ICS 35.080
UNSPSC 43.23.34
CCS L 77

# 团 体 标 准

T/UNP 253—2024

# 语音数据标注系统技术规范

Technical specification for speech data annotation system

全国团体标准信息平台

2024－11－07 发布　　2024－11－07 实施

中国联合国采购促进会　　发 布

全国团体标准信息平台

# 目　　次

前言 ..... III
引言 ..... IV
1　范围 ..... 1
2　规范性引用文件 ..... 1
3　术语、定义和缩略语 ..... 1
3.1　术语和定义 ..... 1
3.2　缩略语 ..... 1
4　系统架构 ..... 1
5　功能要求 ..... 2
5.1　音频导入与管理 ..... 2
5.2　标注工具 ..... 2
5.2.1　时间轴标注 ..... 2
5.2.2　文本转录 ..... 2
5.2.5　分类标注 ..... 2
5.2.6　质量检查工具 ..... 3
5.3　协作与管理功能 ..... 3
5.3.1　多人协作 ..... 3
5.3.2　任务分配与管理 ..... 3
5.4　导出与集成功能 ..... 3
5.4.1　数据导出 ..... 3
5.4.2　与机器学习框架集成 ..... 3
6　性能要求 ..... 3
6.1　高效率性 ..... 3
6.2　并行性 ..... 4
6.3　可靠性 ..... 4
6.4　容错性 ..... 4
6.5　用户友好性 ..... 4
7　数据要求 ..... 4
7.1　输入数据格式 ..... 4
7.2　输出数据格式 ..... 5
8　安全要求 ..... 5
8.1　系统安全 ..... 5
8.2　数据安全 ..... 5
8.2.1　数据完整性 ..... 5
8.2.2　数据储存 ..... 5
8.2.3　数据传输 ..... 5
8.2.4　数据备份 ..... 5

8.2.5 数据恢复 ..... 5
9 运维要求 ..... 5
10 评价与改进 ..... 6
参考文献 ..... 7

全国团体标准信息平台

# 前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由武汉铃声科技有限公司提出。

本文件由中国联合国采购促进会归口。

本文件起草单位：武汉铃声科技有限公司、武汉励为企业管理咨询有限公司、湖北达慕星客商贸有限责任公司、湖北国威时代信息技术有限公司、武汉锐佳讯科技有限公司。

本文件主要起草人：郑伟莲、卢蹁跬、余汉汉、李嘉茵、胡彬。

全国团体标准信息平台

# 引　言

为助力中国企业参与国际贸易，推动企业高质量发展，中国联合国采购促进会依托联合国采购体系，制定服务于国际贸易的系列标准，这些标准在国际贸易过程中发挥了越来越重要的作用，对促进贸易效率提升，减少交易成本和不确定性，确保产品质量与安全，增强消费者信心具有重要的意义。

联合国标准产品与服务分类代码（UNSPSC，United Nations Standard Products and Services Code）是联合国制定的标准，用于高效、准确地对产品和服务进行分类。在全球国际化采购中发挥着至关重要的作用，它为采购商和供应商提供了一个共同的语言和平台，促进了全球贸易的高效、有序发展。

围绕UNSPSC进行相关产品、技术和服务团体标准的制定，对助力企业融入国际采购，提升国际竞争力具有十分重要的作用和意义。

本文件采用UNSPSC分类代码由6位组成，对应原分类中的大类、中类和小类并用小数点分割。

本文件UNSPSC代码为“43.23.34”，由3段组成。其中：第1段为大类，“43”表示“信息技术广播和电信”，第2段为中类，“23”表示“软件”，第3段为小类，“34”表示“实用程序和设备驱动程序软件”。

全国团体标准信息平台

# 语音数据标注系统技术规范

## 1 范围

本文件规定了语音数据标注系统的系统架构、功能要求、性能要求、数据要求、安全要求、运维要求和评价与改进。

本文件适用于规范语音数据标注系统的设计与建设。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语、定义和缩略语

### 3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

**数据标注 data labeling**

给数据样本指定目标变量和赋值的过程。

[来源：GB/T 42755—2023，3.1]

