国家级大学生创新项目结题报告:《个性化语音合成于减轻老年人孤独感的研究与系统实现》

易凯
西安交通大学软件学院
yikai2015@stu.xjtu.edu.cn

姜心雨
西安交通大学公管学院
kirinna@163.com

楼艳兰
西安交通大学管理学院
banansorbananas@gmail.com

于双赫
西安交通大学软件学院
13961835623@qq.com

裘羿乐
西安交通大学软件学院
851606177qyl@gmail.com

庞建业
西安交通大学软件学院
1158230985@qq.com

Abstract

摘要是论文的高度概括，是全文的缩影，是长篇论文不可缺少的组成部分。要求用中、英文分别书写，一篇摘要不少于 400 字。

居中编排“摘要”二字（三号宋体），二字间距为两个字符。“摘要”二字下为摘要正文，每段开头空两字符，小四号。

……

关键词：XXX；XXX；XXX；XXX；XXX

摘要正文内容下，空一行，左对齐，打印“关键词”三字（五号加黑），后接冒号，其后为关键词（五号宋体）。关键词由 3 ～ 5 个组成，每一关键词之间用分号隔开，最后一个关键词后无标点符号。

1 项目依据

1.1 项目实施的目的与意义

当前，我国老年人人口基数以及人口比例不断扩大，同时独居老年人的比例日益增长，老年人的孤独感成为了一个老年人看护的难题，其不仅使老年人更易患上各种疾病，同时也阻碍了社会的和谐化发展。研究发现高龄失能老年人的孤独心理问题严重，处于中等及以上孤独水平的高龄失能老人占 77.9% **?**。孤独感是一种主观上的社交孤立状态，伴有个人知觉到的与他人隔离或不被接纳的痛苦体验 **?**

孤独感对老年人产生消极影响。对社区老年人的调查发现，高孤独感与急症入院相关联。同时孤独的老年人更可能患上高血压。外国学者对社区老年痴呆老人的血管性疾病标志物与孤独感之间的关系进行的研究表明，孤独感与糖化血红蛋白的升高有密切关系。此外，孤独感可能会使老年人发生抑郁，甚至走上自杀之路，同时也可能导致老年人认知功能、生活满意度、幸福感水平下降等。由此可见孤独感对于老年人的身心都有较大的负面影响。

当下，中国老年人对心理看护的需求基数相对而言较大。在第四次中国城乡老年人生活状况抽样调查结果显示，老年产业市场不断升温。其中，很大一部分是老年人照护服务需求持续上升。2015 年，我国城乡老年人自报需要照护服务的比例为 15.3%，比 2000 年的 6.6% 上升近 9 个百分点。类比于西方国家，由于当前中国传统文化观念的影响，老年人对心理看护的自报需求没有对疾病护理的比例高，但是无论是从基数，还是从潜在心理需求人数角度，我们有理由相信这个比例相当可观。

同时，老龄用品市场膨胀，老年用品进一步被应用。据相关资料显示，2015 年，有 5.6% 的老年人使用老龄用品，其中城镇比例为 71.8%。

我们的产品设计，是一种有效减轻独居或者子女疏于联系的老年人孤独感的新探索，我们希望通过该产品，能够为当前大量的老年人带来一些帮助，减轻他们的孤独感，同时降低由于心理原因而引起的疾病的发生比例。

通过实地的调研，较大程度上证明了我们产品本身具有较好的市场前景, 如图 **??**：

将数据绘制成饼状图如图 **??**：

上述调研结果很大程度上说明了当前老年人孤独问题比较突出，进一步证明了之前的资料调研是很有说服力的，总体而言，本产品的前景还是比较广阔的。但是由于街边调研方式的

西安市碑林区部分地区关于该产品的用户意见的调研					
时间：2017 年 3 月 10 日—2017 年 3 月 12 日					
样本容量： 80					
性别	**男**		**女**		
人数	57		23		
年龄区间	**40-50**	**50-60**	**60-70**	**70-80**	
人数	2	13	44	21	
居住方式	**独居**	**与老伴同居**	**与孩子同居**		
人数	9	46	25		
子女是否不在西安	**是**	**否**			
人数	12	68			
是否感到孤独	**很孤独**	**有些孤独**	**基本不孤独**		
人数	0	13	67		
对机器模拟孩子声音的认可度	**对增加陪伴感很有帮助**	**有一定帮助，但帮助不大**	**基本没什么用**		
人数	32	35	13		
如果制作成嵌入式设备，您的可接受价格区间	**200 以下**	**200-300**	**300-500**	**500-1000**	**1000 以上**
	8	42	21	9	2

