会议对话语音数据库

AISHELL-ASR0055-2

产品说明书

北京希尔贝壳科技有限公司

bd@aishelldata.com

目录

1 产品概述

AISHELL-ASR0055-2 会议对话语音数据库共 317 场会议，共 285 有效小时。录音语言，中文；录音地区，中国。会议内容覆盖商务、生活、工作等。以中国北方口音区域为主邀请 106 名发音人参与录制。录制过程在真实会议环境中，录制设备包括头戴式麦克风、1 个真实会议语音记录设备、Android 系统平板、iOS 手机、8 麦线阵和 16 麦圆型麦克风阵列。音频存储格式为 16kHz，16bit。

此数据库经过专业语音校对人员转写标注，并通过严格质量检验，文本正确率在 96%以上。

2 场景与设备

2.1 采集场景

采集场景为小型、中型、大型的三类会议场景。每类会议场景的说话人数在 3-10 人。

会议场景的具体要求如下表：

序号	场景类型	场景大小	场景数（个）
1	**Small**	20m^2 ≥ room	6
2	**Medium**	20m^2<room≤50m^2	4
3	**Large**	Room>50m^2	2

图表 2-1-1

会议场景噪音定义为两类：

序号	类型	内容
1	自带噪音	敲键盘/风扇空调声/轻微的干扰人声/手机铃声
2	人工加入噪声源	鸣笛声等

图表 2-1-2

真实场景实例：

2.2 采集设备

考虑到会议场景和多人对话的特殊性，数据存储格式为 16kHz、16Bit。采集的设备我们选择如下几种：

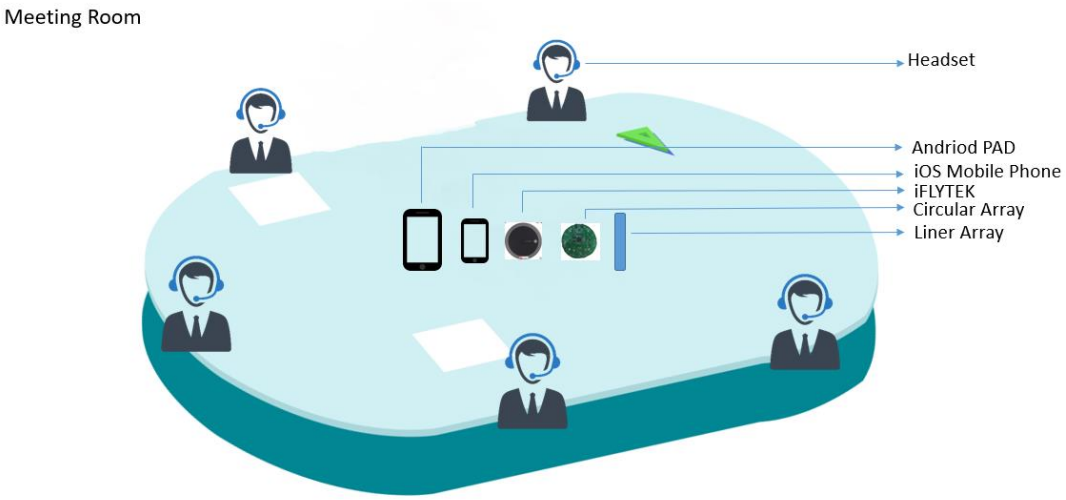
序号	设备
1	讯飞听见 M1
2	Android 系统平板
3	头戴式麦克风
4	iOS 手机
5	8 麦线阵
6	16 麦圆型麦克风阵列

图表 2-2-1

3 采集方法

会议场景采集语音内容以真实项目、业务为核心内容，会议参加人员自然发挥，轻口音普通话语速及中英文不做限制，保证事件真实。场景由实施人员在指定位置采集数据，采集设备摆放合理。

※会议室语音采集位置示意图：



采集现场记录数据内容如下：

记录项	内容
会议室	类型/录制时间/人数
录音人	个人基本信息（性别/年龄/）
录音位	设备/编号
噪音	噪音类型

图表 3-1

4 录音人信息

4.1 性别比例

数据库总人数为 106 人，其中男 46 人，女 60 人。

性别	男性	女性	合计
比例	43.4%	56.6%	100%

图 4-1-1 性别分布

4.2 年龄比例

录音人年龄覆盖 18~64 岁，具体比例如下所示：

年龄	人数	比例
18~24	54	50.9%

25~40	31	29.3%
≥41	21	19.8%
合计	106	100%

图表 4-2-1 年龄分布

5 标注转写规范

数据转写人员根据所听到的音频写出内容，力求使文本内容与音频发音内容保持一致。准则如下：

1）转写的内容必须和听到的语音完全一致，不能多字、少字、错字。

2）数字要转写为汉字形式，如“一二三”，而不是“123”。 注意区分“一”和“幺”，“二”和“两”。

3）句中出现的英文按照发音写出单词，如“thank you”。按拼读朗读的字母，需转写成大写字母加空格的形式。如，“ＮＢＡ”、“ＵＦＯ”。注：汉字间不要有空格。中英混合情况，英文之间要加空格。比如字母和字母间，字母和单词间，单词和单词间都要加空格。汉字和英文间不加空格。

