否形成若干具有代表性的方向频谱模式,并考察这些方向频谱模式在不同个体和数据库中的一致性.本文基于平滑后的HRTF幅度谱构建方向特征,并依次进行标准化、降维、K-means聚类和标签对齐.群体方向结构的分析包括K值评价、群体方向图及跨数据库比较.

设第s名受试者、第e只耳、第d个方向的HRIR为$h_{s,e,d}(t)$,其HRTF幅度谱表示为:

$$H_{s,e,d}(f)=\left|\mathcal{F}\{h_{s,e,d}(t)\}\right| \tag{1}$$

其中,$\mathcal{F}\{\cdot\}$表示傅里叶变换.方向聚类中,对每名受试者分别构建方向-频率特征矩阵,以空间方向为样本,以HRTF幅度谱为特征.特征矩阵经标准化和主成分分析降维后,保留累计解释方差达到95%的主成分,并采用K-means对空间方向进行聚类.不同个体的方向标签通过群体模板进行对齐,模板更新迭代次数为五次.

方向聚类的簇数 K 在 2～6 进行扫描，重点比较 $K=2,3,4$ 的评价结果，并将 $K=3$ 作为主要分析簇数．K 的选择综合考虑轮廓系数、簇内 HRTF 相似性、方向标签熵和空间连续性．由于 HRTF 的方向变化具有连续空间属性，过小的 K 可能只对应最粗粒度分区，过大的 K 则可能增加标签碎片化，因此，将 $K=2$ 作为方向主结构参考，将 $K=3$ 作为便于分析空间细分和跨数据库比较的工作参数，不解释为唯一的物理分界．

1.3 方向聚类评价指标 方向聚类从频谱相似性、空间连续性和跨个体稳定性三个方面进行评价．对于两个 HRTF 幅度谱向量 x 和 y，采用归一化互相关系数 NCC 来衡量二者的频谱相似性：

$$NCC(x,y)=\frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2}\sqrt{\sum_i (y_i-\bar{y})^2}} \tag{2}$$

在每一簇内部，计算同簇方向之间的平均 NCC，用于描述该簇内 HRTF 频谱的一致程度．除频谱相似性外，还从聚类结构清晰度、跨个体标签稳定性和空间分布连续性三个角度对方向聚类结果进行评价．

轮廓系数可衡量聚类后的类内紧密性和类间分离度，表征不同方向簇在特征空间中的可分性．方向标签熵基于同一空间方向在不同受试者中的标签分布计算，可描述该方向在群体中的归属稳定性，较低的熵表示同一方向在不同个体中更倾向于被分入同一方向簇．空间连续性根据相邻空间方向的标签关系计算，可衡量聚类标签在空间上的连贯程度，空间连续性较高表示聚类结果更倾向于形成连续区域，而非在相邻方向之间频繁跳变．

由于 K-means 聚类标签本身没有固定语义，不同受试者或不同数据库的标签需要经过对齐后才能进行比较．以群体模板为参照对单个受试者的方向聚类标签进行对齐，并在左右耳一致性分析中采用标签匹配方法消除标签置换对结果的影响．跨个体方向稳定性采用两类指标表征，一类为任意两名受试者之间的方向标签一致比例，另一类为单个受试者与群体多数标签之间的一致比例．左右耳方向一致性分别从群体多数标签和受试者-方向标签两个层面计算，用于补充说明方向聚类结构在双耳之间的对应关系．

1.4 个体相似性与层次聚类 个体聚类以受试者为样本，分析不同受试者整体 HRTF 幅度谱之间的相似性关系．在统一相似性定义下分别建立三个数据库的个体层次结构，并比较其阈值扫描和簇统计结果．

在个体聚类中，采用与 Xie et al[17] 的典型 HRTF 聚类分析相同的 $MNCC$ 相似性定义、距离转换和平均连接层次聚类方法，用于衡量不同受试者整体 HRTF 幅度谱之间的相似性．由于三个数据库的方向采样网格和方向数量不同，本文在各数据库内部独立构建 $MNCC$ 距离矩阵并进行聚类．跨数据库比较依据阈值扫描趋势、簇统计量和相似性分布统计量．设第 p 和第 q 名受试者在第 e 只耳、第 d 个方向上的特征向量分别为 $x_{p,e,d}$ 和 $x_{q,e,d}$，其中，每个特征向量为对应方向上的 128 维线性频率采样幅度谱，则二者的平均归一化互相关系数为：

$$MNCC_{pq}=\frac{1}{2D}\sum_{e\in\{L,R\}}\sum_{d=1}^{D} NCC\left(x_{p,e,d},x_{q,e,d}\right) \tag{3}$$

其中，D 为参与计算的空间方向数．$MNCC$ 越高，表示两名受试者在双耳和全方向上的 HRTF 幅度谱越相似．根据已有研究，进一步将 $MNCC$ 转换为个体距离：

$$D_{pq}=\sqrt{2\left(1-MNCC_{pq}\right)} \tag{4}$$

受试者间距离矩阵采用平均连接层次聚类分析，以两个簇中所有样本对距离的平均值作为簇间距离，可在极端近邻和极端远邻之间取得相对折中，适用于统一相似性定义下的三库个体聚类结构比较．

1.5 切树阈值与补充统计量 层次聚类树通过距离阈值得到具体个体簇划分．将切树阈值记为 T，并在 0.40～0.70 以 0.02 为步长进行扫描，统计簇数、单受试者簇数量及比例、最大簇大小及比例、非单受试者簇数量和簇内平均 $MNCC$，其中，单受试者簇指仅包含一名受试者的簇．阈值选择需避免两种情况，阈值过低导致单受试者簇过多，阈值过高则可能合并差异较大的受试者．

候选阈值根据单受试者簇比例、最大簇规模

和簇内平均*MNCC*的变化综合确定,以保留非单受试者簇并避免最大簇过度占据全部受试者.该设置用于在相同方法框架下比较不同数据库的个体聚类结果.

