

中文情感语音合成数据库

[AISHELL-TTS0013]



希尔贝壳
A I S H E L L

Copyright

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.....	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.....	6

1 产品概述

希尔贝壳中文情感语音合成数据库 AISHELL-TTS0013 共 27 小时，28883 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。安静录音室录制，设备覆盖高保真，音频采样率为 44.1kHz & 16kHz，比特率 16bit。

此数据库经过专业语音校对人员转写标注，标注包含字、拼音以及韵律，并通过严格质量检验，文本正确率达到 99%。

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。
- 音素覆盖全面。
- 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(18-25 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(18-25 岁)；B(26-40 岁)；C(41 岁以上)。（图表 3-2-2）

	年龄段	人数	比例
A	18-25 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，信噪比达到 30db 以上。发言人按照正常语速，朗读录音文本。

4.2 录制设备

录制设备为高保真。

4.3 录制方法

发音人距离设备 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；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。

5) 拼音以及声调根据实际发音转写，拼音和文本需一致，并考虑变调以及儿化音，如下情况：

拼音应与实际发音一致，如“这”分清楚读的是“zhe4”还是“zhei4”等。

上声的变调，33 变调：了解 23；333 变调：展览馆 223、很勇敢 323。

多个上声 3 要根据紧密程度，划分出语义停顿，再根据上述规则进行变调。

“一、不”变调：单念或语句末尾声调不变；去声 4 前变为 2 等。

儿化音在原来音节加上“r”即可，如花儿 -huar。

6) 韵律标注包含韵律词 (#1)、韵律短语 (#2)、语调短语 (#3) 以及语调组边界 (#4)，音频与语法不一致的情况，以实际音频发音为准；相邻两级之间难以确定的情况，采取低一级标注。

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-TTS0013_中文情绪语音合成数据库.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)
─content.txt	(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)

图表 6-1-1 数据目录结构

6.2 命名规则

6.2.1 目录命名规则（图表 6-2-1）

$\langle \text{USAGE} \rangle / \langle \text{Gender} \rangle / \langle \text{SpeakerID} \rangle / \langle \text{mood} \rangle / \langle \text{wav_ID} \rangle$
 e. g. SPEECHDATA/Feamle/M0002/angrg/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）



$\langle \text{SPEAKER_ID} \rangle \langle \text{WAV_NUM} \rangle$
 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。