4）句中包含的符号，按实际发言人发音转写。如“三 W 点 百度 点 com”。没有发音的符号，需要删掉。 品牌名称，专有名称等按照实际惯用格式转写，如“QQ 空间”、“iPhone”、“喜马拉雅”。

5）标注内容的完整性要与实际发音一致，不得删减。

6) 重叠音，不同的说话人内容标注在对应说话人层，例如有四个说话人的会议，就有四个说话人层， 每个说话人层只标注对应的说话人的说话内容，即便有说话重叠，也只标注这个说话人的，重叠说话人的声音标注在对应的说话人层。

细分场景参照如下：

- 多个人说话时，在重叠处将两个人的声音都进行标注并用 & 符号进行区分；
- 多人重叠，其中一人整句无法听清时，则该人当背景音忽略不标；
- 多人重叠，其中一人说整句话但一部分无法听清时，则只标注听清的部分；
- 多人重复，非主说话人的语气词。语气词：直接标注嗯，啊，呵等，{针对主说话人，出现了非主说话人的清楚语气词需要标注，格式为&嗯&、&啊&；

若主说话人说的语气词，直接正常标注就可以，无需加“&”}[语气词都要有口字旁,除了诶]。

7） 特殊符号标注，详见下表：

序号	标签	说明
1	<sil>	说话人犹豫、拖延时间、停顿、拖长音
2	<->	发出半音卡壳
3	<$>	笑声
4	<_>	切音，开头或者结尾句子截断
5	<%>	咳嗽
6	<#>	人声杂音，如吸气、打哈欠
7	&	重叠音

图 5-1-1 特殊符号说明

6 overlap ratio

会议中常有录音人说话重叠的情况，overlap ratio 的计算方法如下：

Overlap ratio=overlap 总长/录音人说话总时长

本数据库的 overlap ratio 详情如下：

Overlap ratio(%)	场次	占比（%）
<=20	48	15.1
20~40	47	14.8
41~60	54	17.0
>=60	168	53.1
合计	317	100

图 6 Overlap Ratio 的场次分布

7 数据文件目录

7.1 目录结构

数据目录树	
数据目录结构	
| AISHELL-ASR0055-2 数据产品说明书.pdf	**（数据库简介）**
└─DOC	**（文本说明文件）**
| ├─all_wav_list.txt	**（音频列表）**
| ├─spk_info.xlsx	**（录音人信息）**
|	

```
└L                                                          (会议室类型)
  |├L_R005                                                  (会议室编号)
    |├20211111_L_R005S01                                    (会议场次)
      |├wav                                                 (音频文件夹)
        |├20211111_L_R005S01A                               (音频设备)
          ||   ||      20211111_L_R005S01A004N01.wav        (音频文件)
        |├TextGrid                                          (文本文件夹)
          ||   ||      20211111_L_R005S01A.TextGrid         (转写文本)
    |├L_R005.jpg                                            (会议场景图)
```

7-1-1 数据目录结构

7.2 命名规则

7.2.1 目录命名规则

/<VENUE_TYPE>/<VENUE_ID>/<Session_ID>/<WAV>/<Device_ID>/<AUDIO_ID>

e.g. L/L_R005/20211111_L_R005S01/wav/20211111_L_R005S01A/20211111_L_R005S01A004N01.wav

目录	内容	备注
VENUE_TYPE file	Small / Medium / Large	会议场景类型
VENUE_ID file	L_R005	场地编号
Session_ID	20211111_L_R005S01~S08	会议场次
Device_ID	A/P/T/CM	录音设备
AUDIO file	20211111_L_R005S01A004N01.wav	WAV 文件

图表 7-2-1

7.2.2 文件命名规则

/<TIME_ID>/<VENUE_TYPE>/<VENUE_ID>/<MEETING_ID>/<DEVICE_ID>.wav

e.g. 20211111_L_R005S01A004N01.wav

文件	内容	备注
TIME_ID	20211111	录制时间
VENUE_TYPE	_S / M /L_	场地类型
VENUE_ID	R001~500	场地编号
MEETING_ID	S01 ~ 08	会议编号

DEVICE_ID	C01 & C02	设备编号
Channel_ID	N01~16	音频通道

图表 7-2-2

7.2.3 设备信息

设备相对均匀分布在录音人中间，方向遵守由北向南编号。

序号	设备	编号
1	讯飞听见 M1	M01
2	Android 平板	T01
3	头戴式麦克风	A
4	iOS 手机	I01
5	8 麦线阵	L01
6	16 麦圆型麦克风阵列	C01

图表 7-2-3

8 版权声明