# 运用流形学习和内在指标对听觉特征谱进行客观比较

Chen Xu[a]*、Birger Kollmeier[a]和Lena Schell-Majoor[a]

[a]医学物理与*Hearing4all*卓越集群，奥尔登堡大学，德国奥尔登堡 *D-26111*

联系人：Chen Xu

chen.xu@uni-oldenburg.de

医学物理与声学系，第六学院

奥尔登堡大学，26111，德国奥尔登堡

使用流形学习和内在指标对听觉特征谱进行客观比较

将听力受损者划分为不同的听觉特征谱，有助于更好地理解听力损失的成因及后果，并促进基于特征谱的助听器验配。然而，影响听觉画像生成的因素仍未得到充分理解，现有的听觉画像框架也很少经过系统比较。因此，本研究考察了两个关键因素——聚类方法和特征谱数量——对听觉画像生成的影响。此外，研究还系统回顾并比较了八种已确立的听觉特征分析框架，采用内在统计指标和流形学习技术进行评估。评估维度包括内部一致性（即将相似个体归为一组）和簇分离度（即不同组之间有清晰区分）。为确保可比性，所有分析均基于同一个开放获取数据集——扩展版奥尔登堡听力健康记录（OHHR）开展，该数据集包含1,127名参与者（平均年龄 = 67.2岁，标准差 = 12.0）。结果显示，聚类方法和所选特征谱数量均会显著影响生成的听觉特征谱。在纯粹基于听力图的方法中，Bisgaard 听觉画像的聚类性能最佳，而听力测定表型的表现最差。在结合听力图与阈上信息的框架中，Hearing4all 听觉画像表现突出，其特征谱类别数量接近最优（N = 13），且聚类质量较高，这体现在较低的戴维斯–鲍尔丁指数上。

总之，流形学习和内在指标能够系统地比较听觉特征分析框架，并表明 Hearing4all 听觉特征谱是未来研究的一种有前景的方法。

关键词：听觉特征谱；精准听力学；流形学习；内在指标

## 引言

基于纯音平均阈值的传统听力损失分类（例如，世界卫生组织（WHO）对“中度”听力损失的定义；Humes（2019））往往无法体现听力障碍的多样性，导致同一类别中的患者需要完全不同的治疗。为实现更加个性化的诊断和干预，近期研究在听力受损人群中进一步划分了更精细的亚组。本文将采用机器学习方法对这些基于听觉特征谱概念定义的亚组进行基准评估。

听觉特征分析是精准听力学中的一个关键概念（Dreschler 等人，2008年；Van Esch 等人，2013年；Vlaming 等人，2011年；Sanchez Lopez 等人，2018年；Saak 等人，2022年；Cherri 等人，2024），为超越听力图、实现更准确的听力损失诊断和治疗提供了一个颇具前景的解决方案。一方面，它能够捕捉额外信息，例如超阈值缺陷（Kollmeier 1999；Sanchez Lopez 等人，2018年），从而支持更具针对性的诊断。另一方面，它有助于开发基于特征谱的助听设备验配（参见 Sanchez-Lopez 等人，2021年；Kubiak 等人，2022年）。鉴于这些优势，研究听觉特征谱或可为不同诊断测试之间的相互作用，以及它们与康复听力学最相关的听觉功能之间的关系，提供宝贵见解。

听觉特征分析最终旨在支持实际应用，例如改进诊断、个体化治疗、更快速的助听器验配，或针对研究进行目标参与者筛选（Saak 等人，2022年；2025年）。尽管本研究并未直接探讨这些应用，但它们凸显了对听觉特征分析框架进行系统、客观评估的必要性。通过聚焦于内在聚类属性和方法设计选择，本研究为选择适用于不同应用场景的特征分析方法提供了

为选择适用于不同应用场景的特征分析方法奠定基础。

第一类听觉特征谱是完全根据参与者的听力图生成的经典“基于听力图的听觉画像”。在最简单的形式下，根据参与者在 0.5、1、2 和 4 千赫处的纯音平均听阈（PTA4）对其进行分类。PTA4 低于 20 dB HL 的个体被视为听力正常（NH），而高于这一阈值的个体则被归类为听力受损（HI）。本研究将其作为“基线”听觉特征谱。WHO 提供了一套更为详细的基于 PTA4 的分级系统——“正常”“轻度”“中度”“中重度”“重度”和“极重度”（Humes，2019年）。

更复杂的基于听力图的听觉画像会考虑听力图的形状：Bisgaard 等人（2010）通过矢量量化识别出 10 类听力图，也称为标准听力图或 Bisgaard 画像，其中包括七种平坦或中度渐降型听力图（N1 至 N7）和三种陡降型听力图（S1 至 S3）。此外，Cruickshanks 等人（2020）和 Humes 等人（2021）提出的威斯康星州年龄相关听力损伤分类量表（WARHICS）主要依据听力测定阈值，将听力损失分为八个等级，每个等级对应一种不同的听觉特征谱。Dubno 等人（2013）根据动物模型的听力图提出了五种听觉特征谱，称为听力测定表型，包括老年正常型（O-N）、代谢前型（PRE-MET）、代谢型（MET）、感音型（SENS）以及代谢与感觉混合型（MET +SENS）。在此基础上，Parthasarathy 等人（2020）采用数据驱动方法得出四种一般表型：正常听力图、平坦或渐降型听力损失、高频听力损失（HFHL）以及混合型感音神经性听力损失。

第二类听觉特征谱可称为“综合听觉画像”，因为其生成不仅使用听力图，也使用超阈值参数（例如噪声中的语音识别阈值），并采用复杂的数据驱动方法。最早的综合特征谱研究是在“HearCom”项目中开展的（Dreschler 等人，2008；Van Esch 等人，2013；Vlaming 等人，2011）。虽然这些研究未能形成一套可实际应用的听觉特征谱，但为该领域的后续研究奠定了重要基础：Sanchez-Lopez 等人（2018；2020）在丹麦“BEAR”项目中开发了四种不同的听觉特征谱，本文称之为“BEAR 画像”；而 Saak 等人（2022；2025）则在德国“Hearing4all”项目中识别出 13 种不同的特征谱，因此称为“Hearing4all 听觉画像”。

这些听觉特征谱以及前文回顾的特征谱汇总见表 1。该表列出了各框架所包含的特征谱数量。表中还指出，只有世界卫生组织听力损失（HI）等级和WARHICS等级基于流行病学数据集，而其他六种听觉特征分析框架则利用更具体的临床听觉数据集开发。此外，基线、WHO HI 分级、WARHICS等级和听力测定表型均依据专家知识得出（也称为“基于专家的方法”），而其他四种听觉特征分析框架则采用数据驱动方法。

表 1。本研究比较的八种听觉特征分析框架的特征。N = 特征谱数量。n = 生成听觉特征谱所需的患者数量。

| 听觉特征谱 | N | 纳入的参数 | 开发数据集 | 特征分析方法 | 参考文献 |
|---|---|---|---|---|---|

| | | 听力图为基础的 | 全面的 | 流行病学的（n ≥ 4000 患者） | 听力学cal（n ≤ 1000 患者） | 专家为基础 | 数据驱动 | |
|---|---|---|---|---|---|---|---|---|
| 基线 | 2 | × | | | × | × | | - |
| Bisgaard 画像 | 10 | × | | | × | | × | Bisgaard 等人（2010年） |
| 世界卫生组织听力受损分级 | 6 | × | | × | | × | | Humes（2019年） |
| WARHICS 等级 | 8 | × | | × | | × | | Cruickshanks 等人（2020年） |
| 表型 | 5 | × | | | × | × | | Dubno 等人（2013年） |
| 一般表型 | 4 | × | | | × | | × | Parthasarathy 等人（2020年） |
| BEAR | 4 | | × | | × | | × | Sanchez-Lopez 等人（2020年） |
| Hearing4all | 13 | | × | | × | | × | Saak 等人（2022年） |

尽管听觉特征谱在听觉研究中具有诸多优势和重要意义，但仍存在若干局限。首先，影响听觉特征谱生成的因素尚未得到充分探究。Elkhouly 等人（2021）研究了特征谱数量的影响，并指出，在仅包含听力图的数据库中，采用八个或十个特征谱可能最为理想。相比之下，Saak 等人（2022）提出，13

基于对同时包含听力图和超阈值测量的听力学数据库进行模型聚类，所得听觉特征谱达到了最优。此外，Parthasarathy 等人（2020）将高斯混合模型应用于人群数据集，并根据贝叶斯信息准则（BIC）值确定了四个最优特征谱。然而，不同研究得出的最优特征谱数量各不相同，这凸显了进一步研究的必要性。请注意，只有数据驱动方法能够优化特征谱的数量，而专家制定的听觉档案则依赖专家判断来确定这一数量。除特征谱数量外，我们还假设，数据驱动方法中聚类算法的选择也可能影响听觉特征谱的生成，因为不同研究采用了各种算法。例如，Bisgaard 等人（2010）主要使用矢量量化，而 Parthasarathy 等人（2020）采用了高斯混合模型。

其次，据我们所知，目前只有少数研究系统且定量地比较了不同的听觉特征分析框架（Dimitrov 等人，2025年；Xu 等人，2025年）。因此，尚不清楚哪种框架最为合适，因为选择很可能取决于特征分析的具体目标以及可用数据的特点。因此，本文的目标之一是通过基于同一患者数据库，采用统计模型和基于机器学习的模型比较不同的听觉特征谱，弥补这一研究空白，并最终针对特定目的推荐最合适的听觉特征分析框架。本研究的独特之处在于，所有听觉特征分析方法均采用同一个大型、开放获取的数据集——扩展版奥尔登堡听力健康记录（Jafri 等人，2025年）——进行应用和评估。这种统一的框架能够对不同分型策略进行直接、公平的比较，而这是先前研究指出的一项理想特征（Dimitrov 等人，2025年）。

对不同听觉特征分析框架的比较主要考量两个方面：内部一致性，即同一特征谱中的个体聚集程度（即组内方差较低）；以及外部区分度，即不同特征谱中的个体彼此区分的程度（即组间方差较高）。总体而言，我们旨在建立一个客观框架，以公平地比较不同的听觉特征分析框架。具体来说，我们提出采用基于统计学的内在指标（例如 Davies-Bouldin 指数；Davies 与Bouldin，1979年）、主成分分析（PCA；Murphy，2012年）以及流形学习技术（如 t 分布随机邻域嵌入（t-SNE；Van der Maaten 与Hinton，2008年））来进行比较。采用主成分分析和 t-SNE 降低原始数据集的高维度，从而便于在二维空间中直观比较不同群组。这种方法解决了在原始高维数据中对单个参与者进行可视化并比较群组所面临的挑战。在此，降维作为一种分析工具，用于比较听觉特征分析框架，而非作为听觉功能的降维模型。

总之，本研究探讨方法选择如何影响听觉特征分析，以及如何评估由此得到的特征谱。我们旨在探讨以下研究问题，并基于这样的假设：听觉特征谱的构建与评估都会显著影响其临床和科学价值：

1. 不同听觉特征分析框架所生成的特征谱具有多大的一致性和差异性?