Figure 1

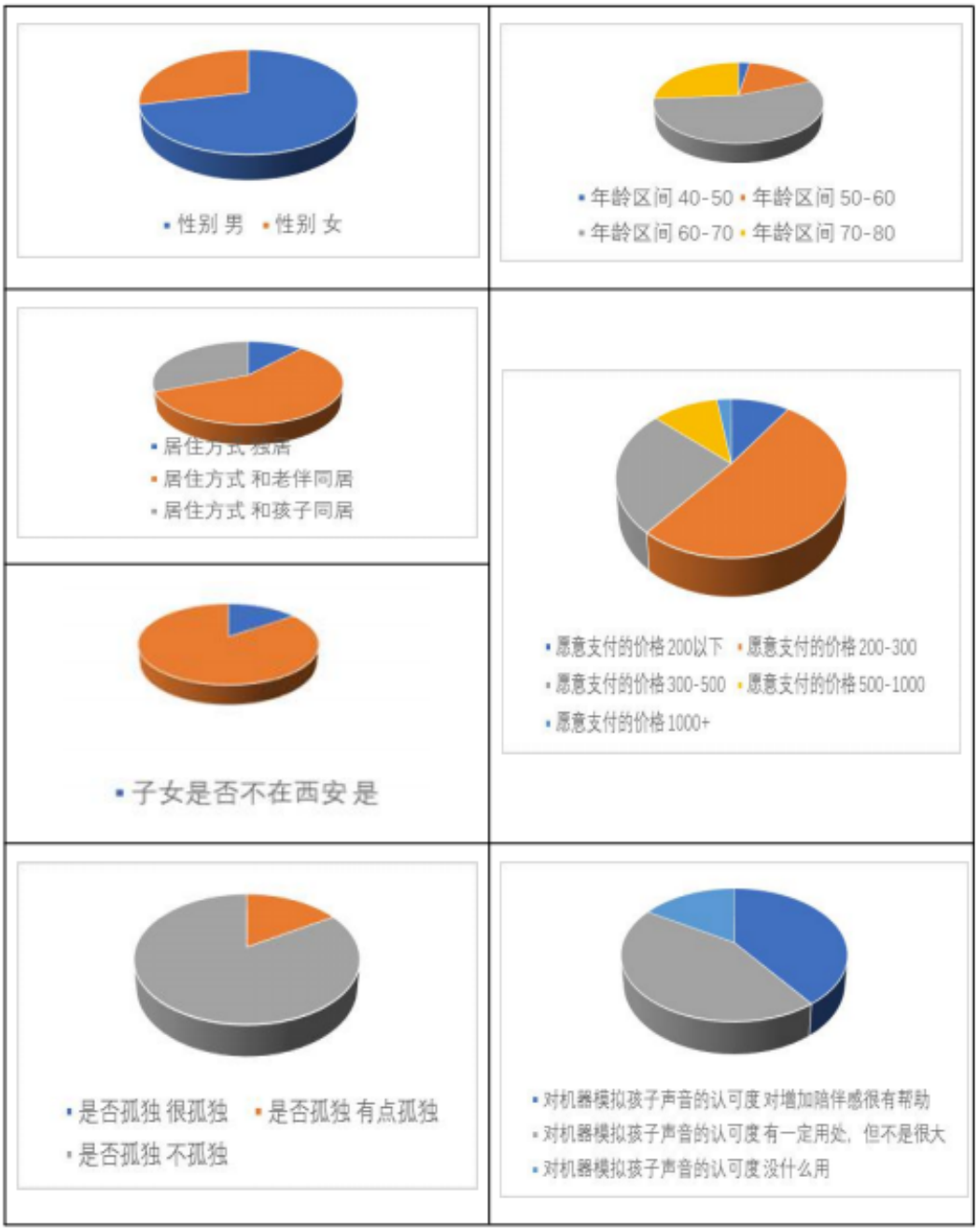
性别 男
性别 女
年龄区间 40-50
年龄区间 50-60
年龄区间 60-70
年龄区间 70-80
居住方式 和老伴同居
居住方式 和孩子同居
愿意支付的价格 200以下
愿意支付的价格 200-300
愿意支付的价格 300-500
愿意支付的价格 500-1000
愿意支付的价格 1000+
子女是否不在西安 是
是否孤独 很孤独
是否孤独 有点孤独
是否孤独 不孤独
对机器模拟孩子声音的认可度对增加陪伴感很有帮助
对机器模拟孩子声音的认可度有一定用处，但不是很大
对机器模拟孩子声音的认可度没什么用

