# 1.研究目的、意义和必要性

近年来，随着深度学习技术在语音、声纹识别领域的发展，纯净环境下的识别性能不断提高。2016年微软的一篇论文中宣称他们的语音识别系统在NIST 2000 Switchboard 任务上实现了和专业速录员相当甚至更低的词错率（WER），随后在2017年，IBM又宣称在同等任务上达到了更好的性能。这些语音任务中的语料具有语音信号质量好，环境简单，没有严重干扰和噪声的特点。

然而随着智能语音在工业、民用领域的应用大量普及，实际部署的环境复杂而多样，噪声和干扰的种类繁多，如：周期性噪声、脉冲噪声、宽带噪声和语音干扰噪声等，这些干扰和噪声的发音原理各不相同，各有特点。同样在安防领域，智能语音系统近年也得到了有效的应用，有很多非常规的拾音手段虽然可以获得声音信号，但由于应用场合多样且常为增强隐蔽性而牺牲清晰度的考虑，获得的声音信号通常含有大量噪音干扰。

复杂环境中，语音识别和声纹识别如何保持很高的识别率，是尚未解决好的难题。

本课题拟在智能语音系统典型的部署场景：智能终端，将研究一种语音识别、声纹识别系统，使其在复杂环境下具有识别率高、鲁棒性强的特性。相比于一般的声音处理系统，本系统在保证稳定性的前提下，突破各类复杂环境下使用的困难，且具有良好的可移植性，可适用于多种智能终端平台。

## 国内外现状和技术发展趋势

我国语音识别研究工作起步于五十年代，但近年来发展很快。研究水平也从[实验室](http://www.chinabgao.com/k/shiyanshi.html)逐步走向实用。从1987年开始执行国家863计划后，国家863智能[计算机](http://www.chinabgao.com/k/jisuanji.html)专家组为语音识别技术研究专门立项，每两年滚动一次。我国语音识别技术的研究水平已经基本上与国外同步，并达到国际先进水平。

尽管如此，在复杂环境下的语音识别方面，早期使用基于模板匹配的模式识别方法有极大的局限性。而到20世纪80年代，GMM-HMM被应用于语音识别中并取得重大成功。在21世纪以来，由于硬件技术的突飞猛进使得机器学习的方法成为语音识别的主流。与传统的GMM-HMM声学模型相比，DNN-HMM声学模型性能更佳。此后又衍生出诸如RNN、LSTM以及基于CTC损失函数等更优秀的模型，但也存在一些问题，最关键的是在复杂环境下噪音的影响极大限制了语音识别系统的识别率，近期研究者采用自适应算法训练鲁棒声学模型，如Gales等利用最大似然线性回归的模型自适应算法、Tran等基于分解的线性输入网络对深度神经网络的声学模型进行噪声环境自适应，对降低噪音的影响达到了很好的效果。

在复杂环境下的声纹识别方面，声纹的概念最早是由bell实验室的L.G. Kesta通过观察语谱图所提出，并由S.Pruzansky通过基于模式匹配的识别方法实现，开启了使用声音辨别身份的先河。同语音识别一样，在20世界80年代之后至今，声纹识别的研究重点转移至基于统计学和机器学习的理论中，但同样存在复杂环境下噪音影响的限制，研究者也提出了很多优化噪音影响的方法，如Abe等提出了将经典谱减法和噪声估计相结合的方法，Vincent等使用基于深度降噪自适应码模型建立带噪语音和干净语音特征间的映射关系进行多目标联合训练。最近主流研究趋势是利用深度学习智能设计得到更好的特征,通过深度学习得到的抽象特征是高维特征在低维的凝练描述，具有更加抽象、鲁棒的效果，能够很好的适应复杂环境下系统的稳定运行。

## 市场需求分析

该系统在实际上解决的是智能语音系统在各个行业应用的痛点，我们针对两个领域，安防和民用领域来展开对该项目产品化后市场前景的分析。

1. 安防领域

目前，我司的主营产品为安防的特种设备，在实际应用中，必须采用非常规的手段获取各种语音讯息，然后对获得的语音讯息进行智能分析处理，如语音识别，声纹识别等，实用性受到了环境的制约，在复杂环境下效果不理想。该项目解决此痛点后，使得该设备的市场前景更为良好。主要体现在以下几点：解决了语音讯息获取的时效性，进一步提高工作效率；和我司的主营设备完美融合，实用性更强；本设备在安防的各个方面均可部署，如国家安全，保密，公共安全等等方面，市场巨大。

1. 民用领域

据《2017年中国人工智能行业分析（智能语音应用篇）》分析显示，智能终端已经渗入到各个领域，代表领域包括金融、医疗、教育、呼叫中心、智能家居、可穿戴设备、无人驾驶、虚拟助理、家庭机器人等。

在如此大的市场需求下，国内多家企业在语音识别技术上投入了大量的精力、物力、财力。虽然，语音识别技术识别的真实应用体验不断提升，但是实际应用打大折扣。在真实使用场景中，如因空间距离、背景噪音、其他人声干扰、回声、方言、口音等问题，识别准确率大打折扣。目前远场识别的错误率是近场识别的两倍左右，所以解决远场以及强噪声干扰情况的语音识别是有待进一步研究的问题。

而该项目“抗扰抗噪语音识别技术”是针对这一问题进行深入研究，使其根据智能语音系统中的声纹识别辨别用户的信息，并利用语音识别进行的智能控制。使得体验更加贴近生活、更加实用能够满足更多用户多样化需求，从而也进一步提升用户黏性，带动市场需求扩大，促进语音智能终端市场良性发展。