2. 特征谱的数量和聚类算法如何影响这些结果?

3. 如何利用统计方法或基于机器学习的方法比较不同的听觉特征分析框架，以评估它们捕捉有意义的个体差异的能力?

这些问题基于这样一个前提：需要更明确、以数据为依据的标准，以提高不同研究中听觉特征谱的可重复性和可比性。

## 材料与方法

### 数据集概述

本研究所采用的数据集是奥尔登堡听力健康记录（OHHR；Jafri 等人，2025年）的扩展版。扩展数据集包含1127名参与者，平均年龄为67.2岁（标准差 = 12.0），其中男性占55.7%，女性占44.3%。所有参与者均接受了包括八项不同测试的综合听觉评估。这些测试包括一份问卷、两项认知评估（言语智力和 DemTect 测试）、SF-12 健康调查、自适应分类响度标度测试、两项噪声下言语测试（哥廷根句子测试（GÖSA）和数字三元组测试（DTT））以及纯音测听。通过这些测试共得出37个参数。听力测定师在符合 IEC 60645-1（2002）标准的隔声室内，使用助听器验配师对气导听力图进行评估。测量涵盖双耳11个听力测定频率，范围为0.25至8千赫，使用 Unity II 听力计和 HDA200 耳机。有关测量的更多细节，请参见 Gieseler 等人（2017）、Saak 等人（2022年、2025年）以及 Jafri 等人（2025年）。本研究展示了如何利用包含开放获取的 OHHR 在内的综合数据集，推导并比较数据驱动的听觉特征。

将听觉特征谱应用于数据集

根据 Bisgaard 等人（2010）、Humes（2019）、Cruickshanks 等人（2020）、Dubno 等人（2013）、Parthasarathy 等人（2020）、Sanchez-Lopez 等人（2020）和 Saak 等人（2022年）提出的分类框架，研究人员依据各框架所规定的参数将参与者分配至相应的听觉特征谱中；这些参数可能包括听力测定阈值，在某些情况下还包括其他超阈值测量。表 2总结了参与者在八种听觉特征分析框架中的分布情况。请注意，在听力测定表型、一般表型和 BEAR 听觉特征中，分别有704名、545名和87名参与者被归类为未确定。

表 2。相应听觉特征分析框架中各类别的参与者人数。有关相应框架的详细信息，请参阅表 1。

| 听觉特征分析框架 | 参与者人数 |
|---|---|
| 基线 | 正常听力：287，听力受损：840 |
| Bisgaard画像 | N1：283，N2：200，N3：253，N4：73，N5：27，N6：7，N7：2，S1：140，S2：88，S3：54 |
| WHO HI 分级 | 正常：256，轻度：312，中度：327，中重度：160，重度：63，极重度：9 |
| WARHICS等级 | 第1级：80，2：55，3：92，4：224，5：232，6：138，7：304，8：2 |
| 表型 | O-N：164，转移前：105，MET：98，敏感度：20，MET + 敏感度：36，未确定：704 |

| | |
|---|---|
| 一般表型 | 正常：119，平坦：36，高频听力损失：240，混合型：187，未确定：545 |
| BEAR | A：431，B：146，C：395，D：68，未确定：87 |
| Hearing4all | 第1级：40，2：18，3：36，4：21，5：66，6：257，7：67，8：20，9：25，10：291，11:190，12: 50，13: 46 |

归一化内在指标

按照 Elkhouly 等人（2021）的做法，采用三种统计内在指标——戴维斯-布尔丁（DB）得分、Calinski-Harabasz（CH）得分和轮廓系数，来评估八种听觉特征分析框架的表现。所有指标均评估内部一致性（即相似个体被归为一组）和群体区分度（即不同群体之间差异明显）（Davies 与 Bouldin，1979年；Caliński 与 Harabasz，1974年；Rousseeuw，1987年；Elkhouly 等人，2021年）。Davies–Bouldin（DB）得分越低，表示特征分析表现越好，最优值为 0；而 Calinski–Harabasz（CH）得分和轮廓系数越高，则表示听觉特征谱越优。轮廓系数的取值范围为 −1 至 1，数值越高表示表现越好。

为估计这三项内在指标的分布，我们从数据集中有放回地抽取 1,000 个样本进行自助抽样，每个样本包含 N = 1,000 个数据点。所有三项内在指标的计算均使用 Python“scikit-learn”软件包完成；内在指标的详细说明见补充材料。

为实现不同听觉特征分析框架之间的公平比较，我们根据特征谱的数量对内部聚类度量进行了归一化。

具体而言，DB分数通过将原始分数除以 $\log_2(N)$ 进行归一化，其中 N 表示特征谱的数量。对于 CH分数和轮廓系数，则通过将各项度量乘以 $\log_2(N)$ 进行归一化。所得数值分别称为归一化DB分数、归一化CH分数和归一化轮廓系数。

影响听觉特征谱生成的因素（聚类方法、特征谱数量）

我们研究了在同一数据集上应用听觉特征分析框架时，可能影响聚类结果的两个因素，即聚类方法和特征谱数量。对于聚类方法，我们比较了两种广泛使用的方法：Bisgaard 画像（Bisgaard et al., 2010）中采用的矢量量化（VQ），以及 Parthasarathy 等人（2020）在通用表型框架中应用的高斯混合模型（GMM）。比较依据为戴维斯-布尔丁（DB）分数。为确保可比性，两种聚类方法均应用于同一数据集，并将特征谱数量固定为 10 个。此选择受方法学限制所驱动：Bisgaard 画像的数量是预先设定的，而只有通用表型框架中的特征谱数量可以通过 GMM 进行调整。我们进行了 t 检验，以评估聚类方法对 DB 分数的影响。为评估特征谱数量的影响，我们采用 GMM 方法，因为其参数可以自由调整以生成听觉特征谱，并将特征谱数量从 2 个变为 15 个。我们绘制了所得 DB 分数随特征谱数量变化的结果，以便进行比较。GMM 模型使用 Python 中的“scikit-learn”包实现。

在扩展版奥尔登堡听力健康记录（***OHHR***）上应用主成分分析（***PCA***）

主成分分析（PCA）是一种无监督机器学习方法，用于降低数据集的维度、在二维 PCA 空间中可视化各个参与者，并直观比较不同的听觉特征谱（Murphy, 2012）。在这些听觉特征分析框架中，使用聚类方法（例如矢量量化、高斯混合模型）根据听者共有的听觉特征对其进行分组。分析使用 R 包“FactoMineR”（Lê et al., 2008）完成，该包采用奇异值分解（SVD）提取前两个主成分（PC1 和 PC2）。我们使用“factoextra”包（Kassambara & Mundt, 2017）生成 PCA 图，其中 PC1 和 PC2 分别显示在 x 轴和 y 轴上。在进行 PCA 之前，对数据进行标准化，使其具有单位方差。最终分析保留了五个维度（N = 5）。

在扩展版奥尔登堡听力健康记录（***OHHR***）上应用***t-***分布随机邻域嵌入（***t-SNE***）

t-分布随机邻域嵌入（t-SNE）是一种无监督流形学习技术，用于降低数据集的维度并可视化高维数据，因此适合直观比较不同的听觉特征谱（Van der Maaten与Hinton，2008年；Van der Maaten，2014）。与保留全局结构的主成分分析（PCA）不同，t-SNE专注于在将数据点从高维空间映射到低维空间时保持其邻域关系。此外，t-SNE是一种非线性降维方法，而PCA是线性的。

使用 R 软件包“Rtsne”开展 t-SNE 分析（Krijthe et al., 2018）。困惑度参数设为 30（Van der Maaten, 2014），theta

参数（用于控制速度与准确性之间的权衡）设为 0.0。未执行初始主成分分析步骤，其他所有超参数均保持默认值（Krijthe et al., 2018）。t-SNE 分析得到的两个输出变量分别记为 t-SNE 维度 1 和 t-SNE 维度 2。