Figure 2

局限 (给路人带来的陌生感) 以及调研地域的局限 (西安较为发达，留守老人相对较少，根据资料分析农村留守老人较城区多数倍，因此潜在需求比调研需求会更大)。综上所述，在中国而言，解决老年人孤独感问题产品的需求度很高。

采用基于 Android 系统的 app 实现相关产品是我们的第一步，在后期会将整个平台调整后移植到嵌入式设备。从初步调研中愿意支付的价格来看，普遍接受的价格相对而言较低，因此我们后期使用嵌入式设备实现，需要严格地控制成本。

2 研究内容

本项目研究内容主要是解决独居老人增多的背景下老年人孤独感的问题，采用了个性化语音合成、深度学习等方法，通过 Android 系统以 app 的形式进行实现。后期，根据实际需要，会将该产品移植到嵌入式设备之上。

项目总体设计上分为四个部分，第一个部分是说话者识别，第二个部分是音素建模，第三个部分是目标对象的语音泛化，第四个部分是语音实时交互。

用户层面的总体流程如下图 **??**:

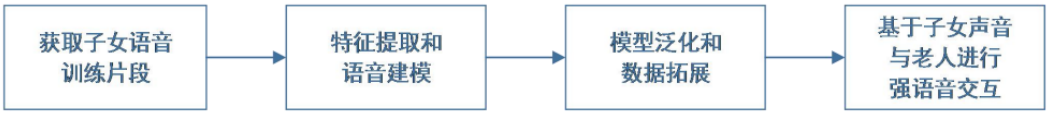

Figure 3

老年人的情感上的空缺往往只是缺少了一个值得信赖的说话对象，而我们的系统不同于传统的系统，要么是做老年人的健康管家，要么就是做老年人没有人情味的指导型机器人，而是以学习子女语音的方式将该应用真正地将这种为老人的心中带来慰藉。

考虑到使用深度学习的相关方法，以及降低产品设计的时间成本，我们决定使用 Android 作为方法实现的载体，也就是以开发 app 的形式完成整个项目的设计。另外，根据之后的实际情况，我们会考虑采用高度集成化的嵌入式设备进行实现。我们会设计两种实现方式，一种是轻便易携式的手环式设备，一种是固定摆放式的小型机器人设备。关于详细的使用方式，设备需求分析以及市场推广方案，将会在“项目研究进展安排—关于嵌入式设备的运用以及市场投放的想法”中进行具体阐述。

3 国内外研究现状和发展动态

3.1 老年人看护研究现状

目前老年人看护已经不仅仅限于人工护理，其与科技方面的融合有进一步发展的态势。老年人手机、老年人腕表、老年人看护机器人等层出不穷，但是在科技产品层面大多是对脉搏等基本情况的检测、定位、紧急语音救助、对话等。

比较典型性的相关产品主要有以下几个：

3.1.1 1.

养老院看护机器人“佐拉”，身高 57cm，可以进行简单的运动、跳舞、读书、讲笑话。此外，其配有传统模式下的语音合成功能，可以识别 19 种语言，护工可以用平板电脑给机器人输入预先编写好的程序，能与老年人进行一对一的交流;

3.1.2 2.

英国的 care-o-bot 机器人会做多种家务，可以通过在 LCD 屏幕上显示不同的表情来慰藉老年人；

3.1.3 3.