# 2. 研究目标

总体目标

1）研究一种语音识别、声纹识别系统，使其在复杂环境下识别率高、鲁棒性强，且具有良好的可移植性，可适用于多种智能终端平台。

2）申请发明专利4项,申请软件著作权2项；晋升职称3人；样品样机3套。

3）针对我司在安防领域的市场布局，在现有拾音设备上，嵌入上述的1）系统，开发出一套智能拾音系统，具有抗干扰、抗噪声能力强的优点。

4）项目完成时，具备申报“杀手锏产品”的条件。

产品技术性能指标

|  |  |
| --- | --- |
| 指标项 | 性能要求 |
| 声纹识别等错误率EER(>3s) | ≤4% |
| 声纹识别Cavg(>3s) | ≤0.05 |
| 语音识别词错误率WER | ≤4% |
| 语音识别响应时间(离线模式) | ≤10s |
| 语音识别响应时间(云端模式) | ≤10s |

# 3. 主要研究开发内容

本课题是基于复杂环境下的语音、声纹识别技术的研究，并开发一套灵活用于各种智能终端的系统。一方面是研究语音增强技术，尽可能的消除噪声和干扰信号，并分离出感兴趣的语音信号。另一方面，需研究在有残余噪声情况的如何提高语音识别和声纹识别的鲁棒性。最后，需考虑研究针对离线和云端情况对智能终端的部署。

## 1）语音增强

语音增强是一个前端模块，目的是分离出有用信号，去除噪声和干扰。本研究点主要分为以下内容：

* 基于信号处理的语音增强算法
* 基于深度学习的语音增强算法
* 基于迁移学习以及环境自适应的多任务学习算法

## 2）鲁棒的语音识别的研究

语音识别的以实现语音的文本转写为目标。通过语音增强处理后得到的特征建立声学模型与语言模型，能够在复杂环境下将语音信号转写为文本信息，提高识别率。研究内容如下：

* 任务导向的声学特征学习和提取
* 基于时序建模和attention机制的语音识别系统

## 3）声纹识别

为了在复杂环境下实现更高的准确率，本项目将从以下几点进行研究：

* 基于深度学习的特征提取
* 基于Triplet Loss等损失函数选取策略优化

## 4）在智能终端的部署

在智能终端的部署包含了云端模式与离线模式。对于我司在安防领域的终端产品采用离线模式，具有安全性高的优点。而在工业与民用领域上则采用云端模式的智能语音系统，相比离线模式有更优秀的识别性能。本要点具体将进行以下研究：

* 软件可移植性
* 云端模式部署方案
* 离线模式部署方案

# 4.项目创新点及拟解决的关键技术问题

## 1）项目创新点

关键技术创新：

* 在适应复杂环境方面：
* 基于小波变换理论，提出改进的语音增强算法。
* 基于噪声注入的生成对抗网络深度学习算法，扩充各种环境条件的训练数据。
* 基于迁移学习等技术，设计任务导向的特征提取算法，增强系统鲁棒性。
* 在语音识别方面：
* 基于CTC方法和attention机制改进了传统LSTM神经网络结构，研究出一种新型LSTM网络结构。
* 在声纹识别方面：
* 结合端到端的CNN、DNN、RNN-LSTM, 改进了speaker embedding方法。
* 结合当前深度学习损失函数，研究出一种新的带有区分性的损失函数。

系统创新：

* 通过分离与系统平台相关的代码，减少与系统平台的耦合度，从而拥有良好的可移植性。
* 结合知识蒸馏、低秩矩阵分解以及稀疏化等方法的优点，研究针对神经网络模型的高效压缩方法。
* 将离线模式的智能语音系统应用到我司的安防终端设备。
* 将云端模式智能语音系统推广到工业领域在内的多行业中，如营业厅的智能终端、智能家居等市场。

## 2）拟解决的技术问题

* 提高语音增强模块对多噪音源下信号特征的提取效果

本课题拟通过两个方面解决该问题。首先利用信号处理方法直接在原始语音信号上尽可能地过滤掉简单噪音；其次再利用深度学习能够对待识别信息的特征进行深度描述的特性，将一些对于识别语音、声纹无关的不易处理的噪音进行剔除。

* 语音、声纹识别系统训练收敛慢和识别性能问题

本课题拟用一种新型LSTM网络结构，根据任务导向调整结构，搭配不同学习策略实现快速收敛并提高识别性能。

* 可移植性问题和终端设备部署问题

为使语音、声纹识别系统能在不同智能终端中运行，使之不依赖于具体的运行环境，需要将与平台相关的代码进行隔离。

终端设备的运算能力和存储空间对比系统服务器都是受限的，为实现识别系统在终端上运行的目标，必须要解决部署问题，本课题拟利用深度学习领域对于模型进行知识蒸馏、神经网络模型中LSTM层低秩矩阵分解以及稀疏化方法实现模型压缩参数，同时解决算力较低，和空间受限的问题。

# 5. 研究方法和工艺技术路线

研究的总体技术路线总结如图1：

语音增强将采用基于信号处理、深度学习的方法，以及基于迁移学习的环境自适应的多任务学习算法。语音识别模块研究任务导向的声学特征学习和提取，并引入时序建模和attention机制。声纹识别模块主要研究结合深度学习的端到端系统和speaker embedding系统。