## 结果

聚类方法和特征谱数量的影响

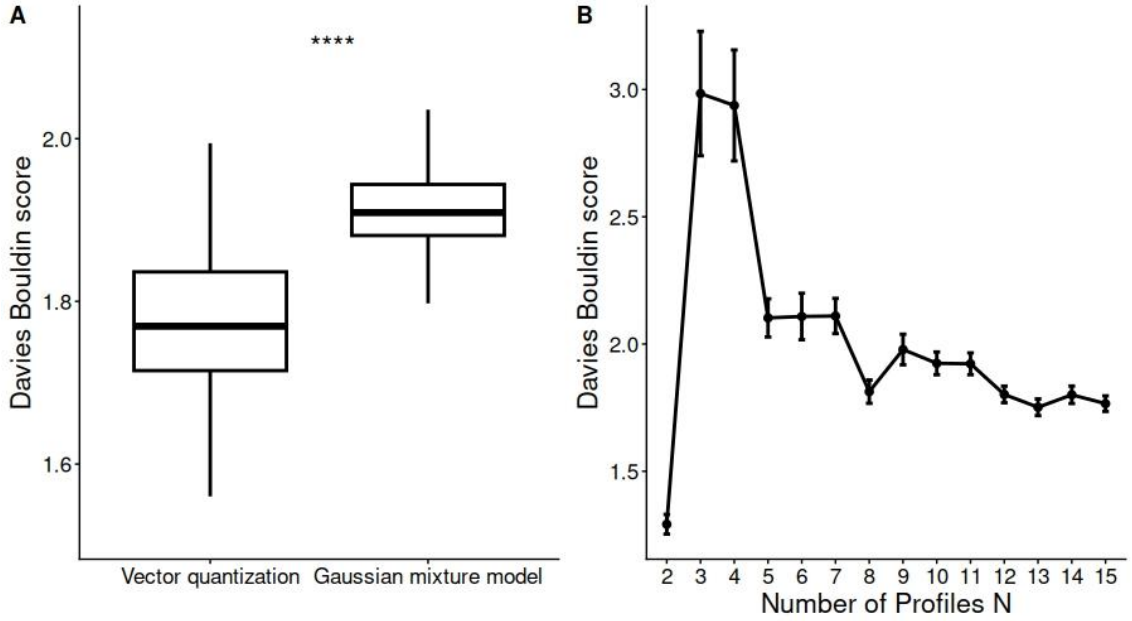



图1 A）两种聚类方法的戴维斯-布尔丁（DB）分数箱线图。B）采用通用的基于高斯混合模型（GMM）的聚类方法时，DB 分数随特征谱数量 N 变化的情况。较低的 DB 分数表示更好的聚类性能，反映出听觉特征谱具有更高的紧凑性和更好的分离度。箱线图表示中位数、第25和第75百分位数以及四分位距（IQR）。须线延伸至距离第25和第75百分位数各1.5 × IQR以内的最极端值。显著性水平以星号表示（* 表示 $p < 0.05$，** 表示 $p < 0.01$，*** 表示 $p < 0.001$，**** 表示 $p < 0.0001$）。

图1A展示了不同聚类方法对聚类性能的影响。对 Bisgaard 画像采用的矢量量化方法（Bisgaard et al., 2010）与一般表型中使用的高斯混合模型

（高斯混合模型）之间的 DB 分数存在显著差异，t 检验结果证实了这一点（p ＜0.05）。在此比较中，特征谱数量固定为 10。图 1B 展示了使用高斯混合模型生成特征谱时，DB 分数随特征谱数量 N 的变化情况。总体而言，随着 N 增加，DB 分数先上升，随后逐渐下降。当 N ＝2 时，DB 分数最低，表明聚类性能最佳。排除 N ＝2 后，最低 DB 分数出现在 N =13，这与 Saak 等人（2022）的研究结果一致。相反，当 N ＝3 时，DB 分数达到峰值，表明聚类性能最差。方差分析结果证实了这一点（p ＜0.05）：图 1B 所示的 DB 分数在 N =2–15 的范围内存在显著差异。

归一化内在指标

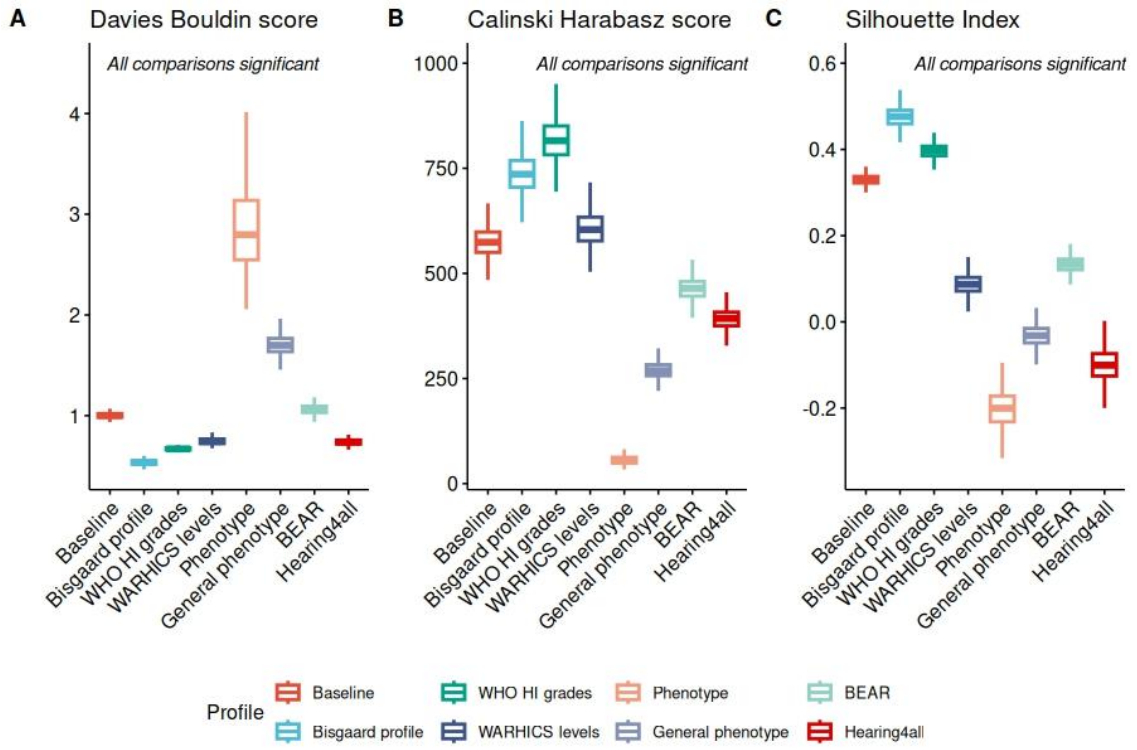



图 2. 八种听觉特征分析框架（即基线、Bisgaard画像、世界卫生组织听力损失等级、WARHICS等级、表型、一般表型、BEAR 和 Hearing4all 听觉特征谱）在三个归一化内在

指标（A. 戴维斯-布尔丁（DB）分数，B. Calinski-Harabasz（CH）分数，以及 C. 轮廓系数）。DB 分数越小、CH 分数越大、轮廓系数越接近 1.0，表明内部一致性（即相似个体被归为一组）和分离度（即不同组之间差异明显）方面的表现越好。箱线图的说明请见图1。

图 2A 展示了不同听觉特征分析框架的归一化DB分数。总体而言，Bisgaard画像的归一化DB分数最低，表现最佳；听力测定表型的归一化DB分数最高，表现最差。Bisgaard画像、WHO 听力损失（HI）等级、WARHICS等级、BEAR 和 Hearing4all 听觉画像的归一化DB分数相近，而一般表型的分数略高。此外，与基线听觉画像相比，Hearing4all 听觉画像的归一化DB分数显著更低。方差分析和事后 t 检验的统计分析证实，所有听觉特征谱之间的归一化DB分数均存在显著差异（$p < 0.05$）。

图 2B 和图 2C 分别展示了归一化卡林斯基–哈拉巴斯（CH）指数和轮廓系数值。WHO HI 分级的归一化CH分数最高，而 Bisgaard画像的归一化轮廓系数值最接近 1.0。这些结果表明，WHO HI 分级和 Bisgaard画像均能有效区分不同组别的参与者，并将相似的参与者聚集在一起。这与图 2A 所示结果一致：WHO HI 分级和 Bisgaard画像的归一化DB分数均相对较低。相反，听力测定表型的归一化CH分数最低，且归一化轮廓系数值离 1.0 最远，表明在所选分型配置下，其聚类分离度相对较低。在其他框架中，基线听觉画像、WARHICS等级和 BEAR 画像

表现相对较好，而一般表型的表现略差。此外，Hearing4all 画像的CH分数和轮廓系数值较低，表明特征谱之间的分离度相对较低。方差分析和事后 t 检验证实，各框架在所有评估指标上的差异均具有统计学意义（$p < 0.05$）。

主成分分析（***PCA***）

图 3 使用主成分分析（PCA）比较了八种听觉特征分析框架。几乎所有听觉特征分析框架都能可靠地识别出位于各子图左侧的听力正常参与者（例如，图 3A 中的 NH 组、图 3B 中的 N1 组、图 3C 中的正常组，以及图 3D 中的第 1 组）。图 4 显示了对第一和第二主成分（ PC1 和 PC2 ）有贡献的因素。总体而言，PC1 和 PC2 分别解释了数据集中 38.2% 和 9.7% 的方差。图 4A 所示的因子载荷表明，对 PC1 贡献最大的六个因素均为听力测定阈值（例如，纯音平均阈值和不同频率下的阈值），这表明 PC1 主要反映阈值水平的听觉敏感度。这与图 3 各子图中听力正常组位于左侧、听力损失最严重的组位于右侧这一观察结果相符。相比之下，图 4B 显示，PC2 的主要贡献因素是源自自适应分类响度标度测试的超阈值参数。因此，PC1 可能代表可听度缺陷，而 PC2 似乎反映了阈上听觉处理的某些方面。

在图 3A、3C 和 3D 中，基线听觉特征谱、世界卫生组织听力损失等级和 WARHICS 等级均呈现出明显的组间分离。然而，在图 3B 中，Bisgaard 画像之间存在显著重叠，尤其是

N 型和 S 型听力图（例如 N1 和 S1）之间，这表明主成分分析可能无法有效捕捉某些特征，例如听力图斜率。

图 3E 和 3F 展示了听力测定表型和一般表型的主成分分析结果。在图 3E 中，大量参与者仍未能归类，且所有组别看起来都高度混杂。这与图 2 中的定量结果一致，听力学表型框架的表现最差。同样，在图 3F 中，许多参与者无法归入任何一般表型，不过其余参与者之间呈现出明显的组间分离。