德国莱尔克斯机器人研究院研制出一种可为老人及儿童提供看护服务的智能机器人。它能与空巢老人对话，开开玩笑，讲讲新奇的事情；能陪伴儿童玩玩游戏、讲讲故事。

国内也有相关类似的产品，其关于老年人看护方面的共性主要是相关的产品要么是专用化的嵌入式设备实现诸如量取血压、脉搏等特定功能，要么就是以机器人的方式实现与老年人的对话，采取换表情或者是讲笑话等方式来逗乐老人。

但是注意到当前产品很少有关注老年人心理健康的，且存在功能单一、价格较高等问题。在国内市场上这种现象更为明显。因此，用于解决老年人孤独感、同时集成多种所需的功能、价格适中的产品是有较大市场需求的。

3.2 个性化语音合成技术的发展现状

个性化语音合成, 就是通过身边的一些录音设备，录取某个人的某些语音片段后，采用TTS(Text to Speech) 语音技术，合成出像某个特定人的说话语音、说话方式和说话情感。

由于某些技术的局限性以及音素特征学习模型的不稳定性，教会机器高仿真度地模仿某特定对象的声音，目前还出于进一步探索阶段。同时也有通过小规模数据来进行模型训练的趋势。

个性化语音合成技术，涉及到语音学方面的很多新技术。包括：音频谱特征转换技术、韵律特征转换技术和个性化参数自适应技术等。在国际上做得相对较好的公认的是讯飞，其主要是以识别的精确度为入手点进行的语音合成，而“个性化”实则不太明显。我们想要利用其目前的研究优势，然后加以改进，应用于我们的产品之中。

另外，值得注意的是，或许是由于之前相关技术的局限性，个性化语音合成技术往往应用于物联网以及智能家居，而我们将其应用于老年人心里看护实际上是有极大创新性意义的一步。

3.3 缓解老年人孤独问题的相关产品

当前，国外的相关企业更为重视老年人的孤独问题：比如数字家庭晚餐，也就是利用屏幕和语音随时和子女进行沟通交流。

此外，“Elli.Q”是当前最为先进的“社会陪伴机器人”之一。其可以通过不同的语调、灯光及动作来表达情绪，还可以根据“主人”的喜好，从而在适当的时间提出建议，比如让主人读书、出去走走、给家人打电话等。据调研，目前国内尚没有相关成熟的产品。

以上都是缓解老年人孤独的一种实现探索方式，国内随着老龄化以及独居老人比例的增长，相信其在国内的需求量会进一步提升。因此，研发缓解老年人孤独问题的相关产品在国内是很有市场前景的。

3.4 孤独感的本质与普遍性

关于孤独感缺失的本质主要有两种观点：缺失观和认知观。缺失观认为个人社会关系网中某种关系的缺乏使归属得不到满足。认知观则表现为对现存人际关系的不满意。我们的产

品通过有效地扩大老年人的归属感，预期能够有效地降低老年人因为某种关系缺乏而产生的归属感，给予老年人心灵上的寄托。

国外较早的研究表明 1100 多万 50 岁以上的美国人感到孤单。而国内研究者对安徽农村老人的研究发现，78.1% 的老年人有中等到严重的孤独感。虽然老年人的孤独感感知可能与文化背景、调查时间、研究样本有关，但实际上是一个不容小觑的问题。

一般地，相关研究表明，体弱多病的老年人更容易感到孤独，而通过前面的分析我们也知道，孤独感的长期存在更容易使老年人身心患病。此外，视力和听力的下降也会使老人产生孤独感，因此我们的 Android 产品充分考虑到了这一点，可以根据用户的需求通过语音的方式进行音量的控制。

3.5 本项目的市场空间量化

汉语语音合成技术一方面带动了语音应用的发展，促进信息大众化和社会化的进步，另一方面又可获得巨大的经济利润，创造出连续发展的语音市场空间，推动汉语语音合成技术的产业化进程，为该技术占领世界市场奠定基础。据专家对未来国内市场的预测，在未来 2～3 年内，语音合成系统的配备率在个人电脑中将达到 25%～30%以上，语音合成系统的个人用户市场潜力为 18 亿～20 亿元人民币，而应用于行业的电话语音查询系统的市场份额将至少在 30 亿～50 亿元人民币以上。

比较典型的有:

3.5.1 畅言 **2000**

目前，畅言 2000 的研制成功已引起 Dell、联想、同创、实达、上海广电等国内外厂商的极大兴趣，这些厂商已准备在其品牌电脑中 OEM 捆绑销售畅言 2000。根据对市场调研和产品 OEM 销售协议，预计该产品年销售总量可达每年 40 万套以上，2000 年销售额指标为 2000 万元，利润 1000 万元。年销售收入总额最终将稳定在 4000 万元以上，实现年利税 2500 万元以上

