



声联网科技
AVSNest Technology

AVS-H家庭环境婴幼儿声音场景和声音事件
数据规格手册

深圳声联网科技有限公司



0755-33349168



0755-33349798



www.avsnest.com



声联网官方订阅号



家庭环境婴幼儿声音场景和声音事件 AVS-H 数据规格手册修正记录		
版本	发布日期	内容描述
1.0	2018-05-16	初稿

AVSNest



重要声明

版权声明

版权归深圳声联网科技有限公司所有，保留所有权利。

商标声明

深圳声联网科技有限公司的产品是深圳声联网科技有限公司专有。在提及及其他公司及其产品时将使用各自公司所拥有的商标，这种使用的目的仅限于引用。本文档可能涉及深圳声联网科技有限公司的专利（或正在申请的专利）、商标、版权或其他知识产权，除非得到深圳声联网科技有限公司的明确书面许可协议，本文档不授予使用这些专利（或正在申请的专利）、商标、版权或其他知识产权的任何许可协议。

不作保证声明

深圳声联网科技有限公司不对此文档中的任何内容作任何明示或暗示的陈述或保证，而且不对特定目的的适销性及适用性或者任何间接、特殊或连带的损失承担任何责任。本手册内容若有变动，恕不另行通知。本手册例子中所用的公司、人名和数据若非特别声明，均属虚构。未得到深圳声联网科技有限公司明确的书面许可，不得为任何目的、以任何形式或手段（电子的或机械的）复制或传播手册的任何部分。

保密声明

本文档（包括任何附件）包含的信息是保密信息。接收人了解其获得的本文档是保密的，除用于规定的目的外不得用于任何目的，也不得将本文档泄露给任何第三方。

本软件产品受最终用户许可协议（EULA）中所述条款和条件的约束，该协议位于产品文档和/或软件产品的联机文档中，使用本产品，表明您已阅读并接受了EULA的条款。

版权所有：深圳声联网科技有限公司



目 录

1. 产品介绍	5
1.1 技术背景介绍.....	5
1.2 数据产品概况.....	5
1.3 数据应用场景.....	5
1.4 数据库的典型使用方法.....	6
1.5 产品优势.....	6
2. 数据规格	7
3. 标注数据格式规范	8
3.1 【LFO 文件头：第 1 部分】.....	8
3.2 【LFO 文件头：第 2 部分】.....	8
3.3 【LFO 文件主体内容】.....	8
4. 附表：数据库概况	9



1. 产品介绍

1.1 技术背景介绍

随着以深度神经网络为代表的人工智能技术的发展，新一轮的人机交互技术热潮正在兴起。在机器视觉领域，由于海量图像和视觉场景数据库的诞生，催生了人脸识别、姿态识别、自动驾驶、无人机等领域的技术革新。目前世界上已有的典型的大规模海量图像和视觉场景数据库包括，ImageNet、MSCOCO 等业界知名的数据库。

在声音场景和声音事件的识别领域，技术的发展已经成熟，然而商业应用滞后于机器视觉领域的应用。在声音领域，目前世界上最著名的音频数据库包括：欧洲的 DCASE（声音场景和事件数据库）和谷歌的 Audio Set（包括各类层次结构化的音频分类数据）等。

在家庭环境领域，能够用来服务于 AI 应用的声音数据极为缺乏，本产品旨在填补这一空白，为全球的智能家庭环境的 AI 应用落地做出贡献。

声音场景（Acoustic Scene）指的是室内、室外、火车站、餐厅、看电影、听音乐等实际生活中人们的有声的生活场景，通过声音信号的识别来辨识这类场景，就是声音场景识别；声音事件（Acoustic Event）指的是根据短时声学特征，利用统计学习的建模方法，对不同的声源所关联的事件，进行类别的分类。例如，对哭声、咳嗽声、脚步声，能够通过声音频率特征的分布规律，进行实时的检测，发现家居环境中的突发性事件、婴幼儿的行为事件、家庭成员的异常活动等。

1.2 数据产品概况

人类婴儿由于出生时过于虚弱，不具备主动接近成人的行为能力，因此在 9 个月之前，其主要通过哭声吸引成人的注意，并向成人表达他的需求。婴幼儿语料库是按照一定采样标准采集的电子数据集，随着大数据时代的到来，语音智能产品已经渗透到移动通信、智能家居、工业生产等很多领域。语音识别技术逐渐趋于成熟，然而，语音识别产品所依赖的语音数据价值变得更加显著，语料库成为重要的基础资源。独有核心技术，让 AI 更进一步。

本语料库采集了近十六万条语音。录音采集人来自中国大陆各地，录制人数 300 多人，录制家庭 300 多家，采取 0-14 个月跟踪家庭录制，男女比例均衡，音频总时长 1275 小时，采集方式为单麦设备。每条音频单独存储为一个文件，并由专业标注人员手工进行两级标注。所有标注数据都是全检后再交付，以保证交付数据的质量。

1.3 数据应用场景

【家庭环境婴幼儿声音场景和声音事件数据库】作为丰富的数据资源，可用于哭声及家庭环境声音频率和特征的分析学习，研究声环境及其同人类活动的相互作用，更好的满足了多种场景的人工智能研发数据支撑。也为婴幼儿智能产品的创新和研发，尤其为互联网行业、母婴行业、智能家居行业提供了更详细的数据和更全面的信息支持。