在图 3G 中，只有少数参与者无法归入四种 BEAR 听觉特征谱之一。然而，我们的结果与 Sanchez-Lopez 等人（2018年；2020年）的报告不同。他们的研究显示出明显的组间分离：A、B、C 和 D 特征谱分别占据主成分分析图中的不同象限，而未确定的参与者则聚集在图中央（详情见 Sanchez-Lopez 等人 2020年发表的图 3）。相比之下，我们的分析显示各组存在重叠。这种差异可能源于两项研究的数据集和检查程序不同。我们的队列规模更大，且可能更多样化，而 Sanchez-Lopez 等人（2020年）可能侧重于更具针对性的原型。此外，Sanchez-Lopez 等人纳入了一些特定测试，例如快速频谱时间调制测试（Bernstein 等人，2016年）和双耳音高处理测试（Santurette 和 Dau，2012年），这些测试并未纳入我们的测试组合。不过，这些特定测试在 Sanchez-Lopez 等人（2020年）的听觉特征谱分类中作用有限，因此，省略这些测试并不能解释此处观察到的差异。图 3H 展示了 13 种 Hearing4all 听觉特征谱。这些特征谱也存在显著的重叠和混杂。

总之，尽管基线听觉画像、世界卫生组织听力损失等级和 WARHICS等级形成的分组相对明确，但听力测定

表型框架似乎一致性较低。不过，需要强调的是，评价分类框架的实用性时，不应只看分离度是否清晰，还应考量由此得到的特征谱所具有的临床价值或解释价值。例如，尽管在听力测定表型框架中观察到大量参与者重叠，表明其判别能力相较于其他特征分析方法较为有限，但它仍可能具有重要的临床价值。具体而言，听力测定表型以听力损失的假设机制为依据，作为描述性框架，用于组织听力测定模式。主成分分析的一项局限在于，将表征限制在两个维度可能不足以清晰区分不同特征谱；因此，未来的分析可纳入更多维度。

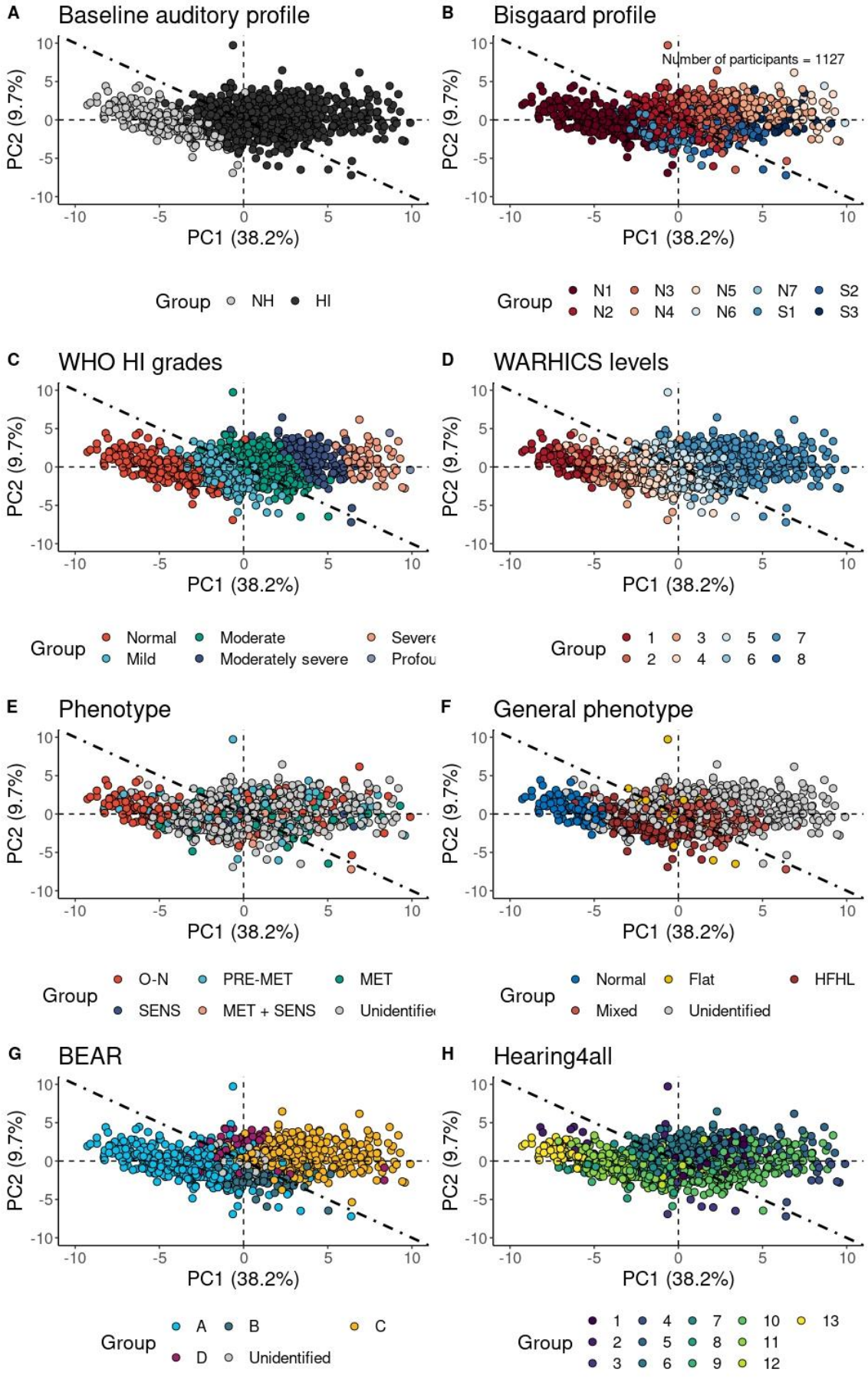

A
Baseline auditory profile
PC2 (9.7%)
PC1 (38.2%)
Group NH HI
B
Bisgaard profile
Number of participants = 1127
PC2 (9.7%)
PC1 (38.2%)
Group N1 N2 N3 N4 N5 N6 N7 S1 S2 S3
C
WHO HI grades
PC2 (9.7%)
PC1 (38.2%)
Group Normal Mild Moderate Moderately severe Severe Profou
D
WARHICS levels
PC2 (9.7%)
PC1 (38.2%)
Group 1 2 3 4 5 6 7 8
E
Phenotype
PC2 (9.7%)
PC1 (38.2%)
Group O-N PRE-MET MET SENS MET + SENS Unidentifie
F
General phenotype
PC2 (9.7%)
PC1 (38.2%)
Group Normal Flat HFHL Mixed Unidentified
G
BEAR
PC2 (9.7%)
PC1 (38.2%)
Group A B C D Unidentified
H
Hearing4all
PC2 (9.7%)
PC1 (38.2%)
Group 1 2 3 4 5 6 7 8 9 10 11 12 13

图 3. 基于 N = 1127 名参与者数据的八种不同听觉特征分析框架的主成分分析（PCA）图。x 轴表示第一主成分（PC1），y 轴表示第二主成分（PC2）。所有面板均采用相同的 PCA。每个面板中的颜色表示根据特定听觉特征分析框架划分的组别。点划线表示对角线。A.) 基线听觉画像（NH：听力正常；HI：听力受损）。B.) Bisgaard 画像（Bisgaard et al., 2010）。C.) 世界卫生组织听力损失（HI）等级（Humes, 2019）。D.) WARHICS 等级（Cruickshanks et al., 2020）。E.) 听力测定表型（Dubno et al., 2013）。F.) 一般表型分类（Parthasarathy 等，2020）。G.) BEAR 听觉特征（Sanchez-Lopez 等人，2020）。H.) Hearing4all 听觉画像（Saak 等人，2022）。

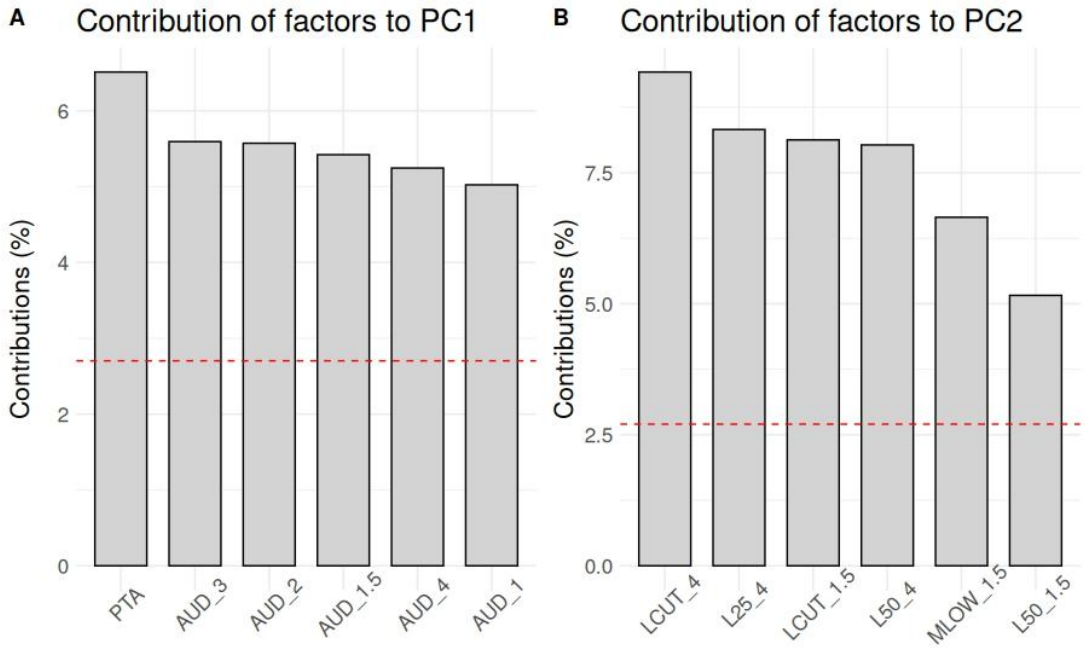



图 4. 各因素对（A）第一主成分（PC1）和（B）第二主成分（PC2）的贡献。仅显示贡献最大的六个因素。红色虚线：平均贡献。子图 A：PTA – 纯音平均听阈；AUD_3/2/1.5/4/1 – 分别为 3、2、1.5、4 和 1 千赫处的听阈。子图 B：LCUT_4/1.5 – 分别为 4 和 1.5 千赫处响度增长函数的转折声级；L25_4 – 4 千赫处的中等响度级（MLL）；L50_4/1.5 – 分别为 4 和 1.5 千赫处的不舒适响度级（UCL）；MLOW_1.5 – 1.5 千赫处的低声级斜率（参见 Oetting 等人。