3.1.2

**语音识别 speech recognition**

将人类的声音信号转化为文字或指令的过程。

[来源：GB/T 21023—2007，3.1]

### 3.2 缩略语

下列缩略语适用于本文件。

AAC：高级音频编码（Advanced Audio Coding）

CSV：逗号分隔值（Comma-Separated Values）

IP：网际互连协议（Internet Protocol）

MP3：动态影像专家压缩标准音频层面3（Moving Picture Experts Group Audio Layer III）

TCP：传输控制协议（Transmission Control Protocol）

## 4 系统架构

系统架构组成如图1。

图1　系统框架图

## 5　功能要求

### 5.1　音频导入与管理

5.1.1　系统应具备清晰、高效的语音输入功能，支持通过麦克风等音频采集设备录制的实时语音输入，并能对内语音进行实时标注。

5.1.2　系统应支持用户对音频进行录制和分析，且系统界面应显示对音频的时域分析结果和频域分析结果，其中：

a)　时域分析结果应包含：周期、峰值、频率、均值、幅值、均方差、相位、方差；

b)　频域分析结果应包含：周期、频率。

5.1.3　系统应支持用户进行多轨道、多音频文件操作，并清晰显示所操作音频的波形图和选中的时间点。系统音频传输应至少包括以下几种功能：

a)　音频控制：切换音频，并对目标音频进行快进、慢放、暂停或录制等操作；

b)　音频缩放：波形图显示大小的调控；

c)　音频剪辑：音频截段，通过选择音频的起始点来操作；

d)　音频调整：调整音频文件的速度、节拍、音调，且具备监听选项。

5.1.4　支持用户上传不同格式的语音文件，如 MP3、WAV、FLAC 等，适应不同来源的语音数据，并安全储存，音频文件或语音数据在输入后不应损坏。

5.1.5　系统所识别过的语音数据应显示音频文件的编号、名称、识别时间、语音处理方式等信息。其中，数据识别的时间应包括年、月、日，精确到秒。

5.1.6　对不同采样率、比特率的音频文件应具有良好的兼容性。

### 5.2　标注工具

#### 5.2.1　时间轴标注

5.2.1.1　支持标注者在音频的时间轴上精确地标记起始点和结束点，用于划分语音片段。

5.2.1.2　支持对不同的语音事件、说话人转换等进行准确标注。

#### 5.2.2　文本转录

5.2.3　提供文本输入界面，支持标注者将音频中的语音内容转录为文字。

5.2.4　支持多种语言的转录，具备自动纠错、拼写检查等功能，提高转录的准确性和效率。

#### 5.2.5　分类标注

5.2.5.1　提供预设的分类体系，支持标注者为音频选择合适的分类标签。

5.2.5.2　分类体系应具有可扩展性，支持管理员根据实际需求添加新的分类标签，满足不同项目的标注要求。
5.2.5.3　支持多级分类标注，对复杂的语音数据，进行多层次的分类标注。

#### 5.2.6　质量检查工具

5.2.6.1　具备自动检测音频质量的功能，如检测噪音水平、音频清晰度、音量大小等，提醒标注者对质量较差的音频进行特殊标注或处理。
5.2.6.2　系统应通过分析音频的频谱、信噪比等参数，自动判断音频的质量，并在界面中给出相应的提示。
5.2.6.3　支持标注者对标注结果进行自查和修正，确保标注的准确性和一致性。
5.2.6.4　提供标注结果的预览功能，标注者应在提交标注结果之前，对自己的标注进行检查和确认，并支持对比多人标注的结果。
5.2.6.5　支持管理员查看不同标注者对同一音频文件的标注结果，发现差异应进行协调和修正，提高标注的一致性。

### 5.3　协作与管理功能

#### 5.3.1　多人协作

5.3.1.1　支持多个标注者同时对同一项目进行标注。
5.3.1.2　具备实时同步功能，当一个标注者对音频文件进行标注时，其他标注者能立即看到标注结果的变化，避免重复标注和冲突。
5.3.1.3　支持标注者之间的交流和协作，允许标注者通过在线聊天、评论等方式进行沟通和讨论，共同解决标注过程中遇到的问题。
5.3.1.4　具备版本控制功能，防止多人协作时出现标注冲突，能记录标注的修改历史，便于追溯和复查。自动为每个标注结果生成版本号，当标注者对标注结果进行修改时，系统会自动创建一个新的版本，并记录修改的时间、内容和标注者等信息。
5.3.1.5　管理员可随时查看标注的历史版本，比较不同版本之间的差异，恢复到之前的版本或进行合并操作，确保标注结果的准确性和可追溯性。