除切树结果外,还计算数据库内受试者对的平均*MNCC*、*MNCC*中位数、平均个体距离和个体距离中位数.为了降低受试者数量差异的影响,平均个体距离先在单个受试者层面计算,再在数据库层面统计,并通过bootstrap方法[24]估计置信区间及数据库间的差值.这些统计量表征数据库内部个体差异的整体水平,而个体距离和*MNCC*表征当前特征、方向集合和相似性定义下的频谱差异,不直接等同于空间听觉感知差异.

2　HRTF方向聚类结果与跨数据库比较

2.1　*K*值的扫描结果　为了比较不同数据库中方向聚类结构的稳定性,在SONICOM,PHRTF-NJU和HUTUBS三个数据库上对$K=2\sim6$进行扫描.方向聚类均基于左耳HRTF幅度谱特征,并采用相同的标准化、主成分分析和K-means聚类方法.评价指标包括轮廓系数、簇内*NCC*、方向熵和群体空间连续性,其中,轮廓系数和空间连续性主要表征聚类结构的稳定性和空间可解释性,簇内*NCC*表征同一簇内方向频谱的相似性,方向熵描述不同个体之间方向标签的不确定性.

如图1所示,三个数据库在*K*的变化过程中表现出相近趋势.随着*K*从2增大到6,簇内*NCC*整体升高,表明增加簇数可以提高局部频谱模式的一致性.同时,轮廓系数和群体空间连续性整体下降,方向熵整体升高,表明簇数过大时,方向标签不确定性增加,空间分区的连续性降低.因此,*K*值的选择需要同时考虑簇内频谱一致性、空间连续性和跨个体标签稳定性.

表2列出了$K=2,3,4$时的主要评价指标.由表可见,$K=2$均具有较高的轮廓系数和空间连续性,在SONICOM,PHRTF-NJU和HUTUBS上的轮廓系数分别为0.4731,0.4849和0.4927,群体空间连续性分别为0.9590,0.9498和0.9386,方向熵分别为0.1928,0.1974和

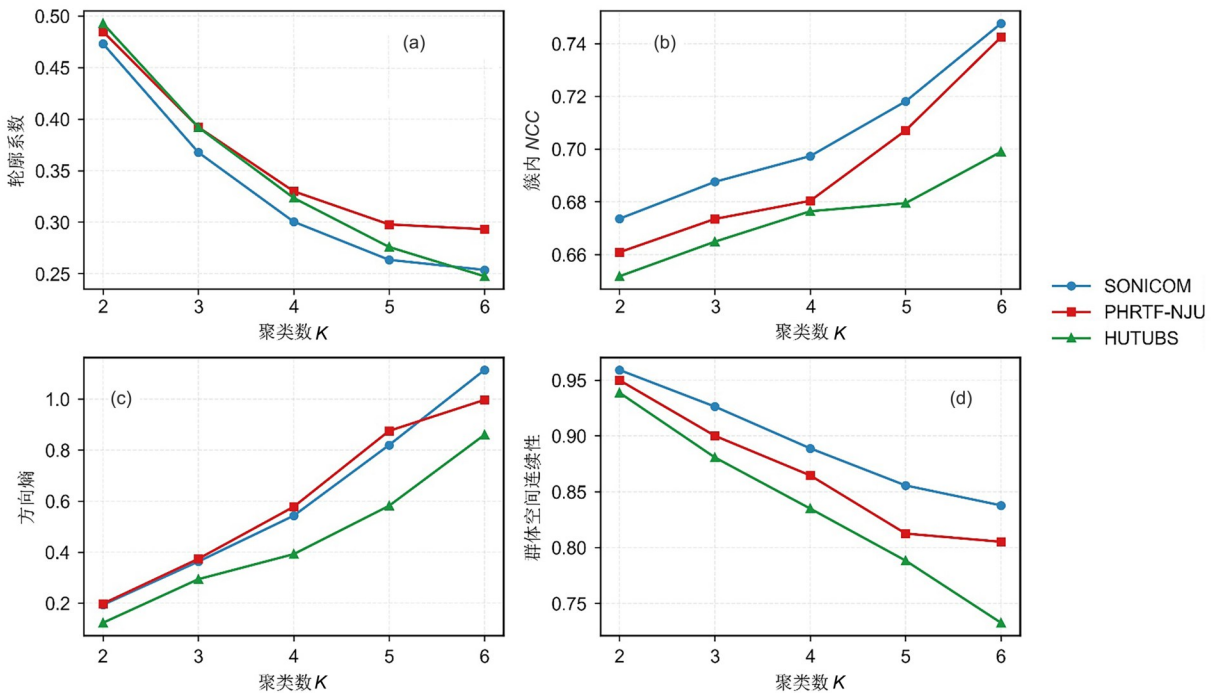


图1　三个数据库方向聚类评价指标:(a)轮廓系数;(b)簇内*NCC*;(c)方向熵;(d)群体空间连续性

Fig. 1　Evaluation metrics of directional clustering across three databases: (a) silhouette coefficient, (b) intra-cluster *NCC*, (c) directional entropy, (d) group-level spatial continuity