（2014 年，关于这些源自响度增长函数的超阈值参数的说明）。

***t-*分布随机邻域嵌入（*t-SNE*）**

图5A–5H展示了将t分布随机邻域嵌入（t-SNE）这一非线性流形学习方法应用于八种听觉特征分析框架的结果。x轴和y轴分别表示t-SNE的第1维和第2维。遗憾的是，与主成分分析不同，t-SNE生成的坐标轴没有明确且可解释的含义（Faust等，2017）。不过，与主成分分析类似（参见图3），仍可粗略地将第1维与总体听力损失（顺序相反）联系起来：所有子图中，正常听力类别都位于最右侧，而平均听力损失最高的类别则位于最左侧。对这些听觉特征分析框架进行更详细的比较后发现，它们与主成分分析的结果也高度一致（见图3）。具体而言，基线听觉画像、WHO HI 分级和 WARHICS等级中的各组明显可分，而听力测定表型中的各组则交错更为明显。比较图3和图5可见，t-SNE似乎能更好地分离各组。例如，在BEAR 听觉特征谱中，图5G里的B组和D组与A组和C组清晰分开，而图3G中，B组和D组则与A组和C组混杂在一起。这些结果表明，正如预期，t-SNE在低维空间中分离各组方面优于主成分分析。与基于线性方差投影数据的主成分分析不同，t-SNE会在低维空间中保留数据点之间的关系，尤其是局部邻域距离。因此，在此情境下，t-SNE可能比主成分分析更适合用于降维。不过，将分析限制在二维可能会掩盖高维表示中存在的类别可分性，这表明今后的研究在应用主成分分析和t-

SNE.

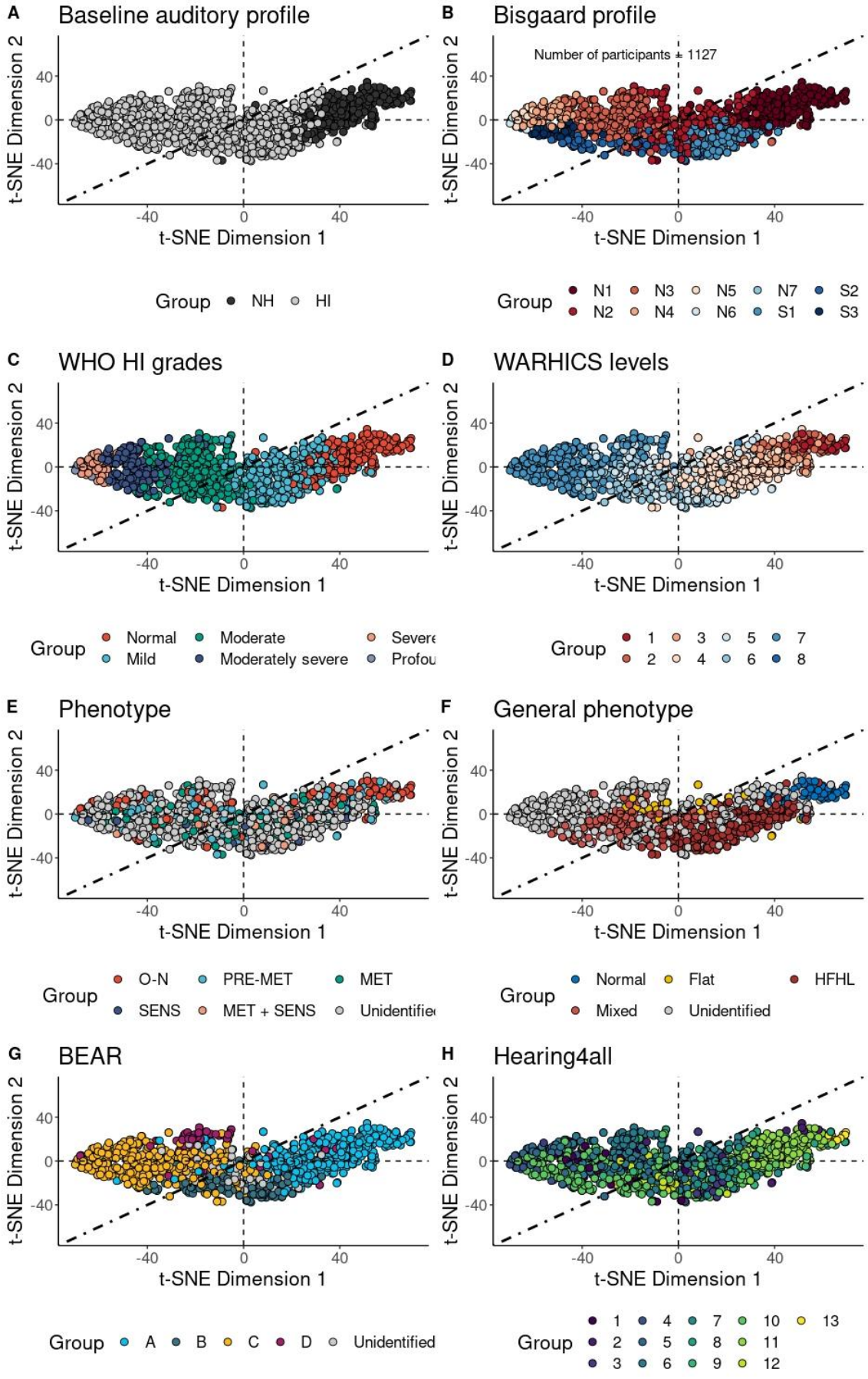

A
Baseline auditory profile
t-SNE Dimension 2
t-SNE Dimension 1
Group NH HI
B
Bisgaard profile
Number of participants = 1127
Group N1 N2 N3 N4 N5 N6 N7 S1 S2 S3
C
WHO HI grades
Group Normal Mild Moderate Moderately severe Severe Profou
D
WARHICS levels
Group 1 2 3 4 5 6 7 8
E
Phenotype
Group O-N SENS PRE-MET MET + SENS MET Unidentifie
F
General phenotype
Group Normal Mixed Flat Unidentified HFHL
G
BEAR
Group A B C D Unidentified
H
Hearing4all
Group 1 2 3 4 5 6 7 8 9 10 11 12 13

图5。N=1127名参与者的t-分布随机邻域嵌入（t-SNE）结果。A.)基线听觉画像（NH：听力正常；HI：听力受损）。B.) Bisgaard画像（Bisgaard 等人，2010年）。C.) WHO 听力损失（HI）等级（Humes，2019年）。D.) WARHICS等级（Cruickshanks 等人，2020年）。E.) 听力测定表型（Dubno 等人，2013年）。F.) 一般表型分类（Parthasarathy 等，2020）。G.) BEAR 听觉特征谱（Sanchez-Lopez 等人，2020年）。H.) Hearing4all 听觉画像（Saak 等人，2022年）。x轴表示t-SNE第1维，y轴表示t-SNE第2维。点划线：反对角线。子图中所示的分组变量与图3中的变量相对应。

## 讨论

### 影响听觉特征谱生成的因素

我们系统地研究了影响听觉特征谱生成的两个主要因素，并采用扩展版奥尔登堡听力健康记录（OHHR）作为共同数据集，从而能够对不同方法进行直接且公平的比较——仅凭基于文献的数值无法实现这一优势。

首先，聚类方法发挥着重要作用，这与我们的预期一致。在特征谱数量相同（N = 10）且使用相同数据集的情况下，与 Parthasarathy 等人（2020）采用的高斯混合模型（GMM）相比，Bisgaard 等人（2010）采用的矢量量化（VQ）方法在听觉特征谱的可分性方面表现更优，证据是其 DB 分数显著更低（见图1A）。VQ 与 GMM 的主要区别在于对聚类重叠的处理方式：VQ 强制采用硬聚类边界，而 GMM 允许聚类重叠（Kinnunen 等人，2011）。这一差异或许可以部分解释本研究中 VQ 的优势，但仍需进一步验证这一发现。此外，VQ 是一种较为简单的非参数方法，而 GMM 是一种具有许多自由参数的参数方法，因此计算量较大。

GMM 还假设数据服从高斯混合分布，但这一假设可能并不适用于我们的情况。未来的研究可以比较这些方法与 Sanchez-Lopez 等人（2020）采用的决策树方法，并评估它们在独立数据集上的表现，这或许会很有启发性。

其次，特征谱的数量会显著影响听觉特征谱的生成。我们的研究结果与 Saak 等人（2022）的研究结果一致，表明 13 个特征谱可获得最佳的聚类性能（见图 1B）。然而，我们不同意 Elkhouly 等人（2021）和 Parthasarathy 等人（2020）的观点；他们分别认为最佳特征谱数量为 10 个和 6 个。这些差异可能源于评估指标的不同：我们使用了 DB 分数，Elkhouly 等人（2021）采用了轮廓系数值，而 Parthasarathy 等人（2020）应用了贝叶斯信息准则（BIC）。此外，各研究采用的聚类方法也有所不同：我们和 Parthasarathy 等人（2020）使用了高斯混合模型，而 Elkhouly 等人（2021）采用了谱聚类（SP）。此外，各研究所依据的数据集也不相同。未来的研究应考察数据集构成和参与者人数等其他因素，因为这些因素可能会影响听觉特征谱的生成。

