

中文普通话麦克风阵列家居场景

语音数据库



希尔贝壳 AISHELL-DMASH Copyright

A I S H E L L

产品说明书

目录

1 产品概述.....	2
2 录音语料.....	2
2.1 语料池的制作.....	2
2.1.1 录音文本的结果设计	2
2.1.2 语料池处理	3
3 发音人信息.....	3
3.1 发音人结构特征.....	3
3.1.1 性别比例	3
3.1.2 年龄比例	3
4 数据录制环境.....	4
4.1 录制环境.....	4
4.2 录制设备.....	4
4.3 录制方法.....	4
5 语音数据转写.....	5
6 数据文件目录.....	5
6.1 目录结构.....	5
6.2 命名规则.....	6
6.2.1 目录命名规则（图表 6-2-1）	6
6.2.2 文件命名规则（图表 6-2-2）	6
7 版权声明.....	7

1 产品概述

AISHELL-DMASH 语音数据库共 20253 小时。录音语言为中文；录音地区为中国。共邀请 511 名录音人，每人在不同日期分别录制 3 次语音，时间跨度为 7-25 天。录音人年龄段为 17-66 岁，此数据库可用于声纹识别、语音识别等研究。

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 录音语料

2.1 语料池的制作

2.1.1 录音文本的结果设计

此数据库录音文本采用每人每次约 330 句，每人 3 次录制共约 990 句，语音内容覆盖语速为快、中、慢的唤醒词和相关交互需求内容；数字；不重复的多领域词汇句子以及自然对话。具体结果结构如下。

领域	内容	编号
唤醒	你好，米雅（快）	0001-0010
唤醒	你好，米雅（中）	0011-0020
唤醒	你好，米雅（慢）	0021-0030
唤醒交互	你好，米雅+交互（快）	0031-0050
唤醒交互	你好，米雅+交互（中）	0051-0070
唤醒交互	你好，米雅+交互（慢）	0071-0090
数字	连续数字串	0091-0110
手机&平板语控	操作控制语句	0111-0170
音箱交互	音箱语控和命令	0171-0210
领域无关	广域文本	0211-0270
自然对话	自然交互语句	0271-330
	合计	330

图表 2-1 文本结构

2.1.2 语料池处理

- 脱敏处理。删除政治敏感、个人隐私、色情暴力等内容。
- 删除 <, >, [,], ~, /, \, = 等符号，数字串部分除外。
- 删除含有中文和英文以外语言的内容。
- 删除单句含有 25 字以上的内容。
- 统一格式。