表 2 方向聚类 K 值评价结果

Table 2 Evaluation results of K value for directional clustering

数据库	K	轮廓系数	簇内 NCC	方向熵	群体空间连续性
SONICOM	2	0.4731	0.6736	0.1928	0.9590
	3	0.3677	0.6876	0.3626	0.9262
	4	0.3004	0.6974	0.5425	0.8888
PHRTF-NJU	2	0.4849	0.6609	0.1974	0.9498
	3	0.3921	0.6735	0.3732	0.9001
	4	0.3299	0.6804	0.5776	0.8646
HUTUBS	2	0.4927	0.6518	0.1223	0.9386
	3	0.3920	0.6649	0.2934	0.8807
	4	0.3235	0.6764	0.3917	0.8348

0.1223,低于 $K=3$ 和 $K=4$ 时的对应结果. 表明在当前特征和评价指标下,三个数据库的方向结构可首先概括为两个较稳定的粗粒度方向簇. 从 HRTF 的形成机制看,这种粗粒度分区与声源相对于耳侧的位置变化有关,同侧、对侧及中线附近方向在头部遮挡、绕射和耳廓滤波作用下会形成不同的频谱变化趋势.

$K=3$ 时,三个数据库的群体空间连续性仍保持在较高水平,分别为 0.9262, 0.9001 和 0.8807;方向熵分别为 0.3626, 0.3732 和 0.2934,虽高于 $K=2$ 时,但仍低于 $K=4$ 时的对应结果. 与 $K=2$ 相比,$K=3$ 能够在原有粗粒度空间分区基础上进一步区分不同方向簇,特别是将主要分布在中线附近、后方或较高俯仰角方向上的过渡方向簇从两个主要侧向方向簇中分离出来. 同时,与 $K=4$ 及更高的 K 值相比,$K=3$ 没有明显破坏空间连续性,也避免了方向标签在个体间过度分散. 这说明在当前特征和评价指标下,$K=2$ 可以概括方向维的粗粒度主结构,$K=3$ 在结构细分程度和结果稳定性之间呈现较好平衡. 综合各项指标,将 $K=3$ 作为后续空间分析和结构解释的工作簇数,但不表示 HRTF 方向结构只能划分为三个固定物理区域.

2.2 $K=3$ 时群体方向结构与示例受试者结果

在 $K=3$ 时,进一步分析三个数据库的群体方向聚类空间分布,结果如图 2 所示. 群体方向图由个体方向标签对齐后得到,用于表征数据库层面的主要空间分区. 跨数据库比较主要关注各方向簇的空间连续性、主要空间分区范围及过渡方向簇的分布特征.

由图可见,三个数据库的群体方向聚类结果均呈现较明显的空间连续性. 三个方向簇没有在球面方向上随机交错分布,其空间分区主要由两个侧向方向簇和一个过渡方向簇构成. 以 SONICOM 数据库为例,$K=3$ 时的群体方向聚类包含两个覆盖范围不同的侧向方向簇以及主要分布在中线附近、后方或较高俯仰角区域的过渡方向簇. PHRTF-NJU 和 HUTUBS 的群体方向图也呈相近的空间分区形式,即主要方向簇在空间上形成连续区域,局部边界和覆盖范围则随数据库有所变化. 过渡方向簇在三个数据库中均表现为具有一定空间连续性的区域,而不是零散方向

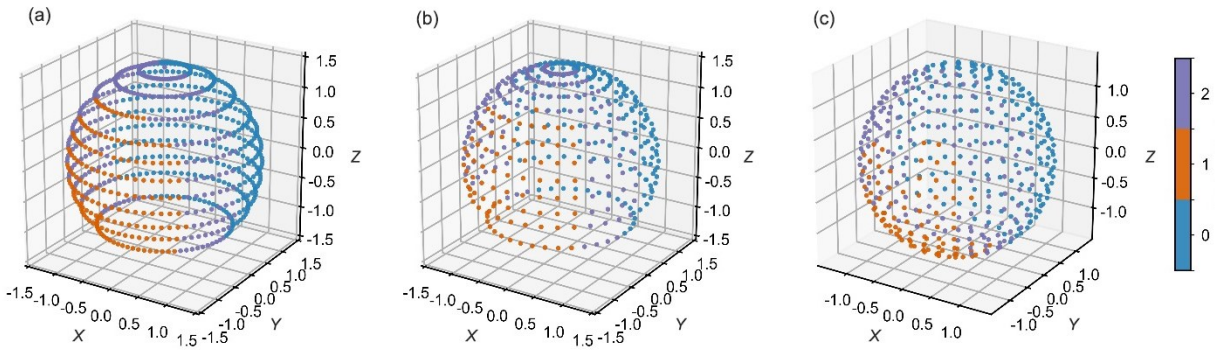


图 2 K=3 时三个数据库群体方向的聚类空间分布:(a) SONICOM;(b) PHRTF-NJU;(c) HUTUBS

Fig. 2 Spatial distributions of group directional clustering with K=3 across three databases: (a) SONICOM, (b) PHRTF-NJU, (c) HUTUBS

点的集合. 根据本文使用的评价指标,$K=3$比$K=2$提供了更细的空间分区. 结合$K=3$时的空间连续性,三个数据库均呈现相近的粗粒度空间分区特征.