听觉特征谱之间的比较

本研究采用的内在聚类分数用于量化特征档案结构的互补方面，例如紧密度和分离度，旨在为听觉特征分析框架的设计提供参考，而非决定其设计。分离度更高、内部一致性更强的特征谱有助于减少特征档案分配中的歧义，并支持制定更有针对性的治疗策略，从而实现更精准的诊断。然而，仅凭较高的内在分数并不能保证临床实用性，因为可解释性和实际应用意义仍至关重要。因此，在精准听力学背景下，应谨慎解读这些分数，将其视为权衡区分度与临床

适用性的参考依据。

本研究使用内在指标对八种听觉特征分析框架进行了定量比较，并通过主成分分析（PCA）和t-分布随机邻域嵌入（t-SNE）对其进行可视化分析。这些归一化内在指标从两个关键方面评估听觉特征分析框架：具有相同特征谱的参与者聚集在一起的程度，以及具有不同特征谱的参与者彼此区分的效果。有效的听觉特征分析框架应力求最大化组间方差，同时最小化组内方差，因此这一点至关重要。研究结果表明，这些方法能够有效比较听觉特征分析框架，并确定最优的听觉特征谱。此外，本研究使用同一数据集——扩展版奥尔登堡听力健康记录（OHHR）数据集——通过不同方法进行听觉特征分析，并计算所得特征分析框架的各种内在指标。

正如预期，基线听觉画像在根据纯音平均阈值（PTA4）将参与者分为两组时往往表现良好，这是一种该领域常用的策略。这可能是因为它体现了一种二元分类方案，使得将高度同质的正常组与听力损失程度各异的组区分开来变得相当容易。图 1B 也支持这一发现：当 $N = 2$ 时，戴维斯-布尔丁（DB）得分最低，表明聚类性能最高。然而，尽管具有这一优势，基线听觉特征对于在精准听力学框架下实现更精确的诊断和优化治疗，支持仍然有限。因此，有必要进一步探索听力受损（HI）听者中的亚组，因为他们并非同质群体，尤其是在超阈值测量方面，例如

Xu 等人（2024b）报告的响度缩放数据，而这些数据并未纳入大多数表现最佳的基于阈值的画像框架。

在所有听觉特征分析框架中，听力测定表型区分不同听者群体的能力最弱。这可能归因于画像内差异和画像间差异。一方面，画像内差异很大：具有相同听力测定表型（例如老年正常型表型）的参与者，在主成分分析和 t-SNE 二维空间中分散分布（见图3E和图5E）。另一方面，画像间差异很小。例如，不同表型的示意边界相互重叠（如 Dubno 等人（2013）的图1所示），而不同表型之间也存在共同特征（如 Dubno 等人（2013）的表1所示）。

对于 Bisgaard 画像，所有方法均一致表明，N 型或 S 型听力图的参与者聚类效果相对较好。然而，区分 N 型和 S 型听力图仍然颇具挑战，如图3B和图5B所示。Bisgaard 等人（2010）报告，N1 的 PTA4 值为 10 dB HL，与 S1 的 PTA4 值（12 dB HL）非常接近。同样，N2 的 PTA4 与 S2 相当，N4 的 PTA4 则与 S3 相似。尽管听力图的斜率不同，这些相近的 PTA4 值仍导致画像间差异较小，这可能是难以在降维后的二维空间中区分 N 型和 S 型听力图的原因之一。

世界卫生组织听力损失（HI）等级和 WARHICS 等级在将参与者划分为不同听觉特征谱方面均表现出色。这一点可从其较低的归一化戴维斯–博尔丁（DB）指数、较高的归一化卡林斯基–哈拉巴斯（CH）指数（图 2），以及主成分分析和 t-SNE 图中观察到的聚类模式（图 3C–3D 和图 5C–5D）得到证实。请注意，某些分组看起来

在 Bisgaard 画像、世界卫生组织听力损失等级和 WARHICS 等级之间具有可比性。例如，N3 Bisgaard画像与“中度”世界卫生组织听力损失等级和 WARHICS 6 级相对应。

这些发现表明，不同的听觉特征分析框架之间存在一定程度的对应关系，这一点也体现在表 1 中。例如，世界卫生组织听力损失等级和 WARHICS 都是基于流行病学听力图数据的专家方法，这在一定程度上解释了它们在二维空间中的表现相对较好，其中一个轴主要代表听力测定信息。例如，Humes 等人（2021年）比较了 WARHICS等级与世界卫生组织听力损失等级，发现若干WARHICS等级落在特定的世界卫生组织听力损失等级范围内。他们认为，WARHICS等级比世界卫生组织听力损失等级提供了更细致的分类。然而，我们的结果呈现出更为复杂的情况。WARHICS等级的归一化CH分数和归一化轮廓系数较低，表明在我们的数据中，聚类的区分度或紧密度较低，这可能反映了群组边界重叠或可分性降低。这或许意味着，尽管 WARHICS 框架提供了更多类别，但在我们的样本中，它未必能提升聚类质量。

还可以通过 PTA4（500 Hz、1 kHz、2 kHz 和 4 kHz 的纯音平均听阈）推断世界卫生组织听力损失等级与 Bisgaard 画像之间的关系。Bisgaard 等人（2010年）在其研究的表 1 中确立了 PTA4 与 Bisgaard画像之间的联系，而 Humes 等人（2019年）则在其研究的表 1 中报告了 PTA4 与世界卫生组织听力损失等级之间的对应关系。这使得我们能够通过 PTA4 对 Bisgaard 画像和世界卫生组织听力损失等级进行比较分析，进一步说明这些框架之间相互关联。

与听力测定表型相比，一般表型能更好地将参与者聚类为不同组别，这一点体现在较小的

归一化DB分数、较大的归一化CH分数以及较高的归一化轮廓系数（详见图 2）。如 Parthasarathy 等人（2020）的图3C所示，与听力测定表型相比，一般表型的听力图边界范围更广——尤其是在低于 2 千赫的频率范围内（见 Dubno 等人，2013年的图1）。这一更宽的范围表明，由于分类标准限制较少，一般表型的聚类性能略好。此外，两种表型之间存在明确的对应关系：老年正常型、代谢型、感音型以及代谢型 + 感音型听力测定表型，分别对应正常、平坦、高频听力损失（HFHL）和混合型一般表型。

归一化DB分数相对较低、归一化CH分数较高且归一化轮廓系数较高，这表明 BEAR 听觉特征能够有效地区分参与者。Sanchez-Lopez 等人（2020年）将四种 BEAR 听觉特征与听力测定表型进行了比较，发现特征A和特征B对应感音型听力测定表型，特征D与代谢型听力测定表型相符，而特征C则代表混合型表型（代谢型 + 感音型）。不过，图3和图5中并未明显呈现这种对应关系，这可能是因为只有少数参与者被归入特定的听力测定表型，而大多数参与者仍未确定。

相比之下，在图 2所示的三项内在指标中，有两项显示，13种 Hearing4all 听觉画像的表现显著优于听力测定表型，但总体上不及 BEAR 画像。不过，一个混杂因素是，BEAR 画像中“未确定”案例的数量相对较多（而 Hearing4all 画像中没有未确定案例），这预计会带来更好的聚类性能，因为“难以分类”的案例被排除在外。Hearing4all 画像表现出较大的特征内差异，这一点从某些特征的四分位距较大可以看出，例如配置文件 6中的言语识别阈值；详见 Saak 等人发表于 2022年的图 3。

这些差异可能导致 Hearing4all 画像相较于 BEAR 画像的表现略弱，前提是这种直接比较本身具有有效性。然而，Hearing4all 框架包含更多的特征谱，这可能增强其在更细致的诊断和治疗规划中的适用性，并带来更好的 Davies–Bouldin指数。

我们认同 Saak 等人（2022）所指出的 Hearing4all 听觉画像的诸多优势，例如画像数量灵活、易于扩展、将 Bisgaard 画像纳入画像构建过程，以及其全面性确保所有参与者都能被归入一个衍生画像。不过，也必须承认其中存在一些局限，包括缺乏专家验证以及可解释性较低。尽管 Bisgaard 画像被明确纳入生成 Hearing4all 画像的输入，但 Bisgaard 画像——以及其他听觉特征分析框架——与 Hearing4all 画像之间的对应关系仍不明确。

基于听力图的听觉特征谱与综合听觉画像

鉴于这两种综合听觉画像（即 BEAR 和 Hearing4all 听觉特征谱）具有更多输入参数，能够更全面地刻画个体患者，因此它们在归一化内在指标、主成分分析和 t-SNE 中并未优于基于听力图的听觉画像，这一点令人意外。一个可能的原因是，所用的评估指标可能无法适应高维数据表示，且倾向于偏好较简单的模型。另一个可能的原因是，在不同听觉特征谱中，阈值参数和超阈值参数不一定以一致的方式共同变化，因此综合听觉画像可能无法始终区分

参与者。尽管已知阈值测量和超阈值测量之间至少存在一定程度的相关性，但文献也充分表明，两者所包含的信息可以彼此独立：例如，听力图完全相同的参与者，其超阈值参数可能存在显著差异，例如言语识别阈值（SRT；Balan 等，2023）或响度增长（Oetting 等，2014；Xu 等，2024b）。反之，听力图不同的参与者也可能具有相似的超阈值参数，言语识别阈值（SRT）便是一个例子（Hoppe 等，2020）。

与综合听觉特征谱相比，基于听力图的特征谱更简单，也更容易生成，因为在大多数临床和研究环境中，听力图都是常规可得的，且独立参数较少。相较之下，综合听觉特征分析需要额外的阈上听力测量，而这些测量在许多诊所尚未成为标准项目，并且可能因实验室而异。然而，听觉特征分析的核心动机之一，是对听力损失进行更详细的描述，超越阈值敏感度所能反映的范围。因此，综合特征谱纳入了阈上加工的其他维度，提供了听力图单独无法捕捉的补充信息，这对于更精准的描述和个体化干预至关重要。例如，噪声中的言语识别被认为是一项重要的听觉功能，Hearing4all 听觉特征谱（采用德语哥廷根句子测试）和 BEAR 听觉特征谱（采用丹麦语噪声中听力测试（HINT））均对其进行评估。将这些测试整合到一套通用的综合听觉特征谱中（例如 Saak 等，2025），或可提供一种语言无关的方法，使这些特定语言测试的结果能够纳入同一框架加以利用，尽管测试本身各不相同，或未必适用于同一受试者。然而，这种整合