#### 5.3.2　任务分配与管理

5.3.2.1　支持管理员根据标注者的技能和工作量，合理分配标注任务。
5.3.2.2　实时监控任务进度，了解每个标注者的工作状态和完成情况。

### 5.4　导出与集成功能

#### 5.4.1　数据导出

将标注好的语音数据以多种格式导出，导出的数据应包含音频文件信息、标注结果及相关的元数据。

#### 5.4.2　与机器学习框架集成

5.4.2.1　支持接口与常见的机器学习框架进行集成，标注好的数据能直接用于模型训练。
5.4.2.2　支持批量导入和导出数据，便于在不同的项目和工具之间进行数据迁移。

## 6　性能要求

### 6.1　高效率性

对需要实时标注、操作的场景，系统应快速响应用户的输入或上传，实时处理语音数据。系统的响应速度应符合表 1 的要求。

**表1　响应速度**

| 项目 | 影响因素 | 响应速度 |
| --- | --- | --- |
| 音频加载时间 | 音频文件大小、格式、网络带宽 | ≤2 s |

表1　响应速度（续）

| 项目 | 影响因素 | 响应速度 |
| --- | --- | --- |
| 实施回放延迟 | 音频处理能力、用户设备性能 | ≤200 ms |
| 标注操作响应时间 | 系统负载、标注数据复杂性 | ≤300 ms |
| 数据保存时间 | 数据存储机制、数据库响应速度 | ≤1 s |
| 搜索和筛选时间 | 数据量大小、索引效率 | ≤3 s |
| 用户界面响应时间 | 前端性能、后台处理能力 | ≤200 ms |

## 6.2　并行性

6.2.1　采用多线程处理技术和分布式计算架构，充分利用多核 CPU 和多台服务器的计算资源，对多个音频文件和标注任务进行并行处理。

6.2.2　支持数据分块处理和增量式标注。对大型语音文件，系统应将其分割成多个较小的数据块，分别分配给不同的标注者进行标注，标注完成后再进行合并。

6.2.3　支持标注者在已有标注的基础上进行修改和补充，无需重新标注整个音频文件。

## 6.3　可靠性

6.3.1　采用智能负载均衡算法，根据服务器的实时负载情况动态分配任务请求到负载较轻的服务器上，确保系统在大规模识别任务或数据处理需求并发的情况下负载均衡，能持续稳定运行。

6.3.2　进行充分的压力测试和性能优化。在系统上线前，通过模拟大量并发用户和大规模数据处理场景，对系统进行全面的压力测试，找出系统可能存在的性能瓶颈和潜在问题，并进行针对性的优化。

## 6.4　容错性

具备自动故障转移和容错机制。某台服务器出现故障（如硬件故障、软件崩溃等）时，负载均衡器能自动检测到故障，并将原本分配到该故障服务器上的任务迅速转移到其他正常服务器上继续处理，确保整个系统的服务不受影响。

## 6.5　用户友好性

6.5.1　系统界面设计应简洁直观，合理规划界面布局，各个功能区域划分明确。

6.5.2　音频导入、标注工具、协作管理等主要功能模块应在界面上有清晰的位置标识，且通过颜色、图标等方式进行区分，方便用户快速定位所需功能。

6.5.3　提供详细的操作指南和帮助文档，包括图文教程、视频演示等多种形式，应满足不同用户的学习需求。操作指南应涵盖从系统登录到完成复杂标注任务的全过程，针对每个步骤提供详细的操作说明和注意事项。

# 7　数据要求

## 7.1　输入数据格式

系统支持的输入数据格式见表2。

表2　输入数据要求

| 数据类型 | 采样率 | 是否支持 |
| --- | --- | --- |
| MP3 | 8000 Hz、16000 Hz | 是 |
| WMA | 44100 Hz、48000 Hz | 是 |
| WAV | 11025 Hz | 否 |
| APE | 44100 Hz | 是 |
| FLAC | 不限 | 否 |
| OCG | 不限 | 是 |
| AAC | 不限 | 否 |

