会议对话语音数据库

AISHELL-ASR0055

产品说明书

北京希尔贝壳科技有限公司
Beijing Shell Shell Technology Co., Ltd.
Add：Room 813, Building No. 4, Shangdi 10th Street, Haidian District, Beijing 100080, P.R.China
Tel：010-80225006　E-mail：bd@aishelldata.com

目录

1 产品概述

AISHELL-ASR0055 会议对话语音数据库共 639 场会议，共 370 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 162 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、高保真麦克风、Android 系统平板、iOS 手机、16 麦面阵和 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 95%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	8
2	**Medium**	20m^2<room≤50m^2	8
3	**Large**	Room>50m^2	4

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	高保真麦克风
5	iOS 手机
6	16 麦面阵
7	16 麦圆型麦克风阵列

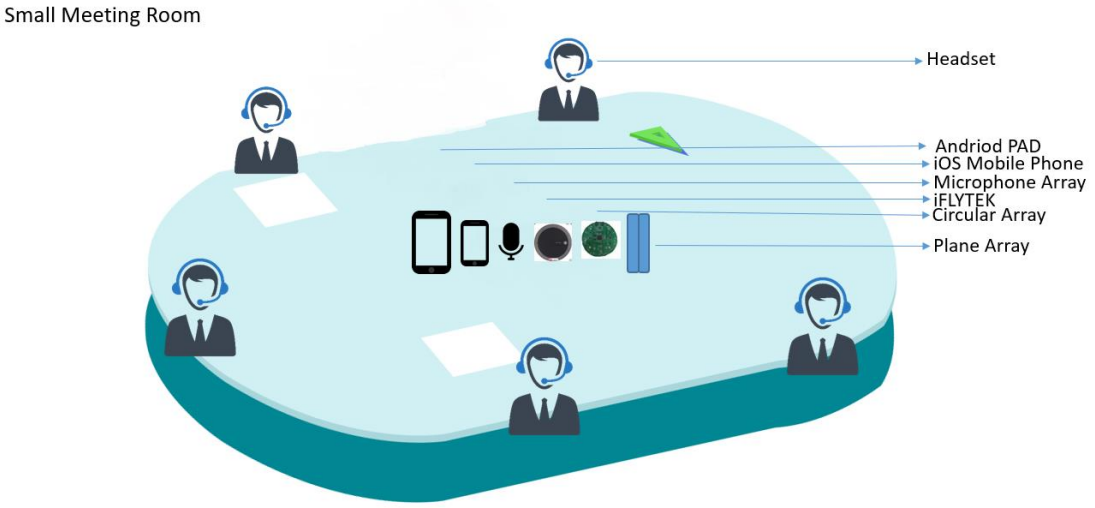
图表 2-2-1

3 采集方法

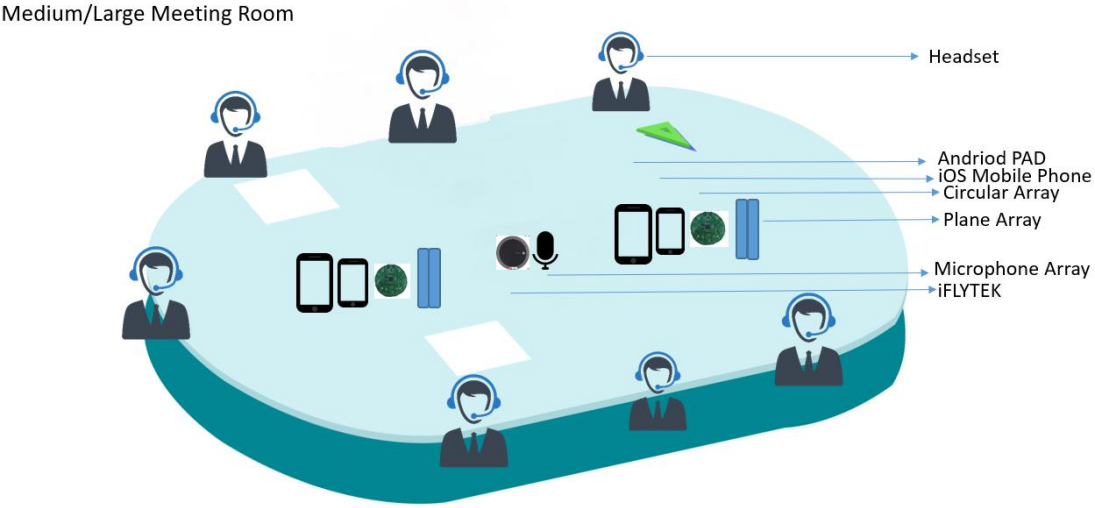
会议场景采集语音内容以真实项目、业务为核心内容，会议参加人员自然发挥，轻口音普通话语速及中英文不做限制，保证事件真实。场景由实施人