需要一套标准化测试组合，其中包含听力图和超阈值测试，以全面评估听觉功能（Van Esch 等人，2013年）。此外，标准化数据结构对于促进机构间的数据共享至关重要（Kollmeier 等人，2024），这有助于通过联邦学习利用多样化的大型数据集，构建由大数据支持的听觉特征谱集合（Saak 等，2025）。

流行病学听觉画像与听力学听觉画像

流行病学听觉画像主要包括 WHO 听力损失（HI）等级和 WARHICS等级。这些特征谱基于包含常模参与者的人群数据集构建，而听力学听觉画像则源自临床或听力学数据集（例如奥尔登堡听力健康记录；见 Jafri 等人，2025年）。流行病学特征谱通常涵盖更广泛的参与者年龄范围（包括年轻成人和老年人），且样本量更大，通常超过 10,000 名参与者。相比之下，听力学特征谱聚焦于更具针对性的样本，例如患有听力障碍的个体或年龄在 50–89 岁之间的个体，其样本量也较小，通常约为 1,000 名参与者；对于研究数据集而言，这一规模已相当可观，但与大规模临床数据相比仍较为有限。

两种流行病学听觉画像都能有效区分当前听力学数据集中的参与者，这一点体现为较低的归一化DB分数，以及较高的归一化CH分数和归一化轮廓系数。这些发现与 Humes 等人（2021年）的研究结果一致，并表明这些特征谱不仅适用于人群数据集，也可扩展至临床数据集。相比之下，使用听力学数据集开发的听觉特征谱，其聚类性能差异显著。例如，有些特征谱（如 Bisgaard画像）表现良好，而另一些（如听力测定表型）则表现不佳。这些

差异可能源于本研究所使用的听力学数据集，即 OHHR 的扩展版。因此，使用不同听力学数据集开发的特征谱，如果其病例构成与当前数据集不够匹配，表现可能不尽理想。

内在指标、主成分分析和 ***t-SNE***

迄今为止，我们的研究是首批使用内在指标对不同听觉特征分析框架进行定量比较的研究之一。具体而言，我们采用了三种内在指标：戴维斯-布尔丁（DB）分数、Calinski-Harabasz（CH）分数和轮廓系数，这些指标总体上得出了相一致的结果。

DB分数通过计算特征谱内距离与特征谱间距离的比值来评估聚类质量（见公式 S1）。DB分数越小，表示聚类性能越好，反映出特征谱内距离最小化、特征谱间距离最大化。相反，CH分数通过计算特征谱间方差与特征谱内方差的比值来评估聚类（见公式 S2）。CH 分数越高，表示聚类效果越好，意味着特征谱间方差最大化、特征谱内方差最小化。DB分数和CH 分数都侧重于评估画像层面的聚类性能，而轮廓系数则能提供更全面的视角。它通过取平均值（见公式 S3），同时提供单个数据点层面和整体画像层面的见解。轮廓系数衡量一个数据点与其所属聚类的相似程度，并将其与该点同其他聚类的相似程度进行比较。鉴于这三种内在指标反映了听觉特征分析框架的不同方面，我们建议未来研究采用这些指标，以比单一指标方法更稳健地评估聚类框架。如果不同指标的结果不一致，则应考虑综合使用所有指标。

我们使用主成分分析（PCA）和t-分布随机邻域嵌入（t-SNE）这两种降维技术，对不同的听觉特征分析框架进行了视觉比较。这两种方法都将原始数据集降至二维，使参与者能够在衍生出的二维空间中按不同组别进行可视化，从而便于有效比较。主成分分析（PCA）广泛应用于早期听觉特征谱比较研究（例如，Sanchez-Lopez 等人，2018年；2020年），方法直接且计算效率高。不过，它对离群值尤为敏感。相比之下，t-SNE在该领域的应用较少，它通过保持数据点之间的相对距离来保留数据的局部结构，因此受离群值的影响较小。我们的研究结果表明，与主成分分析（PCA）相比，t-SNE能更好地分离参与者。主成分分析（PCA）因其简洁和快速仍然很有价值，而t-SNE则提供了互补视角，尤其适用于可视化复杂或具有局部结构的数据。综合来看，t-SNE应被视为一种有效的听觉特征谱视觉比较工具。

局限性与展望

目前，不同听觉特征分析框架之间的比较，基于对一个开放获取的临床数据库——奥尔登堡听力健康记录的扩展。然而，该数据集主要包括患有听力障碍的老年患者。尽管从该数据集得出的特征谱能够提供有价值的见解，但其普适性可能受到限制。为确保研究结果稳健且无偏，有必要使用其他数据集（包括基于人群的数据集）对这些特征谱进行验证。美国国家健康与营养调查（NHANES）是一个合适的选择，已被其他研究广泛使用（Ellis 等人，2021年；Parthasarathy 等，2020；Humes 等人，2023a年、2023b年、2023c年）。NHANES 数据集包括超过 10,000 名参与者，涵盖更广泛的听力能力范围。其

测试组合包括人口统计学问卷和听力测定评估，但缺少许多超阈值听觉评估，例如噪声中语音测试。

其次，不同听觉特征分析框架之间的对应关系仍不明确。尽管已有一些研究探讨了这一问题，例如 Sanchez-Lopez 等人（2020）探索了听力测定表型与 BEAR 听觉特征谱之间的关系，Saak 等人（2022）则考察了 Bisgaard画像与 Hearing4all 画像之间的联系，但若能建立一个统一的对应关系理论框架，将极具价值。这样的框架有助于更好地理解不同听觉特征分析框架之间的关系，并实现它们之间的相互转换或估计。

第三，我们的研究表明，两个因素——特征谱数量和聚类方法——会显著影响听觉特征谱的生成。未来研究可以考察其他因素，例如参与者数量及潜在的交互作用。样本量增加或许能识别出更多特征谱，从而影响特征谱的生成。此外，纳入范围更广的听觉测试特征，也可能进一步影响听觉特征谱的开发。

第四，缺乏用于验证听觉特征谱（尤其是数据驱动画像）的客观验证工具。尽管本研究初步采用三项内在指标，对所考察的听觉特征分析框架进行了基准评估，但哪些数据驱动画像属于冗余或无关内容、因而应予以排除，目前仍不明确。未来研究应着重探索合并相似画像的方法（Saak 等，2025），并找出最相关且可解释的听觉画像，以确保框架精简且具有临床意义。

## 结论

所研究的两个因素均显著影响听觉特征谱的生成：聚类方法和特征谱数量。具体而言，对于包含 10 个特征谱的数据集，矢量量化的表现优于高斯混合模型，得到的 DB 分数更低。此外，随着特征谱数量 (N) 从 2 增至 15，高斯混合模型的聚类性能先下降，随后逐渐提升。当 $N \approx 13$ 时，表现最佳；而 $N = 3$ 时，结果最差。

研究比较了八种听觉特征分析框架，包括基线听觉画像、Bisgaard画像、WHO 听力损失（HI）等级、WARHICS等级、听力测定表型、一般表型、BEAR 听觉特征谱和 Hearing4all 听觉特征谱。其中，Bisgaard 听觉画像在将参与者区分为不同组别方面表现最佳，而听力测定表型表现最差。不过，应结合具体分析设置来解读这些相对差异；差异也可能因预期应用和背景而异。其余六种框架的聚类性能相近。在所考察的两种综合听觉画像中，BEAR 听觉特征谱在三项内在指标中的两项上优于 Hearing4All 画像，但代价是必须排除相当比例的数据集，且其仅提供四种不同类别，这限制了其实用性。
Hearing4All 听觉特征谱在采用最佳类别数（$N = 13$）时，在综合听觉画像中取得了最低的 Davies–Bouldin指数，因此应纳入未来研究的考量。

PCA 和 t-SNE 可视化均将参与者呈现在二维空间中，便于直观地比较各组。值得注意的是，正如预期，t-SNE 在区分参与者聚类方面略优于 PCA。

总之，本研究表明，利用流形学习技术和内在指标，可以有效且高效地比较听觉特征分析框架。本研究首次使用相同的底层数据对这些框架进行定量比较，为听觉画像分类和聚类方法提供了宝贵见解。

**致谢**

本研究由德国研究联合会（DFG，德国研究基金会）通过德国卓越战略资助——EXC 2177/1，项目编号 390895286。

**利益冲突声明**

作者声明不存在任何潜在利益冲突。

**数据可用性**

支持本研究发现的数据已在 Zenodo 上公开，网址为 http://doi.org/10.5281/zenodo.14177903。

**参考文献**

Balan, J. R., Rodrigo, H., Saxena, U., & Mishra, S. K.（2023年）。可解释机器学习揭示了听阈与听力图正常的听者噪声中语音识别之间的关系。《美国声学学会杂志》，154(4)，2278-2288。