1.4 数据库的典型使用方法

一种大规模的声音数据库，可以用于训练和测试，典型的使用流程如下：

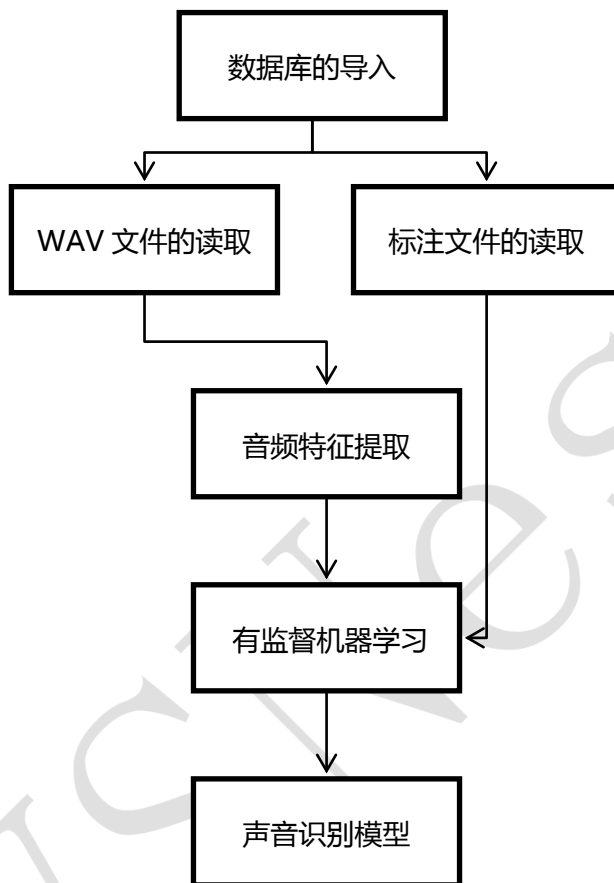


图 1 数据库用于声音识别建模流程图

其中，利用数据库中大量的音频数据文件，可以提取出高效的声学参数，以区分各类不同的声音事件，通过数据标注文件，能够获取声音类别信息，作为监督学习中的不可或缺的标签，进行支持向量机、神经网络等方式的声音建模，最终获得适合家庭实际声场环境的模型。

1.5 产品优势

- 领先的语音采集处理技术及专业标注团队；
- 同时支持科研和教学；
- 支持可视化操作；
- 支持基于机器学习的智能分析；
- 支持科研需求定制开发；
- 人工校对、多层次语音数据标注；



2. 数据规格

数据规格		
符合 LDC (Linguistic Data Consortium) 国际语义数据联盟推荐规范		
采集内容	家庭环境婴幼儿声音场景和声音事件	
音频参数	Wave PCM signed 16 bit, 16000Hz, 256kbps, 单声	
标注内容 A	哭声	婴纯强哭
		婴纯中哭
		婴纯弱哭
		婴混音哭
		婴模糊哭
		非婴哭
	笑声	
	咳嗽声	
	喷嚏声	
	尖叫声	
	婴其他声	
	公共声音	
标注内容 B	婴儿声	
	纯人声	
	纯儿歌声	
	纯环境声	
	其它公共	
采集设备	单麦设备	
产品特点	0-14 个月龄段录音, 男、女标注	
存储方式		
文件格式	Lf0 二进制数据文件	
文件时长	1275 小时	
数据脱敏		
数据敏感项	无	
脱敏操作	无	
志愿者家庭隐私免责协议	有	
产品应用		
适用领域	声信号识别, 声音事件分析, 人工智能应用	
版权归属		
版权所有者	深圳声联网科技有限公司	



3. 标注数据格式规范

3.1 【LF0 文件头：第 1 部分】

长度为 1024 字节。

- 32 字节：预留备用
- 4 字节：本 lf0 共修改了几次（最多可修改 52 次）（软件视图里会显示最后 2 次的修改人和修改时间）（Lf0 文件用记事本打开看到所有的修改记录）
- 7 字节：第 1 次修改人的：姓名；姓名最多 3 个汉字（00 结尾）（例如：张××）
- 12 字节：第 1 次修改人的：修改时间 --- yymmdd:hhmm（年/月/日/时/分）（00 结尾）（例如：170309:1542）
- 7 字节：第 2 次修改人的：姓名
- 12 字节：第 2 次修改人的：修改时间
-

3.2 【LF0 文件头：第 2 部分】

长度为 1024 字节，预留备用

3.3 【LF0 文件主体内容】

不定长度；《*.lf0》中的数据格式为 float 型（占 4 字节）。每 5 毫秒（即一帧）存储一个与声音类别相关的 data 值（与采样率无关）。

- 若当前 wav 为 270 毫秒，则《*.lf0》中存储了 $270/5 = 54$ 个数据，大小= $54*4=216$ 字节。

设：当前 5 毫秒的声音类别值=fLf0

当 fLf0 = 0 时， 存储的 data = $(-1.0e+10)$ （即小于 0 的极小值）

当 fLf0 > 0 时， 存储的 data = $\log(\text{fLf0})$

设：当前 5 毫秒存储的值 = data

当 data < 0 时， 声音类别值 fLf0= 0

当 data > 0 时， 声音类别值 fLf0= $\exp(\text{data})$



4. 附表： 数据库概况

声音数据统计				
总文件数量		文件格式:	WAV	
编码格式:	PCM, 16KHz, , 16bit, 单通道			
数据存储总量:				
总时长:				
有标注时长:		无标注时长:		
性别比例:	男:		女:	
志愿者总人数:	300			
婴儿年龄段:				
采集日期:				
标注人员数量:			标注方式:	主观听辨
标注员年龄分布		性别分布:	男:	女:
育儿专家比例:		育儿经验比例:		
数据校验方式:	多人多次标注; 随机抽查			

表 1 数据库概况