本文方向聚类的主要分析结果基于左耳HRTF幅度谱特征. 对于左耳,左侧声源多为同侧方向,声波到达左耳时受头部遮挡较弱,频谱随方向变化相对平缓,因而左侧方向更倾向于归入覆盖范围较大的侧向方向簇. 右侧方向多为对侧方向,声波受到头部遮挡,需绕射后到达左耳,频谱衰减和高频变化更明显. 中线附近、后方或较高俯仰角方向还会受耳廓滤波和躯干反射等因素的影响,频谱变化更复杂,因此相关方向更倾向归入覆盖范围较小的侧向方向簇或过渡方向簇. 后续左右耳一致性结果也体现了方向聚类结构的耳侧依赖性. 因此,图2不仅呈现了HRTF方向谱特征的粗粒度空间分区规律,也证明这种空间分区会受耳侧因素的影响.

除群体方向外,还检查了各数据库中多个受试者的方向聚类结果,以考察单名受试者内部方向聚类的空间分区形态. 所检查的受试者结果整体呈现较连续的空间分区. 从每个数据库中选取一名受试者进行图示,分别记为SONICOM-S1,PHRTF-NJU-S1和HUTUBS-S1,如图3所示.

由图可见,三名受试者的方向标签均呈现较连续的空间分区,其空间分区形态与相应数据库的群体方向图较为接近. 结合对其他受试者结果的检查,单名受试者的方向聚类与群体模板在粗粒度空间分区上具有一定对应关系,该结果与头部、耳廓和躯干对声波传播的影响随声源方向连续变化的特征相一致.

2.3　跨个体方向稳定性与左右耳一致性　为了量化$K=3$时方向聚类的稳定性,统计方向标签一致性指标. 各指标均为标签一致比例,取值范围为0～1,数值越大表示两组方向标签之间的对应程度越高. 表3包括两类指标. 跨个体方向稳定性考察不同受试者是否具有相近的空间分区结果,其中,两两方向标签一致率表示任意两名受试者之间方向标签的一致程度,个体-群体多数标签一致率表示单个受试者结果与数据库群体方向模板之间的接近程度. 左右耳方向一致性考察左耳和右耳空间分区之间的对应关系,其中,左右耳群体标签一致率从数据库群体层面比较双耳方向图,个体-方向左右耳一致率从受试者和方向样本层面统计双耳标签的一致程度.

三个数据库在跨个体方向稳定性上均达到较高水平. SONICOM,PHRTF-NJU和HUTUBS的两两方向标签一致率分别为85.02%,83.80%和86.97%,个体-群体多数标签一致率分别为89.62%,88.96%和90.92%. 表明在本文的标签对齐和一致性指标下,不同个体的方向聚类结果在群体层面呈现一定的共同结构. 结合群体方向图和示例受试者图,三个数据库的粗粒度方向聚类模式在所分析受试者中具有较高一致性.

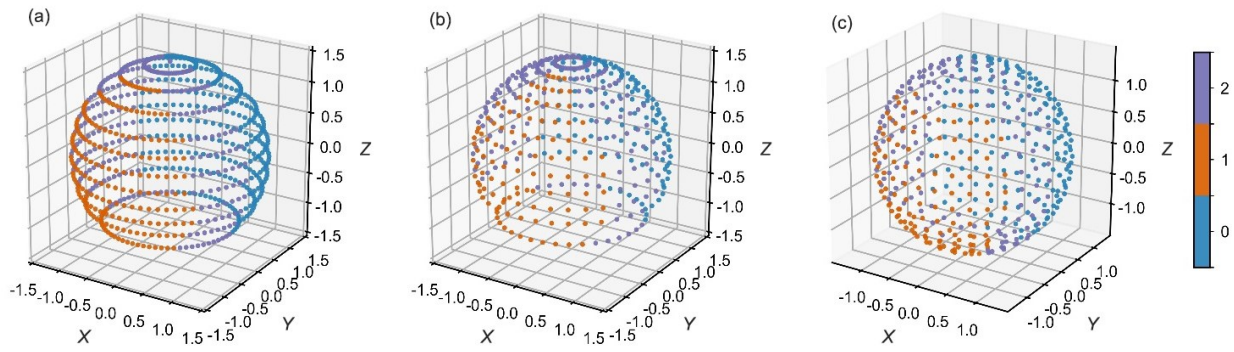


图3　K=3时三个数据库示例受试者方向聚类空间分布:(a) SONICOM;(b) PHRTF-NJU;(c) HUTUBS

Fig. 3　Spatial distributions of directional clustering with K=3 for example subjects across three databases: (a) SONICOM, (b) PHRTF-NJU, (c) HUTUBS

表 3 不同数据库的聚类一致性结果

Table 3 Clustering consistency results across databases

数据库	两两方向标签一致率	个体-群体多数标签一致率	左右耳群体标签一致率	个体-方向左右耳一致率
SONICOM	0.8502	0.8962	0.5763	0.5569
PHRTF-NJU	0.8380	0.8896	0.4815	0.4610
HUTUBS	0.8697	0.9092	0.5341	0.5049

与跨个体方向稳定性相比,左右耳方向一致性较低,SONICOM,PHRTF-NJU 和 HUTUBS 的左右耳群体标签一致率分别为 57.63%,48.15% 和 53.41%,个体-方向左右耳一致率分别为 55.69%,46.10% 和 50.49%,均低于各自的两两方向标签一致率和个体-群体多数标签一致率. 上述差异显示,在同一耳侧内部,不同受试者之间的方向聚类结构具有较高一致性;在左右耳之间,方向标签的对应程度相对降低. 该现象对应左右耳 HRTF 在声学路径上的差异. 对于同一声源方向,左耳和右耳分别对应不同的同侧/对侧关系,并受到头部遮挡、绕射和耳廓滤波差异的共同影响,因此左右耳方向聚类结果会表现出一定的耳侧差异.