Bisgaard, N., Vlaming, M. S., & Dahlquist, M.（2010年）。IEC 60118-15 测量程序的标准听力图。《Trends in Amplification》，14(2)，113-120。Caliński, T., & Harabasz, J.（1974年）。一种用于聚类分析的树突法。《Communications in Statistics-theory and Methods》，3(1)，1-27。Cherri, D., Eddins, D. A., & Ozmeral, E. J.（2024年）。迈向精准听力学：听觉感知与认知能力中的个体差异和特征谱。《听力趋势》，28，23312165241263485。Cruickshanks, K. J., Nondahl, D. M., Fischer, M. E., Schubert, C. R., & Tweed, T. S.（2020年）。一种用于老龄化流行病学研究中听力损失分类的新方法：威斯康星州年龄相关听力损伤分类量表。《American Journal of Audiology》，29(1)，59-67。Davies, D. L., & Bouldin, D. W.（1979年）。一种聚类分离度指标。《IEEE Transactions on Pattern Analysis and Machine Intelligence》，(2)，224-227。Dimitrov, L., Barrett, L., Chaudhry, A., Muzaffar, J., Lilaonitkul, W., & Mehta, N.（2025年）。揭示感音神经性听力损失中的表型：无监督机器学习方法的系统综述。《Ear and Hearing》，46(6)，1401-1411。Dreschler, W. A., Esch Van, T. E., Larsby, B., Hallgren, M., Lutman, M. E., Lyzenga, J., ... & Kollmeier, B.（2008年）。通过“听觉特征谱”刻画个体耳。《Journal of the Acoustical Society of America》，123(5)，3714。Dubno, J. R., Eckert, M. A., Lee, F. S., Matthews, L. J., & Schmiedt, R. A.（2013年）。利用动物模型对年龄相关性听力损失的人类听力测定表型进行分类。《Journal of the Association for Research in Otolaryngology》，14，687-701。Elkhouly, A., Andrew, A. M., Rahim, H. A., Abdulaziz, N., Abdulmalek, M., Mohd Yasin, M. N., ... & Siddique, S.（2021年）。一种面向助听器滤波器组设计与初始配置的纯音听力图新型无监督谱聚类方法。《Applied Sciences》，12(1)，298。Ellis, G. M., & Souza, P. E.（2021年）。利用机器学习和美国国家健康与营养调查对听力损失个体进行分类。《Frontiers in Digital Health》，3，723533。Faust, R., Glickenstein, D., & Scheidegger, C.（2018年）。DimReader：解释非线性投影的坐标轴线。《IEEE Transactions on Visualization and Computer Graphics》，25(1)，481-490。

Gieseler, A., Tahden, M. A., Thiel, C. M., Wagener, K. C., Meis, M., & Colonius, H. (2017). 助听器使用情况下无助听设备噪声中语音识别的听觉贡献与非听觉贡献。《心理学前沿》, 8, 219. Hochmuth, S., Koifman, S., Warzybok-Oetjen, A., Avan, P., Kollmeier, B., & Radeloff, A. (2024). PREciSion 老年性听力损失听力学项目（PRESAGE）：改善早发性年龄相关性听力损失的诊断。《喉科-鼻科-耳科》, 103(S 02). Hoppe, U., Hocke, T., &Iro, H. (2022). 语音感知的年龄相关性衰退。《衰老神经科学前沿》, 14, 891202. Humes, L. E. (2019). 世界卫生组织听力损失分级系统：对年龄相关性听力损失情况下无助听设备交流的评估。《国际听力学杂志》, 58(1), 12-20. Humes, L. E. (2021). 威斯康星年龄相关性听力损失分类系统的进一步评估与应用。《American Journal of Audiology》, 30(2), 359-375. Humes, L. E. (2023a年). 未经筛查的美国成年人的听阈：来自美国国家健康与营养调查2011年–2012年、2015年–2016年和2017年–2020年的数据。《听力趋势》, 27, 23312165231162727. Humes, L. E. (2023b年). 美国成年人听力损失、听力困难和助听器使用的人口数据：美国国家健康与营养调查，2011年–12年、2015年–16年和2017年–20年。《听力趋势》, 27, 23312165231160978. Humes, L. E. (2023c年). 美国成年人自我报告的听力困难和助听器使用的人口数据：美国国家健康访谈调查，2007年–2018年。《听力趋势》, 27, 23312165231160967. IEC 60645-1, 2002. 电声学—听力测量设备—第1部分：纯音测听设备。国际电工委员会标准，日内瓦，瑞士。Jafri, S., Berg, D., Buhl, M., Vormann, M., Saak, S., Wagener, K. C., Thiel, C., Hildebrandt, A., & Kollmeier, B. (2025年). OHHR – 奥尔登堡听力健康记录 [数据集] (1.3.0.0) [数据集]. Zenodo. https://doi.org/10.5281/zenodo.16919812 Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A. P., Brand, M. A., & Bullock, R. (2004). DemTect：一种用于支持轻度认知障碍和早期痴呆诊断的新型敏感认知筛查测验。

《国际老年精神病学杂志》, 19(2), 136-143. Kassambara, A. (2020). ggpubr：“基于 ggplot2 的可用于发表的绘图”。R 包版本 0.4. 0, 438. Kassambara, A. (2021). rstatix：用于基础统计检验的管道友好型框架。R 包版本 0.7. 0. Kassambara, A., &Mundt, F. (2017). 软件包“factoextra”。提取并可视化多变量数据分析的结果，76(2). Kinnunen, T., Sidoroff, I., Tuononen, M., & Fränti, P. (2011). 聚类比较

g

方法：文本无关说话人建模案例研究。《模式识别快报》，32(13), 1604-1617。Kollmeier, B.（1999）。感音神经性听力损失涉及的四个因素。载于《心理物理学、生理学与听觉模型》，世界科学出版社（第211-218页）。Kollmeier, B.（2024）。听力学中开放式电子健康记录数据格式的标准化。载于第36届世界听力学大会（WCA），法国巴黎。Kollmeier, B., & Wesselkamp, M.（1997）。用于客观和主观语音可懂度评估的德语句子测试的开发与评估。《美国声学学会杂志》，102(4)，2412-2421。Krijthe, J., van der Maaten, L., & Krijthe, M. J.（2018）。软件包“Rtsne”。R 包版本 0.13。Kubiak, A. M., Rennies, J., Ewert, S. D., & Kollmeier, B.（2022）。复杂聆听条件下听力能力与语音感知首选播放设置之间的关系。《国际听力学杂志》，61(11)，965-974。Lê, S., Josse, J., & Husson, F.（2008）。FactoMineR：用于多变量分析的 R 包。《统计软件杂志》，25，1-18。Murphy, K. P.（2012）。机器学习：概率视角。麻省理工学院出版社。Oetting, D., Brand, T., & Ewert, S. D.（2014）。针对分类响度标度数据优化的响度函数估计。《听力研究》，316，16-27。Parthasarathy, A., Romero Pinto, S., Lewis, R. M., Goedicke, W., & Polley, D. B.（2020）。大型临床队列中听力测定表型的数据驱动分段。《科学报告》，10(1)，6704。

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É.（2011）。Scikit-learn：Python 中的机器学习。《机器学习研究杂志》，12，2825-2830。Rousseeuw, P. J.（1987）。轮廓系数：用于解释和验证的图形辅助工具

n

聚类分析。《计算与应用数学杂志》，20，53-6

5.

Saak, S., Huelsmeier, D., Kollmeier, B., & Buhl, M.（2022年）。一种用于推导听觉特征谱的灵活数据驱动听力学患者分层方法。《神经病学前沿》，13，959582。Saak, S., Oetting, D., Kollmeier, B., & Buhl, M.（2025年）。通过联邦合并整合听力学数据集中的听觉特征谱。《听力趋势》，29，23312165251349617。Sanchez Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., & Dau, T.（2018年）。用于听觉特征分析及个体听力损失表征的数据驱动方法。《听力趋势》，22，2331216518807400。Sanchez-Lopez, R., Fereczkowski, M., Neher, T., Santurette, S., & Dau, T.（2020年）。面向精准听力学的稳健数据驱动听觉特征分析。《听力趋势》，24，2331216520973539。Sanchez-Lopez, R., Fereczkowski, M., Santurette, S., Dau, T., & Neher, T.（2021年）。迈向基于听觉特征谱的助听器验配：验配依据与初步评估。《听力学研究》，11(1)，10-21。Smits, C., Kapteyn, T. S., & Houtgast, T.（2004）。通过电话开发并验证自动噪声中语音筛查测试。《国际听力学杂志》，43(1)，15-28。Vaden, K. I., Matthews, L.J., Eckert, M. A., & Dubno, J. R.（2017）。年龄相关性听力损失听力测定表型的纵向变化。《Journal of the Association for Research in Otolaryngology》，18，371-385。Van Der Maaten, L.（2014）。使用基于树的算法加速 t-SNE。《机器学习研究期刊》，15(1)，3221-3245。Van der Maaten, L., & Hinton, G.（2008）。使用 t-SNE 可视化数据。《机器学习研究期刊》，9(11)。Van Esch, T. E., Kollmeier, B., Vormann, M., Lyzenga, J., Houtgast, T., Hällgren, M.,

...

&Dreschler, W. A.（2013年）。在一项国际多中心研究中评估初步听觉概况测试电池。《国际听力学杂志》

,

52(5), 305-321。

Vlaming, M. S., Kollmeier, B., Dreschler, W. A., Martin, R., Wouters, J., Grover, B., ... & Houtgast, T. (2011年). HearCom：通信社会中的听力。《声学学报与声学》, 97(2), 175-192. Xu, C. (2025年). 移动设备虚拟听力诊所的关键要素——心理物理学、诊断参数估计与验证（博士论文，奥尔登堡大学出版社（UOLP））。Xu, C., Hülsmeier, D., Buhl, M., & Kollmeier, B. (2024a年). 注意力不集中如何影响通过智能手机进行心理声学评估时自适应程序的稳健性和效率?《听力趋势》, 28, 23312165241288051. Xu, C., Schell-Majoor, L., & Kollmeier, B. (2024b年). 开发并验证用于评估纯音阈值和响度感知的基于智能手机的无监督方法。《国际听力学杂志》, 1-11. Xu, C., Schell-Majoor, L., & Kollmeier, B. (2024c年). 在典型家庭环境中通过智能手机高效评估阈值和响度的可行性。medRxiv, 2024年11月. Xu, C., Schell-Majoor, L., & Kollmeier, B. (2025年). 优化精准听力学的听觉特征分析：聚类框架的比较研究。国际听觉与听力学研究研讨会 ISAAR，尼堡，丹麦。Zenodo。https://doi.org/10.5281/zenodo.16955643