### 7.2 输出数据格式

系统支持的输出数据格式见表3。

**表3 输出数据要求**

| 规范项目 | 要求 | 说明 |
| --- | --- | --- |
| 音频格式 | MP3、WAV | 符合7.1中的要求即可 |
| 采样率 | 16000 Hz、44100 Hz等 | 根据具体需求确定，保持一致性 |
| 比特率 | 128 kbps、256 kbps等 | 根据具体需求确定，保持一致性 |
| 标注准确性 | ≥95% | 根据实际应用场景可调整，例如，100个标注的文字中允许最多5个错误 |
| 时间轴标注精度 | 误差在±50 ms内 | 精确到毫秒级 |
| 标注文件格式 | JSON、CSV等 | 常用数据格式即可 |
| 标注一致性 | ≥85% | 不同标注者对同一音频的标注结果相似程度达到85%以上 |
| 元数据完整性 | 文件名、录制时间、标注者、标注时间、音频时长等信息无遗漏 | 确保元数据的完整 |
| 数据存储大小 | ≤10 MB | 根据存储设备和应用场景确定 |
| 数据编码 | UTF-8等通用编码格式 | 确保不同系统间的兼容性 |

## 8 安全要求

### 8.1 系统安全

8.1.1 系统应具备良好的抗攻击能力、容灾能力和警报功能，在被攻击时应及时提醒用户和管理者并抵御入侵。

8.1.2 系统应具备自我安全检测能力,定期检查系统运行环境,并对损坏和泄露的数据出具分析报告。

### 8.2 数据安全

#### 8.2.1 数据完整性

系统应采用可靠的传输协议（如TCP/IP协议），对传输的数据进行校验和验证。在数据发送端，对要传输的数据计算校验，在接收端收到数据后再次计算校验并与发送端的校验进行比对，如果不一致则自动重传数据，确保数据在传输过程中的准确性和完整性。

#### 8.2.2 数据储存

8.2.2.1 采用冗余存储技术，如分布式文件系统或数据库的主从复制、集群部署等方式，将数据存储在多个节点上。某个存储节点出现故障时，系统应自动切换到其他正常节点。

8.2.2.2 系统应定期对存储的数据进行完整性检查，验证数据是否被篡改或损坏，如发现问题及时进行修复或恢复。

#### 8.2.3 数据传输

数据经过网络传输或交换时不应发生增加、修改、丢失和泄露。

#### 8.2.4 数据备份

每次操作后应对所需的音频数据进行备份。

#### 8.2.5 数据恢复

系统崩溃或服务器故障后不用重新安装系统，且支持用户直接快速恢复崩溃或故障前的数据或文件。

## 9 运维要求

9.1 系统开始运作前，技术人员应对系统的前端设备和各个功能进行调试。

9.2 应进行定期培训，建立成熟的技术团队，预先计划好应急预案。
9.3 应设立监测系统，实时监控关键性能指标，如延迟、响应时间、处理速度、资源利用率等。监测系统应具备集中联动报警功能，将设备故障、系统崩溃、服务器故障等各类故障进行及时预警报警，上传故障发生时间减少系统维护的响应时间。
9.4 发生事故后应进行过程记录、结果的可视化处理，技术维修人员应及时分析事故原因。
9.5 应定期对系统软件和硬件组件进行更新和升级，并在新版本上线前对系统进行控制和测试，对更新过程可能出现的兼容性问题做出预防。

## 10 评价与改进

依据第 5 章、第 6 章、第 7 章、第 8 章、第 9 章的要求确定系统的评价内容，定期开展系统功能、性能、数据、安全、运维方面的评价，审查不合格项，并有针对地采取纠偏措施。

全国团体标准信息平台

# 参 考 文 献

[1] GB/T 21023 中文语音识别系统通用技术规范
[2] GB/T 37973 信息安全技术 大数据安全管理指南
[3] GB/T 42755 人工智能 面向机器学习的数据标注规程

全国团体标准信息平台