

中文情感语音数据库

[AISHELL-ASR0013]

北京希尔贝壳科技有限公司

Beijing Shell Shell Technology Co.,Ltd

Add: Room 3-621, 6F, Zhongguancun Lifangting No. 1, Shanyuan Road, Haidian District, Beijing 100080, P.R.China

Tel: +86 10 80225006 E-mail: bd@aishelldata.com

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	2
3 发音人信息.....	3
3.1 基本信息记录.....	3
3.2 发音人结构特征.....	3
3.2.1 性别比例	3
3.2.2 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.....	4
6 数据文件目录.....	4
6.1 目录结构.....	4
6.2 命名规则.....	5
6.2.1 目录命名规则（图表 6-2-1）	5
6.2.2 文件命名规则（图表 6-2-2）	5
7 版权声明.....	6

1 产品概述

希尔贝壳中文情感语音数据库 AISHELL-ASR0013 共 27.3 小时， 28891 条语音。录音语言，中文；录音地区，中国。以中国北方口音区域为主邀请 100 名发音人参与录制。每个录音人分别录制悲伤、愤怒、恐惧、快乐、中性共 5 种情绪，录制过程为安静室内。录制设备覆盖高保真、安卓系统手机、iOS 系统手机，音频采样率 16Khz，比特率 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，字正确率 100%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份 300 句，每种情绪 60 句的分配方式设计，从语料池中抽取，结构如下。

序号	情绪	备注
1	sad	0001-0060
2	angry	0061-0120
3	fear	0121-0180
4	happy	0181-0240
5	Neutral	0241-0300

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。
- 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 基本信息记录

发音人信息记录内容包括任务编号、性别、口音区域、年龄区间。（图表 3-1）

任务编号	性别	口音区域	年龄区间
W0001	男	北方	A

图表 3-1

任务编号：每个发言人领取 1 个任务编号，每个任务编号对应 1 份录音文本。每个发言人只能参加一次录制。

口音区域：按照发言人原生语言所属区域划分。

年龄区间：A(16-25 岁)、B(26-40 岁)、C(41 岁以上)。

3.2 发音人结构特征

3.2.1 性别比例

数据库总人数为 100 人，男 51 人，女 49 人。（图表 3-2-1）

性别	男性	女性	合计
比例	51%	49%	100%

图表 3-2-1

3.2.2 年龄比例

A(16-25 岁)；B(26-40 岁)；C(41 岁以上)。（图表 3-2-2）

	年龄段	人数	比例
A	16-25 岁	82	82%
B	26-40 岁	10	10%
C	> 41 岁	8	8%
合计		100	100%

图表 3-2-2

4 数据录制环境

4.1 录制环境

安静室内,不包括明显的其他人说话声音及其他噪音,无回音。发言人按照正常语速,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。

4.3 录制方法

发言人距离设备 20~30 厘米,正常音量、语速朗读录音文本。

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-ASR0013_中文情绪语音数据库.docx		(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)
└─SPEECHDATA		(数据文件夹)
├─Female		(性别信息)
├─M0002		(录音人信息)
├─angry		(情绪信息)
	0002W0061.wav	(音频文件)
	0002W0061.txt	(语音内容文本)

6.2 命名规则

6.2.1 目录命名规则（图表 6-2-1）

<USAGE>/<Gender>/<SpeakerID>/<mood>/<wav_ID>

e. g. SPEECHDATA/Female/M0002/angry/0002W0061. wav

目录	内容	备注
USAGE	SPEECHDATA	文件夹名称
Gender	Female & Male	性别信息
SpeakerID	M0001	录音人 ID
mood	angry	情感分类
wav_ID	0002W0061.wav	音频文件

图表 6-2-1

6.2.2 文件命名规则（图表 6-2-2）

<SPEAKER_ID><WAV_NUM>

e. g. 0002W0061. wav

文件	内容	备注
SPEAKER_ID	0002	录音人 ID
WAV_NUM	W0061	WAV 编号

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。



希尔贝壳
A I S H E L L

Copyright