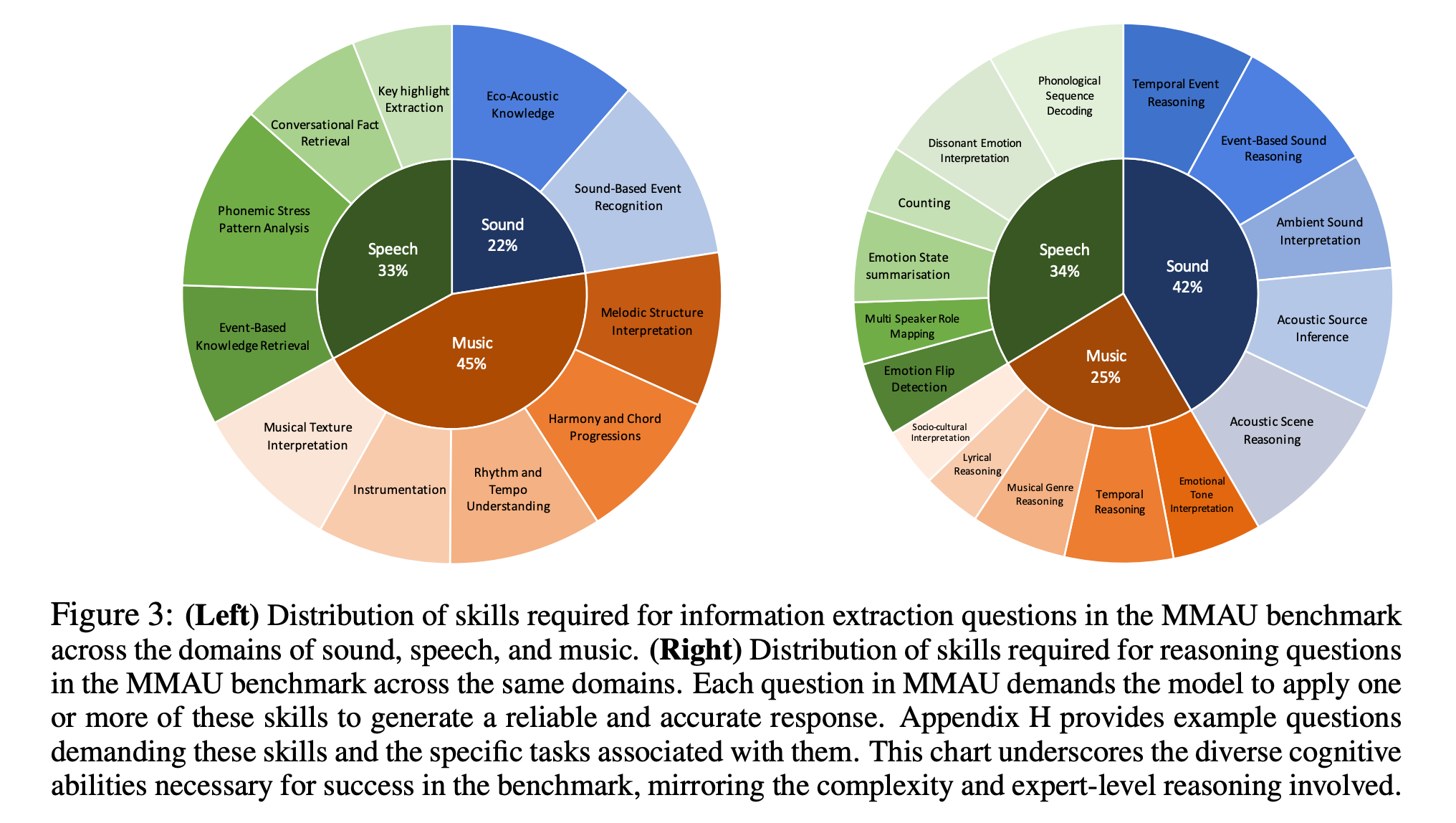
**音频分析 BenchMark 调研-MMAU 篇**

|  |
| --- |
| * 速记：MMAU 是一个规模为 10,000 条、多域（语音/speech、环境音/sound、音乐/music）、涵盖 27 种信息抽取与推理技能的“多任务音频理解与推理基准”，旨在评估音频语言模型在复杂认知任务上的能力 —— 当前顶尖模型[r1-aqa](https://github.com/xiaomi-research/r1-aqa)准确率仅约 65%，仍有巨大提升空间。 * Paper: https://arxiv.org/pdf/2410.19168 * Leaderboard: https://sakshi113.github.io/mmau\_homepage/ * 数据集地址： * Pfs: /mnt/pfs-mc0p4k/asr/team/group\_data/speaker/3oa\_data/MMAU/data * Lpai: /lpai/dataset/asr-group-data/0-1-0/group\_data/speaker/3oa\_data/MMAU/data |

1. **简介**

* MMAU 是一个主打音频分析类任务的评测 benchmark，任务覆盖度较为全面，且使用 ACC 作为指标呈现， ACC 的计算方式是通过子串匹配；
* 任务分布：