3 HRTF 个体聚类结果与跨库一致性

3.1 个体层次聚类结构与阈值扫描结果 在方向维分析的基础上,进一步从受试者维度考察三个数据库中的个体相似性结构. 基于 *MNCC* 距离矩阵和平均连接层次聚类,在 0.40~0.70 以 0.02 为步长进行切树阈值扫描,在候选阈值下统计簇结构,并给出多维尺度分析(Multi-Dimensional Scaling,MDS)二维投影结果,如图 4 所示,其中,单受试者簇指仅包含一名受试者的簇.

图 4 所示的阈值扫描曲线表明,三个数据库具有相似的层次聚类变化规律. 在阈值较低时(0.40),PHRTF-NJU 的 60 名受试者均为单受试者簇,HUTUBS 的单受试者簇比例为 85.42%,SONICOM 的单受试者簇比例为 65.83%,此时切树条件较严格,难以形成稳定的非单受试者簇. 随着阈值的增大,各数据库的单受试者簇比例均下降,最大簇规模逐渐增加. 在阈值较高时(0.70),HUTUBS 的 96 名受试者已合并为一个簇,PHRTF-NJU 的最大簇规模达到 55,SONICOM 的最大簇规模达到 267,表明过高阈值会降低聚类结果的分辨率.

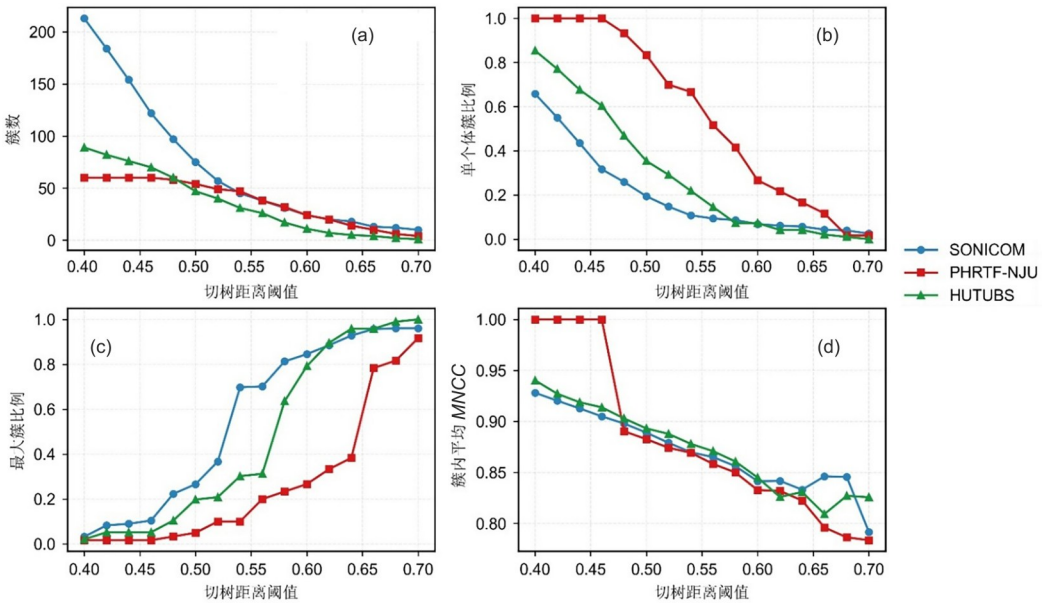


图 4 三个数据库个体聚类阈值扫描曲线:(a)簇数;(b)单受试者簇比例;(c)最大簇比例;(d)簇内平均 *MNCC*

Fig. 4 Threshold scan curves for subject clustering across three databases: (a) number of clusters, (b) singleton-cluster ratio, (c) largest-cluster ratio, (d) average intra-cluster *MNCC*

上述共同趋势显示,三个数据库的个体聚类均呈现从碎片化到逐步合并的层次变化过程.

候选阈值综合考虑单受试者簇比例、最大簇规模和簇内平均*MNCC*确定,以兼顾单受试者簇数量、主簇规模和簇内相似性. SONICOM, PHRTF-NJU和HUTUBS的候选阈值分别为0.50,0.62和0.54,对应的簇统计见表4. 该结果显示,三个数据库在个体维上具有相似的层次变化规律,但具体阈值和簇规模受数据库内部相似性尺度影响.

图5为三个数据库个体间距离矩阵的MDS二维投影结果,图中每个点表示一名受试者,点之间的相对位置由双耳全方向*MNCC*距离经二维投影得到,颜色表示在候选切树阈值下得到的个体簇标签. 由于MDS将高维距离关系投影到二维平面,图5主要用于辅助分析受试者间的相对分布和簇的空间关系.

图5和表4显示,在候选阈值下,三个数据库的受试者没有全部表现为相互独立的单受试者簇,也没有合并为单一大簇. 聚类结果呈主簇、小簇和单受试者簇并存的结构. 部分受试者在距离空间中形成相对集中的群体,其整体HRTF幅度谱较接近;同时,仍有一部分受试者分布在主簇之外,形成小规模簇或单受试者簇,对应数据库内部相对分散的个体样本. 因此,个体聚类结果刻画了受试者相似性由局部小群体向较大群体合并的层次关系.