图表 1

图 1研究的总体技术路线图

语音增强

语音识别

声纹识别

训练

测试

实验性能评估

在线性能评估

训练

语

料

采

集

线上语

料

采

集

参数调整

语料的准备也是本课题研究的重要基础之一。对于不同应用场景，需要采集足够时长的语料。然后把采集的语料分为训练和测试集不重叠的两部分。

对于三个子模块，先选取初始的模型超参数（神经网络的规模，GMM的混合数等等），然后用训练集语料各自进行训练。为了避免过拟合的问题，一般将训练集再分出一部分作验证集。根据目标函数在验证集上的变化来控制训练的迭代次数，防止过拟合。在训练完成后，可用测试集来评估子识别系统的性能。通过调整尝试的初始参数和选取的模型框架，不断地优化。

部署方面，为使系统可以适用在多种智能终端，需要封装与终端硬件无关的核心代码并提供针对各种平台的接口，根据终端是否可以联网，设计离线系统和在线云端系统两套解决方案。

实时方面，在上线后，需评估性能，并可在线上累计语料，通过不断更新语料库优化系统性能。

下面分别针对语音增强、语音识别、声纹识别和智能终端部署方案设计分别阐述其具体研究方法。

## 1）语音增强模块

* **基于信号处理的语音增强算法**

目前传统的基于信号处理的语音增强算法由于其功耗低、计算速度快以及适应性强等优点，因而普遍用做很多智能硬件的前端语音处理方法。传统上的语音增强算法包括谱减法、最小均方误差（MMSE）和维纳滤波器等，其在有加性噪声的条件下拥有不错的效果。其次由于在日常交互环境中会有很多的混响产生，针对混响问题，也有一些算法被提出。这些算法虽然对很多平稳噪声的抑制效果比较显著，但是对于一些非平稳噪声的却不能得到一个令人满意的效果，基于此，本项目将研究一种环境自适应的语音增强系统以使得前端语音增强得到更加鲁棒的效果。本项目中将研究使用小波变换以及线性滤波器等方法，结合不同方法的优点进行语音前端预处理以达到一个令人满意的鲁棒性。



图 2 小波分析

小波分析是一种有效的去噪技术，它在时域和频域都具有良好的局部化性质，既能够把握信号的整体，同时又能够提取局部信息。因而成为分析非平稳语音信号的工具。带噪语音信号经小波分解后,其中有效语音信号集中在大尺度低频小波系数上,而噪声的能量则分布于全部尺度空间,因而噪声的小波系数通常小于语音的小波系数。据此特性，设定合适的阈值,将噪声的小波系数设置为0,保留语音的系数,即可达到有效地去除语音信号中的噪声的目的。

* **基于深度学习的语音增强算法**

随着深度学习的发展，很多相应算法也随之应用到了语音增强技术中来。基于深度学习的语音增强技术可以获得比传统的基于信号处理的语音增强技术更好的效果，基于深度学习的方法大致可以分为基于映射的方法和基于掩蔽的方法，这两种方法都是通过使用神经网络来学习一个含噪语音到干净语音的非线性函数，以此来达到去噪或者去混响的目的。先前的研究都是基于能量谱或者振幅谱来进行增强，并没有考虑语音相位的作用，本项目将结合振幅谱和相位谱来进行整个系统的训练，这也将在很大程度上提高增强的效果。其次由于传统的全连接的网络并不能充分发挥神经网络的潜能，本项目中将会重新设计一种基于生成对抗网络的框架，这种框架能够进一步提升系统的性能。

另外，人为设计有噪语料，可以提高神经网络的泛化能力。



图 3 噪声注入的DNN训练

虽然深度学习在语音增强上会产生一个比信号处理方法更好的效果，但是单单使用深度学习的算法来进行语音增强很多时候却不能处理一些使用信号处理方法能够进行去除的噪音混响等干扰，所以本项目将结合这两种方法的优点，深层次的探索两种方法的互补性。我们先使用信号处理的方法来进行一个语音的预处理，然后将处理后的语音作为神经网络的输入，所得到结果对比单单使用深度学习或者信号处理的方法性能会有较为明显的提升。

* **基于迁移学习以及环境自适应的多任务学习算法**

迁移学习是指一个学习算法可以利用不同学习任务之间的共性来共享统计的优点和在任务间迁移知识，其目标是用从一个环境中学到的知识来帮助新环境中的学习任务。因此，迁移学习不会像传统机器学习那样假设训练数据与测试数据服从相同的数据分布，本项目中将使用迁移学习的方法进行无监督数据标注，这可以大大节约语料成本。

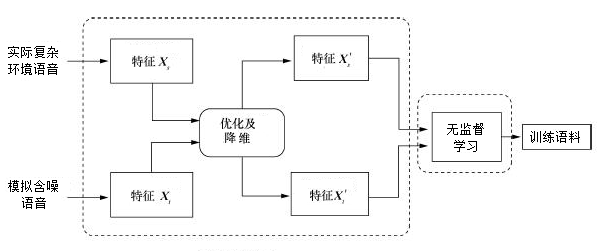


图 4 迁移学习算法

在多任务深度网络中，低层次信息的共享有助于减少计算量，同时共享表示层可以使得几个有共性的任务更好地结合相关性信息，这些任务都利用同样的语音数据，因此它们之间相互依存、相互促进，任务特定层则可以单独建模任务特定的信息，实现共享信息和任务特异性的统一。