员在指定位置采集数据，采集设备摆放合理。

※小会议室语音采集位置示意图：



※中/大型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 162 人，其中男 66 人，女 96 人。

性别	男性	女性	合计
比例	41%	59%	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	90	55.6%
25~40	37	22.8%
≥41	35	21.6%
合计	162	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，并在线下用 Praat 软件将重叠音分层标注；
- 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。比如 ABC 重叠，B 无法听清，则标注为：1 说话内容&c 说话内容；
- 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。比如 ABC 重叠，B 部分能听清，则标注为：1 说话内容&b 听清内容&c 说话内容。
- 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）角色区分，要求将参会人员全部区分标注，为操作简便，可如下处理：

- 标注文本开头加上角色对应的数字。如 A—1、B--2 以此类推。比如录音人 A 说“你吃饭没”则文本最终标注为“1 你吃饭没”；
- 多人重叠标注时，后面的角色用小写字母代替对应的角色，如 A—a、B--b 以此类推。比如 A、B、C 三人重叠，A 先说，则标注为“1 说话内容&b 说话内容&c 说话内容”。每个说话人对应一份标注文本。

注：如无法区分角色则不标注角色，但是说话内容需要正常标注，标注内容前加“&”。

8） 特殊符号标注。

6 数据文件目录

6.1 目录结构

数据目录树
数据目录结构 **| AISHELL-ASR0055 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)** **| ├─spkrinfo.xlsx (录音人信息)** **|** **└─TextGrid (文本文件夹)** **| ├─20200807 (期号文件夹)** **| | | | S_R001S08M01.TextGrid (音频内容文件)** **|** **└─SPEECHDATA (数据文件夹)** **| ├─20200807 (期号文件夹)** **| | | | S_R001S08M01N01.wav (音频文件)**

6-1-1 数据目录结构

6.2 命名规则

6.2.1 目录命名规则

/<USAGE>/<MEETING_ID >/<Speech_ID>

e.g.SPEECHDATA/20200807/S_R001S08M01N01.wav

目录	内容	备注
USAGE	SPEECHDATA	文件夹名称
MEETING_ID	20200807	会议期号
SPEECH_ID	S_R001S08M01N01.wav	WAV 文件

图表 6-2-1

6.2.2 文件命名规则

/<ROOM_ID>/<MEETING_NUM>/< DEVICE_ID >/<CHANNEL_ID>.wav

e.g. S_R001S08M01N01.wav

文件	内容	备注
ROOM_ID	S/M/L_R001~500	会议室编号
MEETING_NUM	S01~08	会议场次
DEVICE_ID	M01 & 02	设备编号
CHANNEL_ID	N01~16	信道编号

图表 6-2-2

6.2.3 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 & 02
3	头戴式麦克风	A01
4	高保真麦克风	H01
5	iOS 手机	I01
6	16 麦面阵	L01 & L02
7	16 麦圆型麦克风阵列	C01 & C02

图表 5-2-3

7 版权声明