不同数据库的个体聚类结构也存在差别. SONICOM的受试者数量最多,在候选阈值下形成75个簇,其中,最大簇包含74名受试者,同时仍有54个单受试者簇,表明其内部既存在较大的相似个体群体,也保留较多分散个体. PHRTF-NJU在候选阈值0.62下形成20个簇,最大簇包含20名受试者,单受试者簇数为13;考虑到PHRTF-NJU总受试者数为60,显示其在较高阈值下仍保持一定分散性. HUTUBS在候选阈值0.54下形成31个簇,最大簇包含29名受试者,单受试者簇数为21,簇规模分布介于SONICOM和PHRTF-

表4　个体聚类候选阈值与簇统计

Table 4　Candidate thresholds and cluster statistics for subject clustering

数据库	阈值	个体数	簇数	单受试者簇数	单受试者簇比例	最大簇规模	最大簇比例	簇内平均*MNCC*
SONICOM	0.50	278	75	54	0.1942	74	0.2662	0.8891
PHRTF-NJU	0.62	60	20	13	0.2167	20	0.3333	0.8319
HUTUBS	0.54	96	31	21	0.2188	29	0.3021	0.8780

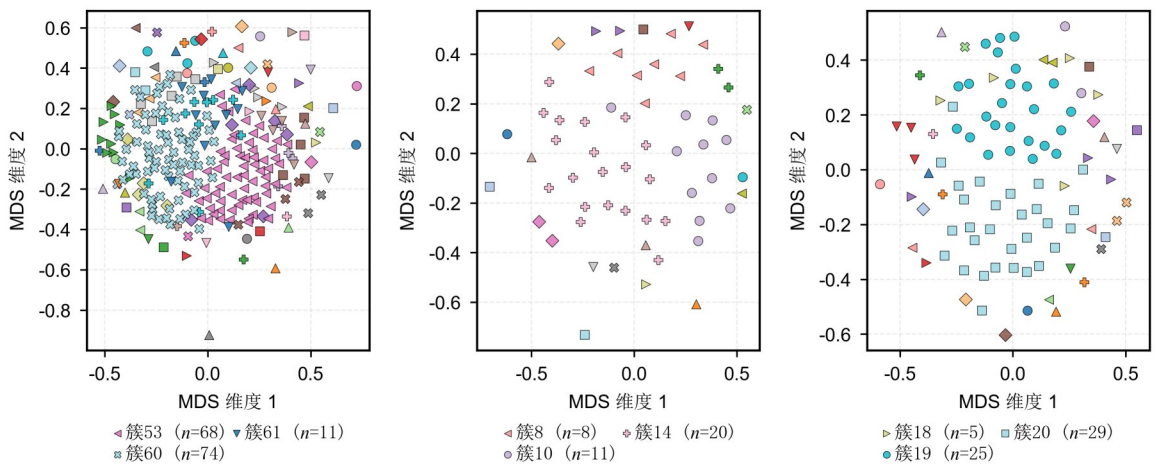


图5　三个数据库个体聚类MDS二维投影:(a) SONICOM(T=0.50);(b) PHRTF-NJU(T=0.62);(c)HUTUBS(T=0.54)

Fig. 5　Two-dimensional MDS projections of subject clustering across three databases: (a) SONICOM (T=0.50), (b) PHRTF-NJU (T=0.62), (c) HUTUBS (T=0.54)

NJU之间. 上述差异表明,三个数据库虽然都能形成由主簇、小簇和单受试者簇共同构成的层次结构,但候选阈值、最大簇规模和单受试者簇比例并不相同,体现出各数据库内部个体相似性尺度存在差别.

3.2 个体相似性指标的补充比较 为了补充候选阈值分析,统计数据库内所有受试者两两组合的平均*MNCC*和平均个体距离. *MNCC*越接近1,表示两名受试者的整体HRTF幅度谱越相似;由*MNCC*转换得到的个体距离越大,表示整体差异越大. 使用bootstrap方法估计平均个体距离的95%置信区间,用于描述统计结果的波动范围.

表5显示,三个数据库的平均*MNCC*均处于较高水平,但数值尺度存在差别. SONICOM和HUTUBS的平均*MNCC*分别为0.8339和0.8257,平均个体距离分别为0.5679和0.5862;PHRTF-NJU的平均*MNCC*为0.7772,平均个体距离为0.6629. 在当前特征和相似性定义下,PHRTF-NJU的平均*MNCC*低于SONICOM和HUTUBS,平均个体距离高于另外两个数据库,该结果与候选阈值分析一致.

表5 个体间相似性补充指标

Table 5 Supplementary metrics for inter-subject similarity

数据库	平均*MNCC*	中位*MNCC*	平均个体距离	中位个体距离	平均个体距离95%CI
SONICOM	0.8339	0.8469	0.5679	0.5534	[0.5609,0.5755]
PHRTF-NJU	0.7772	0.7871	0.6629	0.6525	[0.6522,0.6760]
HUTUBS	0.8257	0.8296	0.5862	0.5837	[0.5790,0.5942]

进一步的bootstrap差异分析显示,PHRTF-NJU与SONICOM的平均个体距离差为0.0951,95%置信区间为[0.0817,0.1085];PHRTF-NJU与HUTUBS的平均个体距离差为0.0768,95%置信区间为[0.0637,0.0910];HUTUBS与SONICOM的平均个体距离差为0.0182,95%置信区间为[0.0077,0.0287]. PHRTF-NJU与SONICOM,HUTUBS的平均个体距离差置信区间均未包含0,表明在本文分析条件下,PHRTF-NJU的平均个体距离高于另外两个数据库.

综上,三个数据库在个体维上均呈现相似的阈值响应规律,但候选阈值、平均*MNCC*和平均个体距离存在差别,显示不同数据库在统一相似性定义下仍保留各自的个体相似性尺度.