基于此，本项目首先通过将多种口音的干净语音语料与实际环境中测到的空间脉冲响应做卷积得到模拟的带噪语音。然后在实际的复杂声学环境下（例如混响环境下）收录语音语料，以便获得充分的训练数据。并在这些模拟和实际的带噪语音数据的基础上，通过无监督迁移学习等算法从互联网上的其他无标签语音数据进行自动标注与筛选，从而获取训练模型时所需的大规模语音语料。其次，由于语音中包含丰富的信息，因此语音信号可以用来做多种不同的任务，比如有语音增强、语音识别、声纹识别等任务。本项目设计多任务深度学习算法，设计语音增强、语音识别等多任务的目标函数，通过全局最优化训练模型参数，进而达到语音特征对于复杂环境、多任务的鲁棒性性能。

## 2）语音识别系统

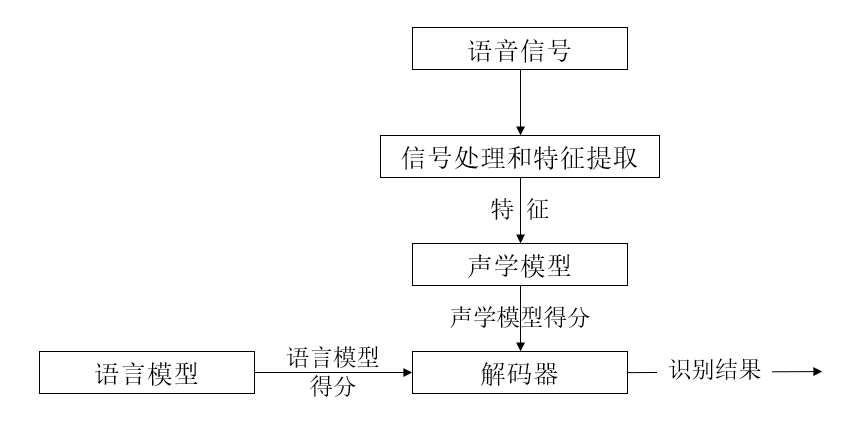


图 5 语音识别流程

经典语音识别系统主要由四部分组成：信号处理和特征提取、声学模型、语言模型和解码器。目前基于端到端技术的语音识别，在语料充分的情况下，性能更优于先前经典的语音识别模型，结合当前主流语音识别发展进程，本课题重点研究在多噪声、干扰的复杂环境下在特征提取和声学模型部分的优化。

根据语音信号的时序性，遵循时序建模的思想，使用RNN-LSTM模型对传统声学模型进行逐步优化，为不同的鲁棒性任务提供一个通用的性能优良的语音识别系统。

* **任务导向的声学特征学习和提取**

先利用前面提到的语音增强中信号处理子模块，处理简单噪音、提升信噪比。然后根据任务导向，训练不同适应性的特征提取神经网络模型，实现不同的识别任务选用不同模型提取不同特征。并进行对特征信息的进一步处理增强，如特征降维，特征拼帧及跳帧等一系列特征处理技术，由此获取适用于任务更加具有表达能力的特征。



图 6 任务导向的特征学习和提取

* **基于时序建模和attention机制的端对端语音识别系统**

关于声学模型，主要有两个问题是一直以来需要不断优化的，分别是特征向量序列的可变长和音频信号的丰富变化性。可变长特征向量序列问题传统上使用隐马尔科夫模型方法来解决。而音频信号的丰富变化性是由说话人的各种复杂特性或者说话风格与语速、环境噪声、信道干扰、方言差异等因素引起的，这就要求声学模型需要有足够的鲁棒性来处理以上的情况。

传统的GMM-HMM和DNN-HMM为语音识别声学模型打下了坚实的基础，但无法利用历史信息来辅助当前任务。本课题拟使用RNN-LSTM模型，来有效利用历史信息，将历史记忆持久化。并结合当前最先进的Attention机制，提升识别性能。

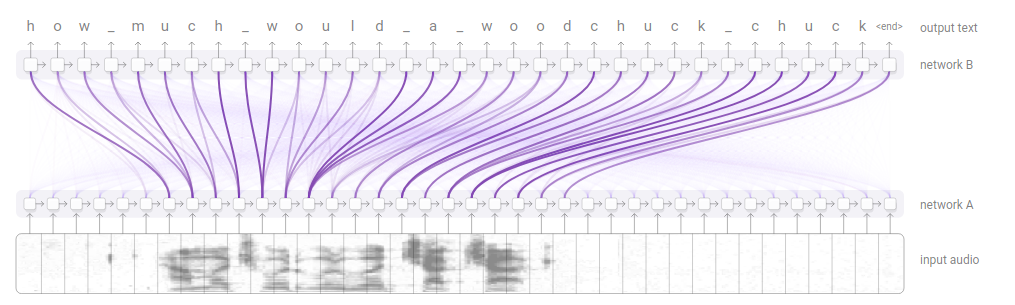


图 7 结合CTC和Attention的LSTM网络

传统声学模型是对固定长度上下文语音帧区段建模，实现每次识别一个声学标签。

本课题拟结合CTC准则，以对不定长的任意语音段和对应的声学标签序列建模，实现从完整音频帧序列直接生成待识别声学标签序列（如图7，直接识别完整序列）。

然后，引入attention机制的LSTM框架，在对序列性语音特征的编码时，利用attention机制得到多个记忆信息，然后在后端使用LSTM进行解码预测每个待识别词时，上述的记忆信息就会更加关注于对其识别最有效的信息上，从而增加每个独立词的识别性能。