3.5.2 博思智能中文平台

该平台是在以汉语语音合成为核心的语音平台上，将汉语语音识别、汉语语音合成、手写识别、扫描输入、机器翻译、汉字输入法等最先进的中文信息处理技术进行集成，制定统一的接口规范和接口标准，将我国在中文信息处理方面的各局部优势凝聚成整体优势，具有我国自主知识产权的智能化中文信息平台。该平台不但包含了普通电脑用户所关心的各种中文信息处理应用模块，还支持软件开发厂商和编程人员的二次应用开发。根据调研，市场对该平台的年需求可达 50 万套以上，预计年销售量在 30 万套，年销售收入 3500 万元，年利税总额 200 万元，投资利润率为 204%，投资回收期 0.7 年，总资产报酬率为 62%。

3.5.3 声讯平台改造

如前所述，汉语语音合成技术在声讯服务的电话查询系统中有得天独厚的技术优势。目前，各行业声讯服务系统正在实施技术改造或系统换新，例如现在全国正在进行的 160/168 二次改造工程。160/168 二次改造最主要的经济效益来源体现在以提供技术而获得的声讯增值服务中。目前，全国 160/168 声讯年收入约 35 亿元人民币，采用语音合成技术后，其增值服务按 10%计算（根据电信方面的有关专家估计，在未来 1 ～ 2 年内将至少达到 10%～20%），则每年将据此获得信息费 3.5 亿元人民币。预计 5 年以内，增值服务将达到声讯服务的 80%，整个声讯收入将增长到目前的 5 倍以上。

4 创新点与项目特色

该项目的创新点与项目特色有以下几个方面的内容:

4.1 项目的出发点是基于老人的孤独问题

当前，目标人群为老年人的国内相关产品，主要针对的问题要么是老年人的疾病监测与简单疾病的预防；要么是通过语音交互的方式对老年人生活进行简单沟通交流 (交互语言不够智能化，同时语音不够自然)，但是考虑到当前国内老年人比例逐渐增多，同时独居老人的比例也在不断地增大，老年人的孤独问题呈现出一种很严重的态势。通过之前的研究分析，老年人心理上的孤单，显著提升了老年人的疾病的发生比率，而同时疾病又进一步增加了孤独感的比例，这是一种恶性循环，同时也说明了关注老年人精神健康的重要性。

4.2 产品使用了当前最先进的相关技术

在 Speaker Recognition 研究领域，当前最前沿同时也是效果最好的当属于微软公司开源的 Speaker Recognition API；至于在语音合成领域，当前国内外公认的较好的产品是讯飞语音合成相关接口。这些前沿的最新技术，我们都将深入学习，然后将其修正，适应性地使用于我们的产品之中。

4.3 一种缓解老年人孤独感的新方式的探索

我们沿着项目的出发点，提出了泛化的强语音交互模型，并且将尝试通过 Android 进行实现。据我们的查阅的资料显示，我们的语音泛化的方法在当前还没有被应用于实际领域，相对而言难度较大，我们独辟蹊径，想要通过学习子女的语音音素特征，然后将其进行泛化到整个数据集上，实现以子女的声音来与老人交流的新方式。同时，为了有效地规避风险，我们设置了不同的仿真度。

5 技术路线

以下从技术角度对该产品进行相关说明：

5.1 系统概述

我们的产品构思将会分为四个部分进行实现，分别为说话者识别 (speaker recognition)、音素建模 (phoneme modeling)、模型泛化 (model generalization) 和强语音交互 (strong voice interaction)。

强语音交互即不直接建立问题与回答之间的映射关系，引入情感化的分析，如问题是今天几号？系统不会直接告诉你今天几号，而是会告诉你：今天是母亲节。强语音交互是一种更加智能化、人性化的语音交互方式。在本产品中使用的数据泛化，将数据集中的特征进行抽象，并且模拟用户的语音与其他用户进行交流，增强交互体验，这是目前手机语音应用市场上几乎没有涉及的，增强了用户间交流的兴趣与听觉感官体验。目前，该产品主要针对于老

人孤独感问题，通过调研分析，其也将极大提升心理体验。后期我们会考虑引入嵌入式设备，为老年人带来更多的便利以及更好的交互体验。

5.2 设计约束

本产品中数据集收集与测试，是在用户已知的情况下打开相应的权限功能来进行的，保证了用户数据自身的安全。

5.2.1 软件

开发环境 ：Ubuntu + win7 + Android Studio 2.3 + VS Code.