4 讨论

4.1 方向聚类结构的共同性与局部差异 *K*值评价结果为方向聚类结构的讨论提供了主要依据. 在$K=2$时,三个数据库均具有较高的轮廓系数和群体空间连续性,方向熵处于较低水平. 上述指标表明,在本文评价指标下,方向标签在空间上较连续,在不同个体之间也具有较高稳定性,因此,$K=2$可作为方向维粗粒度主结构的参考. 在$K=3$时,簇内*NCC*比$K=2$时略有提高,说明细分后簇内方向频谱相似性增强;同时,群体空间连续性仍保持较高水平,方向熵也低于$K=4$时的结果. 上述指标变化表明,该簇数在结构细分和标签稳定性之间呈现较好平衡. 基于上述指标,采用$K=3$进行空间分析和结构解释,但不将其作为HRTF方向结构的唯一物理簇数.

三个数据库在方向维上呈现相近的粗粒度结构,这一现象与HRTF随声源方向变化的基本声学规律相一致. 对于给定耳侧,声源从同侧、前后方向、中线附近到对侧方向变化时,头部遮挡、绕射路径、耳廓方向性滤波和躯干反射等作用也随之连续变化. 相邻方向的HRTF频谱通常具有较高相似性,而同侧、对侧以及中线附近方向之间会形成相对明显的频谱差别. 因此,即使三个数据库在受试者数量、方向采样数量和测量设置上不同,方向聚类仍可在较粗粒度上形成相近的连续空间分区.

图2和图3为上述结构提供了空间分布依据. 三个数据库的方向标签均呈现连续分区,与$K=3$时较高的群体空间连续性一致,簇内*NCC*表明同一方向簇内部具有较高的频谱相似性. 因此,方向聚类结果不仅是空间位置上的分区,也对

应HRTF幅度谱随声源方向变化形成的频谱结构. 同时,左右耳一致性低于跨个体方向稳定性,说明方向聚类结构还受到耳侧因素影响. 不同数据库在方向熵、局部边界和方向簇覆盖范围上存在差别. 由此,方向聚类的共同性主要体现在粗粒度方向结构和整体变化趋势上,具体边界仍可能受到方向采样网格、方向数量、受试者构成、测量系统和预处理流程等因素影响.

4.2　个体聚类结构的一致性与数据库差异　个体聚类刻画的是受试者之间在双耳、全方向HRTF幅度谱上的整体相似性. 阈值扫描曲线显示,三个数据库均呈现相近的层次变化规律. 随着切树阈值增大,单受试者簇比例下降,最大簇比例上升,簇内平均*MNCC*保持在较高水平. 这些指标表明,受试者间相似性在平均连接层次聚类中形成由小簇向较大簇逐步合并的层次关系. 由于该过程基于*MNCC*相似性和由*MNCC*转换得到的个体距离,个体聚类结果主要表征整体HRTF幅度谱的相似性层次结构.

候选阈值下的簇统计和个体相似性补充指标反映了数据库之间的数值尺度差异,三个数据库在候选阈值下均呈现主簇、小簇和单受试者簇并存的结构. 各数据库的候选阈值、最大簇比例、单受试者簇比例和簇内平均*MNCC*存在差异,说明其内部个体相似性尺度不同. 其中,在当前特征和相似性定义下,PHRTF-NJU的平均*MNCC*低于SONICOM和HUTUBS,平均个体距离较高.

PHRTF-NJU与另外两个数据库在平均*MNCC*、平均个体距离和候选阈值上的差异,可能与受试者构成、受试者数量和方向采样设置有关. 测量环境和测量系统等因素也可能产生影响. 由于没有进行受试者数量匹配、方向数量对齐或相同方向子集下的控制比较,所以不能区分各因素的独立贡献,因此,将上述指标差异作为当前分析条件下的现象描述和后续研究线索. 本文不将其直接归因于某一单一因素,也不将其解释为数据库优劣或感知效果差异.

4.3　对HRTF建模、数据库组织和多数据库融合的启示　方向维聚类结果可为HRTF方向分组建模提供结构参考,HRTF的方向频谱模式可以在一定程度上归纳为若干空间区域,因此在空间音频建模中,可按方向簇分别提取频谱特征、选择代表方向或构建分区模型. 与直接对全部方向采用同一模型相比,分区处理有助于分别描述同侧、对侧和过渡方向区域内的频谱变化,可为降低全方向建模复杂度提供参考. 对于HRTF插值、典型方向筛选和数据驱动个性化模型,方向聚类结构也可作为方向先验,用于约束模型在相邻或相近频谱区域内保持更一致的变化规律.

个体维聚类结果可为典型HRTF筛选和匹配式个性化提供依据. 层次聚类和相似性分布能够用于识别相对集中的个体群体、较小规模的个体簇以及单受试者簇,为确定候选HRTF范围提供参考. 在个性化HRTF建模中,不同数据库在个体维上呈现的相近层次变化规律可为跨数据库扩充候选HRTF提供依据,相关结构还可用于建立相似性检索空间和设计匹配式个性化策略.

本研究也可为多数据库融合前的结构诊断提供参考,方向聚类的共同结构和个体聚类的相似性尺度差异有助于在数据合并或联合建模前区分跨数据库共有规律和需要保留的数据库特异性.

4.4　局限性　本文主要从HRTF幅度谱出发分析方向维和个体维的聚类结构,没有纳入相位、双耳时间差和动态头动线索等信息,因此,本文结果主要对应幅度谱特征空间中的组织规律,其与完整空间听觉感知之间的关系仍需要进一步验证. 此外,三个数据库在受试者数量、方向采样数量、方向网格和测量设置上存在差异. 本文没有进行受试者数量匹配、方向网格对齐或相同方向子集下的控制分析,因此,数据库间个体维数值差异的解释力度有限,相关结果可作为现象描述和后续研究线索.