## 3）声纹识别系统

* **基于深度学习的特征提取**

提高系统短语音识别准确度的关键在于，在有限的语音数据中挖掘出隐含的说话人信息。深层神经网络拥有多层非线性抽象表达能力，使其学习得到的特征可以携带对分类更有效的信息。

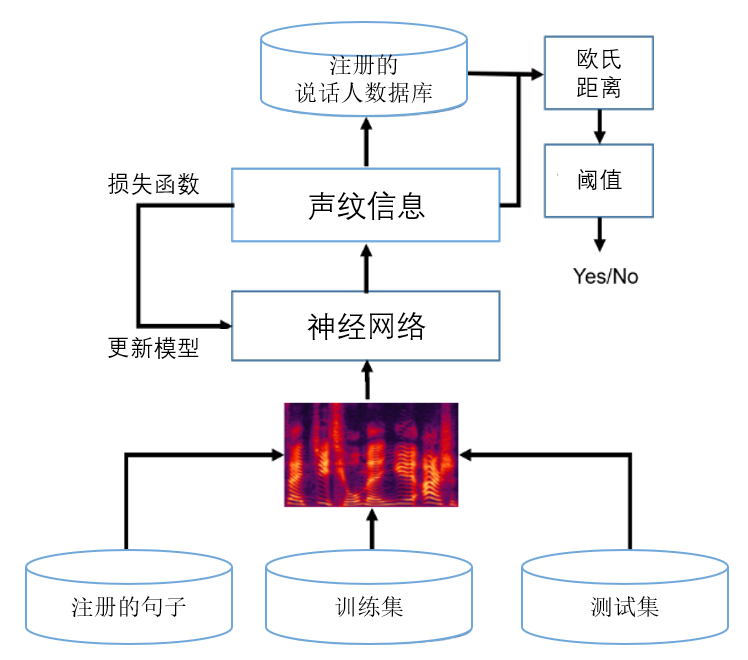


图 8 声纹识别

基于深度学习的声纹识别是通过构建一个更加直接的系统（输入端是语音信号，输出端对应说话人的标签），来联合优化整个系统。通过最近的研究工作来看，在深度神经网络部分，CNN、DNN、RNN、LSTM等端到端模型等都取得了很好的识别效果。

本研究将利用深层网络的特征抽取能力获取深层次特征（即embedding特征）以得到更好的声纹表达，以此来更好地对声纹进行区分和判别。

Embedding特征算法近期的研究工作表明，基于深度神经网络得到的speaker embedding算法能够得到更加具有鲁棒性的特征。对于训练和测试不匹配、噪音、测试语音过短等情况都表现出了良好的特性。

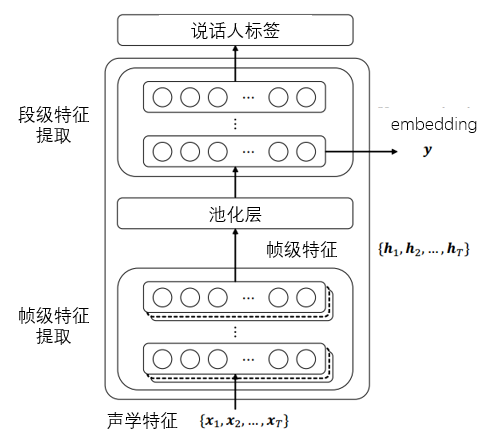


图 9 speaker embedding系统

常用的speaker embedding实现如图9所示，该模型利用神经网络进行提取帧级特征，然后通过池化层提取段级embedding信息，将语音映射到一个固定维度的超向量，从而可以通过欧氏距离或者余弦相似度来衡量声纹的相似度。

提取帧级特征的浅层神经网络的结构比较简单，本课题将结合深度学习对特征的抽取能力， 研究不同类型神经网络的融合提取出更加精炼，更具区分性的特征，以此来提高识别率。

* **基于Triplet Loss等损失函数选取策略优化**

常用的softmax损失函数在区分性上一般。最近图像识别上提出的带有区分性的深度学习损失函数，诸如：angular-softmax，triplet loss等，在图像分类上取得了非常高的准确率。

Triplet Loss是深度学习中的一种损失函数，用于训练**差异性较小**的样本，如人脸等， Feed数据包括锚（Anchor）示例、正（Positive）示例、负（Negative）示例，通过优化锚示例与正示例的距离**小于**锚示例与负示例的距离，实现样本的相似性计算。

Triplet Loss中锚（Anchor）示例的选取，以及新出现带有区分性的其他损失函数，目前还有改进的空间。本研究拟在之前基础上结合angular-softmax等，改进损失函数，提高模型性能。

## 4）智能终端部署方案

完整的复杂环境语音识别系统如图10描述，首先前端语音增强模块对多噪音、多干扰的语音先进行信号处理，然后根据不同任务使用不同任务导向的特征提取子模块提取得到任务导向特征，再使用不同任务模块进行智能处理得到识别结果。



图 10 复杂环境语音识别系统

* **可移植性**

考虑到系统的可移植性，本课题中将分离与硬件相关的代码，通过统一的接口与硬件平台交互，减少了与硬件平台的耦合度。采用分层的软件技术，将系统分为平台接口层、逻辑层和输出层。当需要移植到新的平台时，只需修改与硬件平台相关的代码即可实现兼容。