3 发音人信息

3.1 发音人结构特征

3.1.1 性别比例

数据库总人数为 511 人，男 211 人，女 300 人。（图表 3-1-1）

性别	男性	女性	合计
A 比例	41.3%	58.7%	100%

图表 3-1-1

3.1.2 年龄比例

录音人年龄覆盖 17-66 岁，具体分布如下所示：A(16-25 岁)；B(26-40 岁)；C(41 岁以上)。（图表 3-1-2）

	年龄段	人数	比例
A	16-25 岁	481	94%
B	26-40 岁	15	3%
C	> 41 岁	15	3%
合计		100	100%

图表 3-1-2

4 数据录制环境

4.1 录制环境

安静室内，无回音，不同场景分别加入家居和会议场景噪声。发言人按照慢速、正常语速以及快速朗读录音文本。

4.2 录制设备

录制设备覆盖高保真、安卓系统手机、iOS 系统手机、安卓系统平板电脑以及半径为 5 厘米的圆形麦克风阵列板。

4.3 录制方法

共计 7 个点位，分别包含录音人近讲，后 1.5 米、0.25 米、1 米、左右各 3 米、5 米。设备采集位置示意图如下所以：

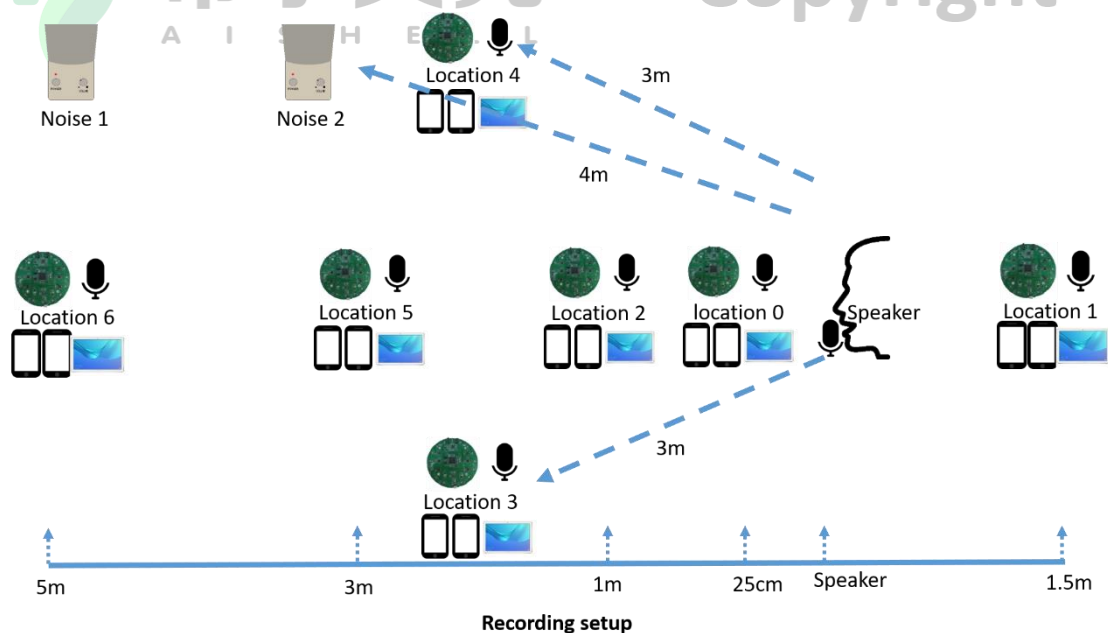


图 4-3-1

注：Noise1 为风扇噪声；Noise2 为电视、音乐、办公噪声

5 语音数据转写

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。

一般准则如下：

- 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) 音频中有英文发音的应写成相应的汉字或英文。具体分为以下几种情况：

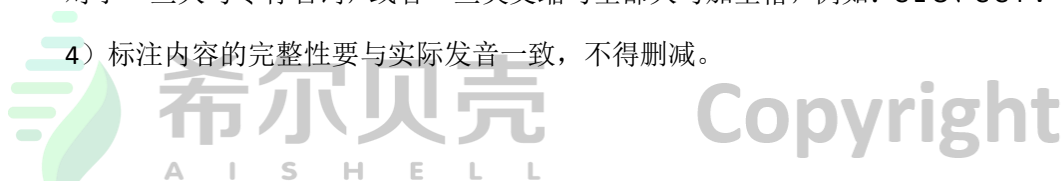
网址中包含的所有的字母或单词，均为大写。例如：发音内容为“www.abc.com”，应转写为“三 W 点 A B C 点 com”

发音中包含的英文单词，转写时全部为小写。

发音中包含的英文字母，转写时全部为大写。

对于一些大写专有名词，或者一些英文缩写全部大写加空格，例如：CEO、CCTV 等。

- 4) 标注内容的完整性要与实际发音一致，不得删减。



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-DMASH.pdf	(数据库简介)
└─DOC	(文本说明文件)
├─wav_list.txt	(音频列表)
├─spkrinfo.xlsx	(录音人信息)
└─SPEECHDATA	(数据文件夹)
├─0548	(录音人编号)
├─F0548	(第一次录制)
├─548A0.25M	(设备及点位信息)
F0548_548A0.25M_1_0019_normal.wav	(音频文件)
├─F0548.txt	(语音内容文本)

-S0548	(第二次录制)
-T0548	(第三次录制)

6.2 命名规则

6.2.1 目录命名规则（图表 6-2-1）

<USAGE>/<SpeakerID>/<Time>/<Recording_info>/<wav_ID>

e. g. SPEECHDATA/0548/F0548/548A0.25M/F0548_548A0.25M_1_0019_normal.wav

目录	内容	备注
USAGE	SPEECHDATA	文件夹名称
SpeakerID	0548	录音人编号
Time	F0548	录制次数
Recording_info	548A0.25M	录制信息
wav_ID	F0548_548A0.25M_1_0019_normal.wav	音频文件

图表 6-2-1

6.2.2 文件命名规则（图表 6-2-2）

<Time><Speaker_ID><Device_ID><Point_ID><Channel_ID><WAV_NUM><Speed>

e. g. F0548_548A0.25M_1_0019_normal.wav

文件	内容	备注
Time	First (F)、Second (S)、Third (T)	录制次数
SPEAKER_ID	0001-0550	录音人 ID
Device_ID	A: 安卓系统手机、I: iOS 系统手机、MIC: 高保真麦克风、 PAD: 安卓系统平板、PCM: 16 麦圆形阵列板	设备信息
Point_ID	0.25M、-1.5M、1M、3M、L3M (左)、R3M (右)、5M、	点位信息
Channel	1-16	信道信息
WAV_NUM	0019	WAV 编号
Speed	normal、fast、slow	语速

图表 6-2-2

7 版权声明

本文内容禁止转载，AISHELL(北京希尔贝壳科技有限公司)对本文拥有修改权、更新权及最终解释权。



Copyright



Copyright