运行环境 ：考虑到当前大部分 Android 设备的版本都大于 5.0，因此我们的 app 向下兼容至 Android API21(Android5.0) 的设备。

硬件环境 ：根据产品实际后期需要进一步确定。

5.2.2 接口

出于安全性以及用户体验的考虑，会有摄像头、话筒使用、读取联系人等请求。

5.2.3 用户界面

操作简化的符合老年人实际需要的用户界面，同时该 app 与用户交互的主要方式是语音。

5.2.4 产品质量

便捷易用，安全可靠，兼容性好，可扩展性强。

5.3 软件的总体结构图

从上述的总体设计流程图可以看到，我们的项目产品本身分为两个部分。一部分是训练部分，另外一部分是应用部分。

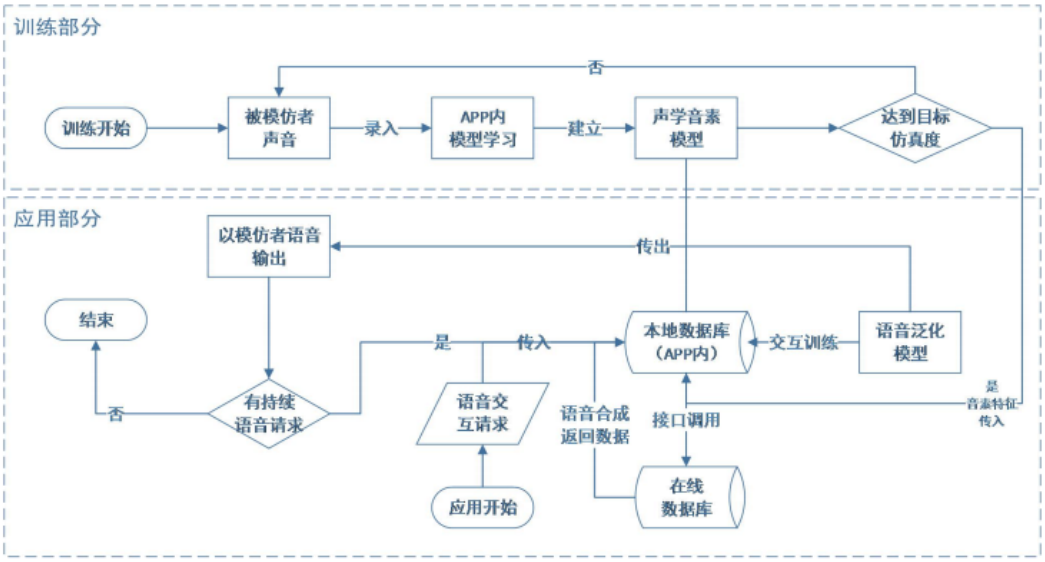

Figure 4

5.4 模块设计

5.4.1 说话者识别

该部分的内容主要是通过我们的 app 产品识别说话者的有效声音信息，通过训练提取出其特征。

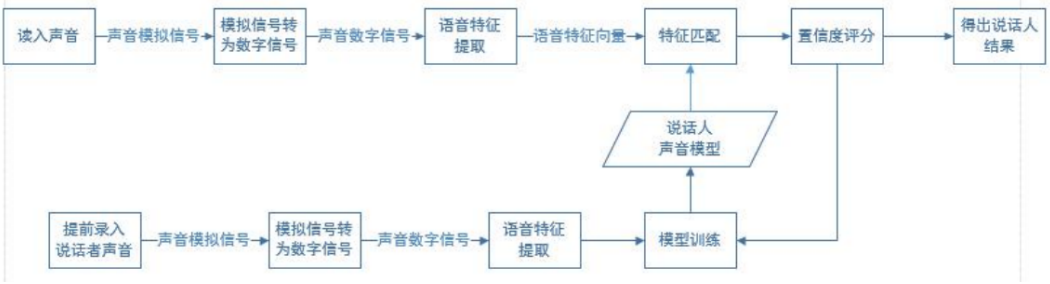

Figure 5

通过上述模型，得到较为稳定的说话者的音素特征，即指定识别人的语音。

5.4.2 音素建模

线性预编码 (LPC) 预处理器从 WAV 音源的原始 PCM 数据中计算出 LPC 系数，将其转为 LPC 向量，LPC 向量可提高音频数据的稳定性和抗干扰能力，更有利于语音识别工作。

