

中科院声学所DNNVAD引擎

代码模块接口设计说明

目录

[1.概述 5](#_Toc91666797)

[1.1简介 5](#_Toc91666798)

[1.2 读者对象 5](#_Toc91666799)

[1.3术语定义 6](#_Toc91666800)

[2.总体设计 6](#_Toc91666801)

[3.接口设计 9](#_Toc91666802)

[3.1 引擎接口设计 9](#_Toc91666803)

[3.1.1 引擎初始化接口 9](#_Toc91666804)

[3.1.2 引擎分段接口 10](#_Toc91666805)

[3. 1.3 引擎设置结果回调函数 10](#_Toc91666806)

版 本 历 史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本/状态 | 作者 | 参与者 | 起止日期 | 备注 |
| 1.0 | 陈向东 | 龚云波  黎塔 | 2021/04/28 |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 1.概述

1

1.1简介

DNNVAD的主要功能：前端功能，对输入的原始语音进行有效音截取，使得输入到识别引擎中的语音为有效音从而减少噪音对识别引擎的干扰。本引擎主要基于DNN模型的VAD算法。对原始语音进行特征提取并输入DNN模型，根据DNN模型的输出概率值并结合维特比算法，对原始语音逐帧判别是否为有效语音，最后根据起尾点判别算法将原始语音中的有效语音截取出来。

## 1.2 读者对象

本文档是对识别引擎的内部代码流程和相关技术点的整体说明，因此读者最好满足如下几点条件后再读此文档：

* 有一定的C和C++语言基础
* 了解语音识别过程和原理
* 了解声学所识别引擎实现原理

## 1.3术语定义

DNN(Deep Neural Networks）：深度神经网络

VAD（Voice Activity Detection）：有效音检测

# 2.总体设计

2

DNNVAD主要模块主要流程包括特征提取、DNN模型推理、维特比寻址、起尾点判断，大致流程图如下：

引擎总体设计流程如图2.1所示：

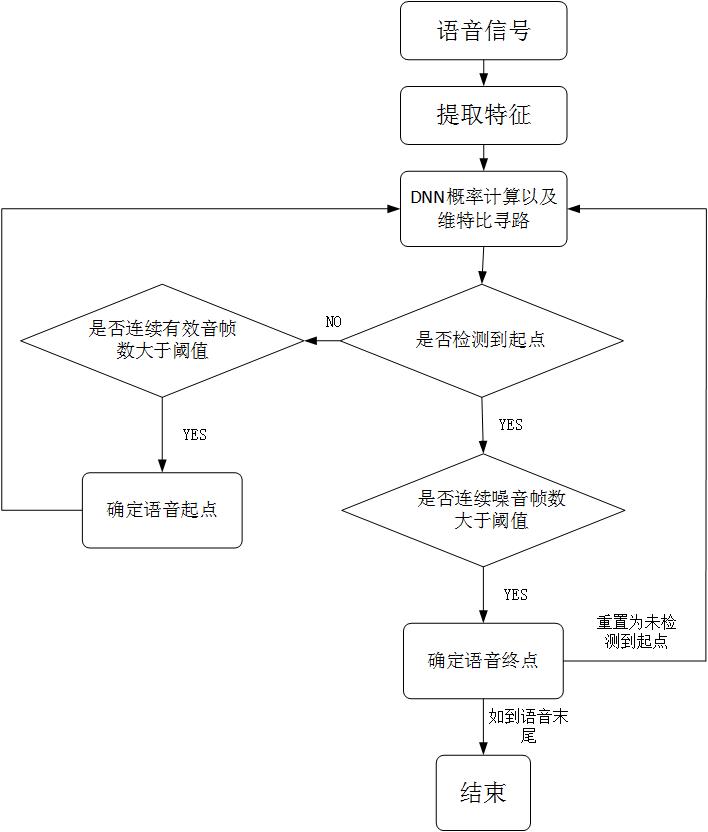


图2.1 引擎总体设计流程

语音数据送入引擎后，经过“特征提取->DNN概率计算->维特比寻路->结果判决”，最终得到是否唤醒的结果。

* 特征提取：输入的语音数据分别经过预加重、分帧、加窗、FFT、梅尔滤波、对数运算、DCT后形成每帧40维的MFCC语音特征，语音特征数据将被送入DNN概率计算模块；
* DNN概率计算：在引擎初始化阶段加载DNNVAD模型，构建神经网络；将特征送入神经网络中，计算出后验概率，送入结果判决模块；
* 维特比寻路:利用神经网络输出的后验概率，进行维特比寻路，判别输入的语音帧是语音或者非语音；
* 结果判决：根据连续的语音帧或者非语音帧是否超过阈值，来判断语音的开始以及结束。

# 3.接口设计

3

## 3.1 引擎接口设计

## 3.1.1 引擎初始化接口

|  |  |
| --- | --- |
| 函数原型 | **DNN\_VAD(char \*config);** |
| 功能 | 加载配置文件 |
| 参数说明 | **char \*config:配置文件位置** |
| 返回值 |  |
| 说明 |  |

## 3.1.2 引擎分段接口

|  |  |
| --- | --- |
| 函数原型 | **int doVAD(float \*Feature\_in, int FrameNum, int IsFinished);** |
| 功能 | 接收数据并获取分段结果 |
| 参数说明 | **float \*Feature\_in：语音特征** |
| **int FrameNum：语音特征帧数** |
| **int IsFinished：是否送入完毕，默认1** |
| 返回值 | 正常为0，异常为-1 |
| 说明 |  |

## 3. 1.3 引擎设置结果回调函数

|  |  |
| --- | --- |
| 函数原型 | **int SetVadCallback(Dnn\_vad\_callback pFunc, void\* pObj);** |
| 功能 | 将VAD输出的结果通过回调函数的形式返回出来 |
| 参数说明 | Dnn\_vad\_callback pFunc：定义的回调函数句柄 |
| void\* pObj：占位空指针，为识别引擎服务 |
| 返回值 | 正常为0 |