图 11 系统分层结构设计图

* **云模式部署方案**

云端模式包含前端系统和后端系统，其中前端系统是一个可移植的轻量级系统，可轻松嵌入至各种类型的智能终端设备中，该设备只需对周围环境中说话人的语音信号进行拾音处理和压缩编码，通过智能终端的网络发送至后端的云端服务器，由运算能力强大的云端服务器中各种复杂的神经网络模型进行语音识别处理

在云端模式下，因为终端设备和云端服务器的传输带宽受限， 要保证终端设备识别任务的时效性。

本课题的智能终端工作在云端模式时，不直接向云端服务器传输原始音频信号，而是先使用语音增强模块提取特征，然后终端只向云端服务器传输提取后特征，因为提取得到的特征远远小于原始音频，所以能够保证传输的时效性。

在终端进行特征提取，就要考虑终端算力问题，所以需要在不降低各个模块有效性条件下，对各个基于神经网络的任务导向特征提取子模块的模型在实现时，尽量精简特征提取子模块的结构，实现传输和特征提取之间的平衡。

* **离线部署方案**

离线模式的语音识别系统完全由嵌入至终端设备的智能语音系统独立完成，主要应用在安防领域的不便连网的设备中，也可以作为在线模式的一个备用模式，以防止出现网络异常情况下系统出现问题，提高系统的鲁棒性。

离线部署方案需要将完整系统全部存入智能终端中，这样算力和存储空间限制会成为主要问题。

针对该问题，前端特征提取模块与云端模式下前端的保持一致。



图 12 系统压缩

对于后端语音、声纹识别网络模块则先利用知识蒸馏方法压缩网络深度，然后利用低秩矩阵分解方法将最消耗算力和空间的某些不重要的LSTM层转化为非线性函数变换得到一个在不损失准确率条件下更适合部署在终端的模型，最后再通过稀疏化方法，将神经网络模型中不必要的神经元连接去掉，实现网络剪枝进一步降低算力和空间需求，实现在终端设备上离线部署。

# 6. 总经费预算、资金筹措方式及来源

整体项目计划总投资总额为990万元，其中市财政拨款为245万元，其余745万元来自公司自筹。

项目经费预算如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **资金来源预算** | **预算金额** | 其中 | | | |
| **2018年** | **2019年** | **2020年** | **2021年** |
| 国家部委拨款 |  | 0 | 0 | 0 | 0 |
| 市财政拨款 | 245 | 27 | 70 | 75 | 73 |
| 区县财政配套 |  | 0 | 0 | 0 | 0 |
| 主管部门配套 |  | 0 | 0 | 0 | 0 |
| 单位自筹 | 745 | 8 | 229 | 254 | 254 |
| 银行贷款 |  | 0 | 0 | 0 | 0 |
| 其它 |  | 0 | 0 | 0 | 0 |
| 来源合计 | 990 | 35 | 299 | 329 | 327 |
| **资金支出预算** | **预算金额** | **其中市财**  **政拨款** | **计算依据** | | |
| 人员费 | 745 | 0 | 工资 | | |
| 设备费 | 101.4 | 101.4 | 服务器 信号发生器 信号分析仪 数字示波器等 | | |
| 材料费 | 96.6 | 96.6 | 语料采购 材料、辅助材料、低值易耗品、元器件、试剂、实验动物、部件、外购件等 | | |
| 测试化验加工费 | 20.5 | 20.5 | PCB制版 加工 | | |
| 燃料动力费 | 0 | 0 |  | | |
| 差旅/会议/国际合作交流费 | 22.5 | 22.5 | 差旅费 | | |
| 档案/出版/文献/信息传播/知识产权事务费 | 4 | 4 | 知识产权4件，每件1万 | | |
| 专家咨询费 | 0 | 0 |  | | |
| 其它费用 | 0 | 0 |  | | |
| 支出合计 | 990 | 245 |  | | |

# 7. 项目进度安排

2018年10月-2019年05月：

主要进行信号预处理方法中小波变换等技术的研究，同时进行语料收集。

2019年05月-2019年11月：

在上面信号处理方法研究基础上，进行深度学习方法特征提取框架的研究。

2019年11月-2020年05月：

按语音、声纹识别中任务导向，研究调整深度学习特征提取框架，得到任务导向特征提取模型。

2020年05月-2020年11月：

以前面研究的抗噪抗干扰以及任务导向特征提取模型，结合时序建模和attention机制开发新型LSTM语音识别网络。

2020年11月-2021年05月：

以前面研究的抗噪抗干扰以及任务导向特征提取模型，进行噪声干扰环境下声纹识别研究。

2021年05月-2021年11月：

进行样机试产、调试，解决部署问题，以及样机上线测试。

# 8. 预期的经济、社会效益分析及成果应用、产业化衔接情况

在经济效益方面，本项目于2018年10月启动，2020年末形成样机并开始市场推广。产品单价预计初期20万元/套，随后根据市场的情况进行调整。预计从2020年开始，头三年销量分别为30台,50台，80台。

从2020年到2022年，预计总产值：5000万元，营业收入：5000万元，技术性收入：0万元，净利润额：1500万元，上缴税金：600万元。

在社会效益方面，本项目可向社会提供40个就业岗位。同时，因为本项目完全符合天津市科委制定的新一代人工智能科技重大专项项目的要求以及产业政策。项目的实施必将促进我国科技领域的发展以及在安全领域中发挥积极有效的重要作用。

复杂环境语音智能综合系统将应用于公司安防设备中，在保证产品性能的稳定性和可靠性的前提下扩展产品的功能，增强产品的智能化和高效性。该系统可广泛应用于安防、语音导航、智能设备控制等领域，还可与其他自然语言处理技术相结合，构建出更加复杂的应用，例如语音到语音的翻译等。