2. **评测形式**

* MMAU 使用选择题形式来让大模型预测答案，之后通过子串匹配的方式计算 ACC，评测中官方推荐使用的 prompt 是：

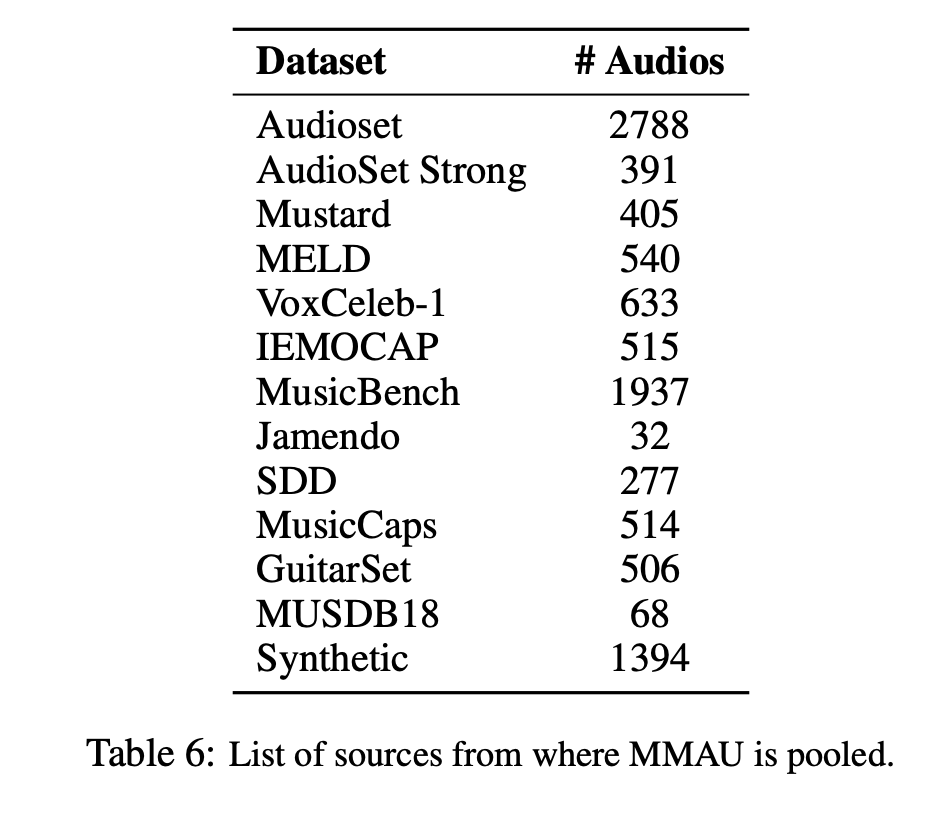
|  |
| --- |
| python f"""{question} Select one option from the provided choices.\n{choices}. |

3. **数据样例**

**[同步块-无权限下载此内容]**

4. **数据构造方式**

4.1 **数据组成**



5. **BenchMark 涵盖的任务**

|  |
| --- |
| 论文中给出的任务定义相对较为模糊，有待进一步合并整理 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 类型 | 能力 | 任务 | 任务归类 | 样例 |
| Sound | 时序事件推理 | 识别音频中的事件次数/时长 | 音频事件检测 |  |
| 声源识别 | 识别声音来源 | 音频事件检测 |  |
| 声学知识 | 音频场景检测 | 音频场景检测 |  |
| 环境音/音乐风格解析 | 背景音信息抽取 | 音频描述？ |  |
| 声学场景推理 | 根据音频回答相关问题：音频描述 | 音频描述？ |  |
| 音频事件推理 | 音频事件因果推理 | 音频事件因果推理 |  |
| 音频事件检测 | 音频事件检测 | 音频事件检测 |  |
| Speech | 负面情感检测 | 多说话人场景的负面情感检测 | 情感识别 |  |
| 基于事件的知识抽取 | 提取对话的核心知识 | 语音问答 |  |
| 计数问题 | 说话人计数 | 语音问答 |  |
| 语音重音模式分析 | 识别语音中的重音模式 | 语音问答 |  |
| 情感状态总结 | 多说话人情感识别 | 情感识别 |  |
| 对话事实抽取 | 根据对话内容提取关键事实 | 语音问答 |  |
| 多说话人角色判断 | 识别语音中每个说话人的角色 | 音频描述？  语音问答？ |  |
| 发音序列解码 | 识别单词的发音顺序 | 语音问答 |  |
| 情感转换检测 | 识别对话中的说话人情感是否有变化 | 情感识别 |  |
| 对话意图检测 | 识别对话意图 | 语音问答 |  |
| Music | 时序推理 | 识别音乐/歌声中的时序信息 | 音乐描述 |  |
| 旋律推理 | 旋律/曲风识别 | 曲风识别 |  |
| 歌词推理 | 歌词的情感、主题、含义识别 | 语音识别 |  |
| 社会文化解读 | 识别音乐出现的社会背景 | 曲风识别 |  |
| 音乐旋律解读 | 推断旋律的组织结构和进展过程，以理解其模式、形式和情感表达。 | 音乐描述 |  |
| 和旋解读 | 和旋识别 | 和旋识别 |  |
| 节奏理解 | 节奏识别 | 节奏识别 |  |
| 歌声理解 | 歌声理解 | 曲风识别 |  |
| 乐器识别 | 乐器识别 | 乐器识别 |  |
| 音乐情感识别 | 音乐情感识别 | 情感识别 |  |

6. **其他**

* MMAU 分为 mini 集（1000 条）和全量集（10000 条），全量测试集不完全开源，仅开源了问题+音频，答案闭源，如果想测试需要提交模型推理结果到[官方](https://eval.ai/web/challenges/challenge-page/2391/overview)来进行评测；