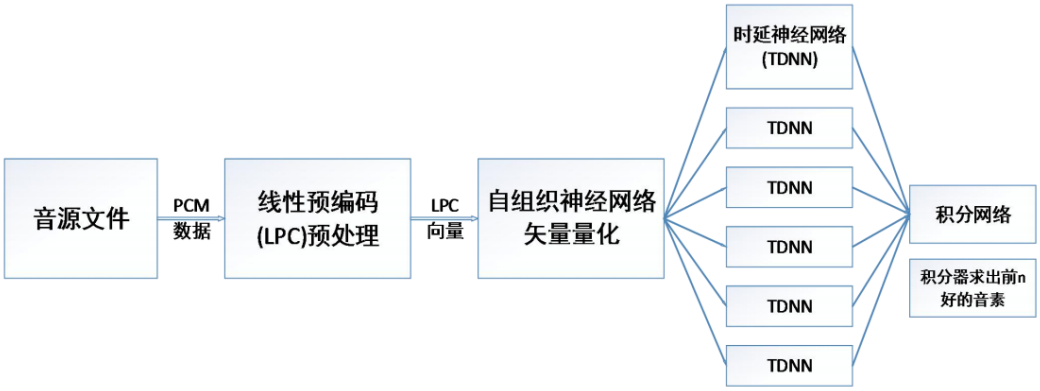

Figure 6

采取自组织神经网络 (SOM) 对 LPC 向量进行量化以降低 TDNN 处理时的复杂度,SOM 模型带来对噪音更好的健壮度。

对于每一类音素，使用一个 TDNN 来进行识别。

最后将所有 TDNN 的输出结果汇总到一个积分网络中的多层感知器中进行处理。

5.4.3 模型泛化

此部分是该项目的重点，也是该项目的最大创新点。

模型泛化即是将训练好了的子女的语音进行泛化，然后根据老人实际的语音交互需要，通过泛化模型生成语音来进行后续的强语音交互实现。

模型泛化的实现基本思路是：

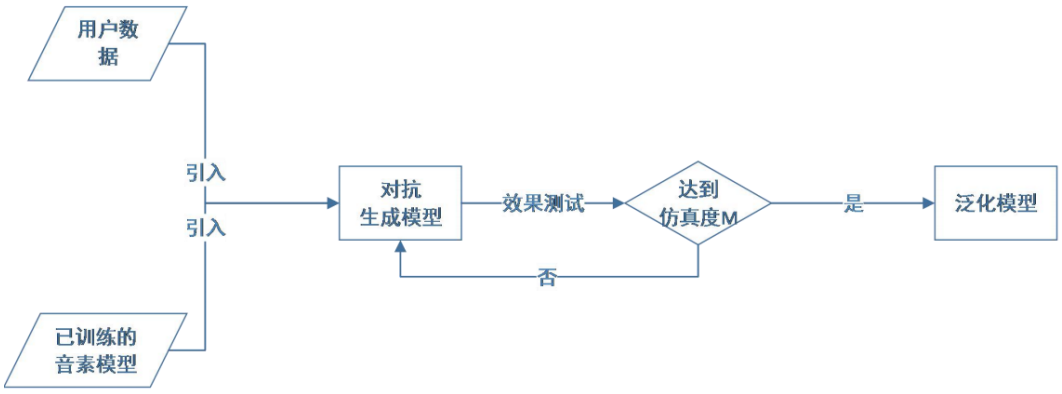

Figure 7

对抗生成模型作为当前的热点内容，将其引入到语音的模型泛化尚无先例，具有创新性的一步。同时，该部分设计流程需要将数据的交叉验证，迁移学习等引入对抗生成模型的训练。

5.4.4 强语音交互

该部分主要是进行设备和用户之间的在线与离线式即时交互，用户与用户间的在线交互。

该强交互实现是基于讯飞语音开放平台的，讯飞语音是以语音交互为核心的 AI 平台，由于讯飞语音精度高，在市场上被广泛使用，因此我们决定进行学习应用。

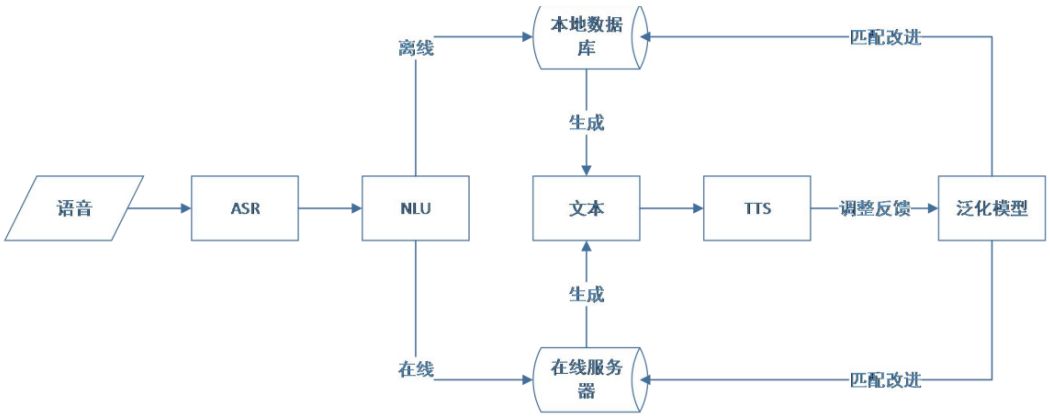

Figure 8

语音由用户输入后，做语音识别（ASR），自然语言理解，并且通过在线与离线的方式进行生成，离线方式用本地的数据库，在线方式用服务器的数据，并与之前的泛化数据进行模式匹配，并得到改进的需要的反馈语言语句，改进数据库中的数据并生成相应的语言文件（context），并且输出语音（TTS），用这一状态下的语音输出参与到泛化数据的训练中，并且给予下一状态下的语音交互的一个反馈，不断加强改进强交互的状态结果，通过这种训练与改进加强与用户间的交流方式。

5.5 数据库设计概述