# 9. 研究工作基础及条件

## 1）现有技术和工作基础

我司在安防领域，对信号处理技术的研究使用有多年经验，具备对信号进行抗噪抗干扰处理的核心技术。在语音识别领域开发过针对维吾尔语、哈萨克语、普通话的语种、语音识别。

我司已具备信号抗扰抗噪处理技术，以及语音、声纹识别的技术积累，并在项目组中有多位领域内专家博士以及硕士工程师，具备技术、研究和人才基础。

慧言科技（天津）有限公司依托天津大学语音识别技术在语音识别、声纹识别等核心算法上有了一定的突破，已经形成了自己的引擎，且正在逐步完善。在发音机制及控制的生理计算建模、言语习得、远场语音识别、声纹识别、智能语音交互等语音相关方面的研究成果处于国内领先地位

## 2）实施条件

我司已配备多台GPU人工智能服务器， 和多家高技术企业有密切技术合作关系，并且之前在人工智能语音识别领域的研究中，已积累大量的语音训练预料。

我司已在安防领域布局多款智能终端设备多年，曾研发基于2G系统的语音分析系统。该系统于2015年研发完成，基于该系统的产品已经形成销售具有良好的安防市场，本课题实现的抗扰抗噪语音、声纹识别系统可直接嵌入到我司安防终端设备中，实现一体化拾音、识别终端，实现快速人工智能赋能落地。

## 3）现有知识产权方面

专利：

一种自软语言语义推理方，201711296575.2

一种针对自然语言进行情绪识别的方法，201711294193.6

一种基于影音定位技术的服务机器人降噪方法，201811107223.2

一种基于振幅和相位信息的语音情感识别方法，201811212955.8

论文：

(1) Jianwu Dang, Qinghua Hu, Hui Feng, Chinese speech information processing in the Internet age: From representation to understanding, Addressing Global Science Issues 2014 Supplement, 2014, 53-58.

(2) Jianwu Dang, Jianguo Wei, Kiyoshi Honda, Takayoshi Na, A study on transvelar coupling for non-nasalized sounds, The Journal of the Acoustical Society of America, 2016, 139(1)：441-454.

(3) Jianwu Dang, Aijun Li, Donna Erickson, Atsuo Suemitsu, Masato Akagi, Kyoko Sakuraba, Nobuaki Minematsu, Keikichi Hirose, Comparison of Emotion Perception among Different Cultures, Acoustics of Science and Technology, 2010, 31(6)：394-402.

(4) Wu xiyu, Jianwu Dang, Control strategy of a physiological articulatory model for speech production, Journal of Chinese Linguistics, 2015, 43(1B)：337-363.

(5) Songgun Hyon, Jianwu Dang, Hui Feng, Hongcui Wang, Kiyoshi Honda, Detection of speaker individual information using a phoneme effect suppression method, Speech Communication, 2014, 57：87-100.

(4) Xiyu Wu, Jianwu Dang, Stavness Ian, Iterative method to estimate muscle activation with a physiological articulatory model, Acoustic Science & Technology, 2014, 35(4): 201-212.

(6) Li Qiang, Dang Jianwu, Han Yahong, Image decomposing for inpainting using compressed sensing in DCT domain, Frontiers of Computer Science, 2014, 8(6): 905-915.

(7) Lili Guo, Longbiao Wang, Jianwu Dang, Linjuan Zhang, Haotian Guan, A feature fusion method based on extreme learning machine for speech emotion recognition, ICASSP 2018, 2666-2670.

## 4）研究队伍

项目负责人齐心：是天津市“1-3-1”创新型人才，共申请美国及欧洲发明专利40余项，已授权中国发明专利20余项，在国际期刊及会议上发表学术论文十余篇。齐心博士专业特长包括：移动通信技术，信息安全技术，人工智能技术等，是移动通信、人工智能领域内专家，在抗扰抗噪方面有多年经验，对于语音、声纹识别技术更是有着深入研究。

慧言首席科学家党建武：国家千人计划特聘专家，973项目首席科学家、中国计算机学会语音对话与听觉专委会主任、日本北陆先端科学技术大学院大学智能信息处理实验室主任，天津大学计算机学院教授、博导、前任院长。他研发出的人的发音机制及控制的生理计算模型一直在该领域处于领先地位。是国内最早一批研究人工智能、神经网络等技术的权威专家，为本课题语音、声纹识别领域技术突破提供保障。

目前本项目团队在发音机制及控制的生理计算建模、言语习得、远场语音识别、声纹识别、智能语音交互等语音相关方面的都有多年研究开发经验。团队在反声纹欺骗国际标准评测Anti-spoofing 2015中获得国际第二的优异成绩，在复杂环境下的语音识别国际评测比赛CHiME 2015中获得全球第四的成绩。

## 5）产学研合作情况

与清华、天大建立起了密切的产学研合作关系，培养多名硕士研究生人才。研究方面,在人工智能领域、语音信号处理都有着深度密切的技术合作。

# 10. 承担单位简介及合作方式

## 1） 因诺微科技（天津）有限公司

因诺微科技（天津）有限公司成立于2012年。公司总部位于天津，北京设有研发中心。公司专注于通信、人工智能、大数据等在信息安全领域的研发及应用。

公司创始人均毕业于清华大学电子工程系，曾在诺基亚、高通、中国电子科技集团等大型企业任职，从事信息技术领域的科研、技术标准化和产品研发工作，拥有很强的技术实力和丰富的公司管理及产业化经验。目前公司有员工80余人，人才结构合理，研发实力强劲。研发团队人数占比超过30%，本科以上学历占85%以上，博士3人，硕士14人，高级职称4人，中级职称4人，核心研发人员均来自清华大学电子工程系。