5　结论

本文在统一的特征提取和相似性定义下,分析了SONICOM,PHRTF-NJU和HUTUBS三个数据库的方向聚类和个体聚类结构.

方向聚类结果表明,在本文特征与评价指标条件下,三个数据库的方向聚类结果随聚类数变化呈相近趋势. 划分为两个方向簇时可概括出相对稳定的粗粒度结构;划分为三个方向簇时,在保持较高空间连续性的同时,可进一步分离出过渡

方向簇. 因此,本文将三个方向簇作为空间分析的工作簇数,而不将其视为唯一的物理分界. 群体方向图和示例受试者方向图均呈现较连续的空间分区,三个数据库在主要方向频谱结构上具有相近特征,但局部边界、方向簇覆盖范围和左右耳一致性仍存在差异.

个体聚类结果表明,三个数据库在*MNCC*相似性定义下均形成由小簇向较大簇逐步合并的层次结构,不同数据库的候选阈值、簇规模和个体距离分布存在差异. 当前分析条件下PHRTF-NJU的候选阈值和平均个体距离高于SONICOM与HUTUBS,平均*MNCC*低于另外两个数据库. 由于本文没有进行受试者数量和方向数量对齐后的控制比较,上述差异主要作为数据库使用和跨数据库比较的补充信息.

上述结果可为方向分组建模、典型HRTF筛选、匹配式个性化及多数据库融合前的结构诊断提供参考,后续可在统一方向集合和受试者数量的条件下,结合人体测量参数、听音实验或声源定位模型,进一步检验聚类结构的声学与感知意义.

参考文献

[1] Schissler C, Nicholls A, Mehra R. Efficient HRTF-based spatial audio for area and volumetric sources. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(4): 1356－1366.

[2] Ranjan R, Gan W S. Natural listening over headphones in augmented reality using adaptive filtering techniques. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(11): 1988－2002.

[3] Vickers D, Salorio-Corbetto M, Driver S, et al. Involving children and teenagers with bilateral cochlear implants in the design of the BEARS (both EARS) virtual reality training suite improves personalization. Frontiers in Digital Health, 2021(3): 759723.

[4] Xie B S. Spatial sound: Principles and applications. The 1st Edition. Boca Raton: CRC Press, 2022.

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

[6] Blauert J. Spatial hearing: The psychophysics of human sound localization. Cambridge: MIT Press, 1997.

[7] Wenzel E M, Arruda M, Kistler D J, et al. Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America, 1993, 94(1): 111－123.

[8] Simon L S R, Zacharov N, Katz B F G. Perceptual attributes for the comparison of head-related transfer functions. The Journal of the Acoustical Society of America, 2016, 140(5): 3623－3632.

[9] Li S, Peissig J. Measurement of head-related transfer functions: A review. Applied Sciences, 2020, 10(14): 5014.

[10] Engel I, Daugintis R, Vicente T, et al. The SONICOM HRTF dataset. Journal of the Audio Engineering Society, 2023, 71(5): 241－253.

[11] Brinkmann F, Dinakaran M, Pelzer R, et al. A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses. Journal of the Audio Engineering Society, 2019, 67(9): 705－718.

[12] 刘嘉伟,吕朋博,林志斌. 面向深度学习的个性化HRTF数据库的构建与分析. 南京大学学报(自然科学), 2025, 61(5): 752－762.

[13] Daugintis R, Barumerli R, Geronazzo M, et al. Listener acoustic personalization challenge LAP24: Head-related transfer function dataset harmonization. IEEE Open Journal of Signal Processing, 2025(6): 950－964.

[14] Fahn C S, Lo Y C. On the clustering of head-related transfer functions used for 3-D sound localization. Journal of Information Science and Engineering, 2003, 19: 141－157.

[15] So R H Y, Ngan B, Horner A, et al. Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: Cluster analysis and an experimental study. Ergonomics, 2010, 53(6): 767－781.

[16] Tame R P, Barchiesi D, Klapuri A. Headphone virtualization: Improved localization and exter-nalization of non-individualized HRTFs by cluster analysis//The 133rd Convention of the Audio Engineering Society. San Francisco, CA, USA:

Audio Engineering Society,2012,Convention Paper:8757.

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

[18] Guo Z Y, Lu Y G, Zhou H L, et al. Anthropometric-based clustering of pinnae and its application in personalizing HRTFs. International Journal of Industrial Ergonomics, 2021, 81: 103076.

[19] Wang J, Wang J, Qian K, et al. Binaural sound localization based on deep neural network and affinity propagation clustering in mismatched HRTF condition. EURASIP Journal on Audio Speech and Music Processing, 2020, 2020(1): 4.

[20] MacQueen J B. Some methods of classification and analysis of multivariate observations//Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA, USA: University of California Press, 1967: 281—297.

[21] Rousseeuw P J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 1987, 20: 53—65.

[22] Majdak P, Iwaya Y, Carpentier T, et al. Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions//The 134th Convention of the Audio Engineering Society. Rome, Italy: Audio Engineering Society, 2013: 1—11.

[23] Glasberg B R, Moore B C. Derivation of auditory filter shapes from notched-noise data. Hearing Research, 1990, 47(1/2): 103—138.

[24] Efron B. Bootstrap methods: Another look at the jackknife//Kotz S, Johnson N L. Breakthroughs in Statistics. New York, NY, USA: Springer, 1992: 569—593.

(责任编辑　杨可盛)