由于使用了微软 Speaker Recognition API 和讯飞语音开放平台，因此数据库将会分为两部分进行实现。

一部分是本地数据库，另一部分是在线服务器。其中本地数据库主要存储的是模仿者的音素信息以及需要进行交互的未训练信息；另外一个是在线服务器，其中主要是调用接口的相关测试平台，其上有大量的用户语音数据信息，进行模式匹配后将未训练待交互信息传

入在线服务器，得到训练后将数据传回本地数据库，通过强交互语音模型结合模仿者的音素特征做出语音的反馈。

5.6 安全性设计说明

关于安全性主要是为了产品需要获取用户相关信息的综合考量，详情可以参考设计约束—接口中的部分内容。

5.7 综合考虑

稳定性和可扩展性

该模型在对待模仿者音素特征进行训练的过程中，仿真度会随着训练次数的增加不断增强，但达到良好状态的训练时间以及次数需要严格控制在用户可以接受的范围之内。

该 app 支持符合用户实际需要以及可实现的功能扩展。

性能分析

满意的结果是对于用户的语音交互请求可以即时地进行语音反馈。根据用户实际的网络条件，可以允许有一定程度的反馈延时。

复用与移植

A. 简单的交流方式模型可复用

B. 测试的数据集可复用

C. 训练泛化的技术方法可复用

防错和出错处理

A. 对目标声音过小的处理

B. 方言问题的特征泛化

C. 交流方式与语气的问题需要注意

6 拟解决的问题

1. 如何将现有的最为先进的技术运用于当前的 Android 应用开发之中；

2. 如何设计一种适用于语音模型泛化的新方法，并且顺利地在 Android 系统上进行实现；

3. 由于机器学习往往是进行大规模数据的处理，如何进行机器学习算法的卷积压缩与优化，使其能够高效的在手机端能够实现；

7 预期成果

完成符合项目内容的 app，实现的主要内容是说话者识别、音素建模、语音泛化、强语音交互四个部分，同时将机器学习模型的相关算法进行压缩优化，在该应用上进行高效地实现。考虑该产品在嵌入式设备上的实现并制定相应规划。

?