公司自成立以来，发展迅速，经营状况良好，从业人员达到80余人。2015-2017年的营业收入分别为800万元、2300万元、4500万元，增长性极强，截止到2018年10月份，已签订的合同总额突破8000万元，预计全年营业额突破亿元。公司是以自有技术为根基，广纳贤才，研发团队由最初的5人发展到30余人，研发投入不断增加，2018年预计投入500万元。

公司拥有国家级高新技术企业、天津市高新技术企业、武器装备科研生产单位三级保密资格（军工三级保密资格）、装备承制单位资格、国家军用通用标准质量管理体系、软件企业、ISO9001质量管理体系认证等企业资质，并拥有多项发明专利和软件著作权等知识产权。

历经数年，公司研发团队在语音信号处理等方面积累了丰富的经验，并实现了关键技术验证，同时与多家国内知名高校形成产学研合同协议，如清华大学、浙江大学、南开大学、天津大学等，在人才和技术储备上打下了良好的基础，形成了一个“三脚架”的稳固架构，为公司提供了一个很好的平台。

## 2） 慧言科技（天津）有限公司

慧言科技（天津）有限公司于2018年4月在天津滨海高新区成立，公司主营业务范围包括智能语音处理技术及其行业应用产品等。公司研发团队在复杂环境下的语音识别、声纹识别、智能语音交互、人工智能等领域积累了丰富的经验，同时与多家国内知名高校形成产学研合作。

公司致力于发展成为国内外主要的智能语音科技公司之一。结合生产销售企业，立足市场，从应用和用户侧角度出发，围绕细分领域场景进行语音产品研发，旨在提供语音技术细分领域的实际应用和解决方案，为用户带来 “锦上添花”、“雪中送炭”的定制化体验，提高人们的生活质量。

## 3） 双方分工

因诺微：负责项目管理，硬件平台研发，系统开发等工作；

慧言：负责理论分析，关键模型推导，和系统仿真验证等工作。

# 11. 项目负责人及主要成员简介

齐心：工学博士、2006年7月毕业于清华大学信息与通信工程专业，师从移动通信领域的领军人物——国家未来移动通信（FUTURE）项目总体组专家王京教授。专业特长包括：移动通信技术，信息安全技术，人工智能技术等。2015年，被天津市团市委推荐为天津青年企业家协会会员；2016年，所参与的国家863FUTURE项目（第四代移动通信系统TD-LTE关键技术与应用）获得了国家科学技术进步奖特等奖；2017年，被评选为天津市“1-3-1”创新型人才。

齐心博士曾任诺基亚及诺基亚西门子在IEEE 802.16m国际标准化会议中的首席代表、IEEE 802.16m国际标准化会议分会主席，在3GPP LTE国际标准化会议中的RAN1代表，获得诺基亚西门子公司2008年度“全球最佳发明人”称号。在清华大学无线中心工负责研发的基于软件无线电的C-RAN系统，部分成果在英特尔2011年度全球开发者大会作为主题宣传片展出。共申请美国及欧洲发明专利40余项（已有近20项获得美国和欧洲专利局授权），已授权中国发明专利20余项，在国际期刊及会议上发表学术论文十余篇，其中两篇被SCI检索。

党建武：慧言科技首席科学家，国家千人计划特聘专家，973项目首席科学家，中国计算机学会“语音、对话与听觉”专委会主任，负责慧言公司语音技术的研发工作。1984年毕业于清华大学获得工学硕士学位，1992年获得日本静冈大学工学博士学位。之后工作于日本ATR人类信息通讯研究所、加拿大滑铁卢大学、日本北陆先端科学技术大学院大学、法国科学院格勒诺布尔语音通讯研究所、天津大学。2018年4月做为联合创始人成立了慧言科技（天津）有限公司并担任首席科学家，负责公司研发管理。他用机械模型、生体实验以及计算模型等方法首次严谨地证明了梨状窝在人的语音生成及个性化中的作用，这一研究被誉为开拓性的基础研究。他研发出的人的发音机制及控制的生理计算模型一直在该领域处于领先地位。他组织和分担过美日、中日多项国际合作研究开发项目，发表了数十篇领域内权威期刊论文及二百余篇国内外会议论文。

# 12. 风险分析

风险分析从以下几方面来说明：

A. 技术风险

软件的安全性：通过专利保护、加密等方式对软件版权进行有效的保障；

关键人员的流失：通过合理的激励方式保证关键人员的积极性，通过对技术文档进行归档和对参与人员签署保密协议来确保在关键人员离职的情况下核心技术不流失，通过合理的梯队建设确保研发团队的可持续发展。

B. 市场风险

项目时间风险：通过迅速启动项目研发，严格控制研发流程，依靠人才及资源积累，保持技术的领先时间约一年；

项目成本风险：确保产品在既定时间内上市推广，产品的核心部分为软件，通过快速扩展市场、提高销售量、降低研发成本，最大程度上降低产品的成本价格，提高价格竞争优势。

C. 政策风险

通过和合作单位、客户的紧密合作、交叉合作、技术互动，并且双方共同对产品进行规范测试来确保项目的正常运行。