

中科院声学所叠音检测引擎

开发测试

目录

[1.概述 1](#_Toc89095512)

[1.1简介 1](#_Toc89095513)

[1.2 读者对象 1](#_Toc89095514)

[1.3术语定义 1](#_Toc89095515)

[1.4 参考资料 1](#_Toc89095516)

[3.测试工程 1](#_Toc89095517)

[3.1 测试工程 1](#_Toc89095518)

[3.1.1测试工程目录结构说明 1](#_Toc89095519)

[3.1.2 测试工程编译说明 1](#_Toc89095520)

[4.测试环境 1](#_Toc89095521)

[4.1 测试工程 1](#_Toc89095522)

[4.1.1测试环境目录结构说明 1](#_Toc89095523)

[4.1.2测试环境使用说明 1](#_Toc89095524)

[4.1.3测试环境结果查看 1](#_Toc89095525)

[4.1.3测试环境配置文件 1](#_Toc89095526)

版 本 历 史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本/状态 | 作者 | 参与者 | 起止日期 | 备注 |
| 1.0 | 张英楠 |  | 2021/04/28 |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 1.概述

1

1.1简介

叠音自动检测在语音技术领域中具有十分重要的地位。该技术是指从声音信号中寻找出多个说话人语音的重叠部分。进行有效的叠音自动检测能够对声音信号中多个说话人互相打断对话的情况进行检测。

## 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
* 了解语音信号处理过程和原理

## 1.3术语定义

DNN(Deep Neural Networks）：深度神经网络

# 3.测试工程

3

## 3.1 测试工程

测试工程组要完成对叠音检测引擎接口的集成，实现识别引擎对录音的转写功能。提供给用户对如何集成识别引擎，引擎具体工作流程和测试一个参考实例。

### 3.1.1测试工程目录结构说明

参考提供测试环境testproj测试工程用例。

testproj目录结构说明

- main.cpp 叠音检测引擎调用

- usage.cpp 读取文件相关函数实现

- Makefile\_Overlap 叠音检测引擎测试工程的makefile文件

### 3.1.2 测试工程编译说明

直接在编译环境中编译Makefile\_Overlap文件即可

# 4.测试环境

4

## 4.1 测试工程

测试环境为识别引擎完成转写等功能所依赖的运行环境。

### 4.1.1测试环境目录结构说明

|- NewTest\_Overlap 引擎的工作目录

|-- fe\_source引擎相关配置文件目录

|--- detect\_overlap 叠音检测引擎配置文件及模型目录

|--- extract\_feat 特征提取配置文件目录

|--- vad 有效音检测配置文件及模型目录

|-- front\_end.cfg 测试用例配置文件

### 4.1.2测试环境使用说明

打开引擎脚本run.sh修改相应参数：

* 修改wavList语音列表
* 修改nThread开启线程数
* 修改OutputFile输出文件名
* 启动引擎启动脚本run.sh;

### 4.1.3测试环境结果查看

配置好detect\_overlap.cfg、run.sh中的相关参数，直接运行run.sh即可。

1. “cd Overlap \_Test/bin”进入bin目录下；
2. 根据具体需求，配置detect\_overlap.cfg文件中的相应参数；
3. 根据具体需求，配置run.sh文件中的相应参数；

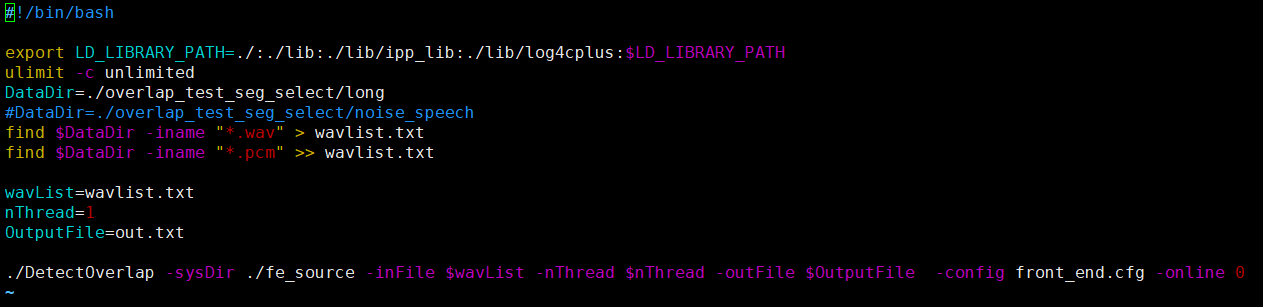
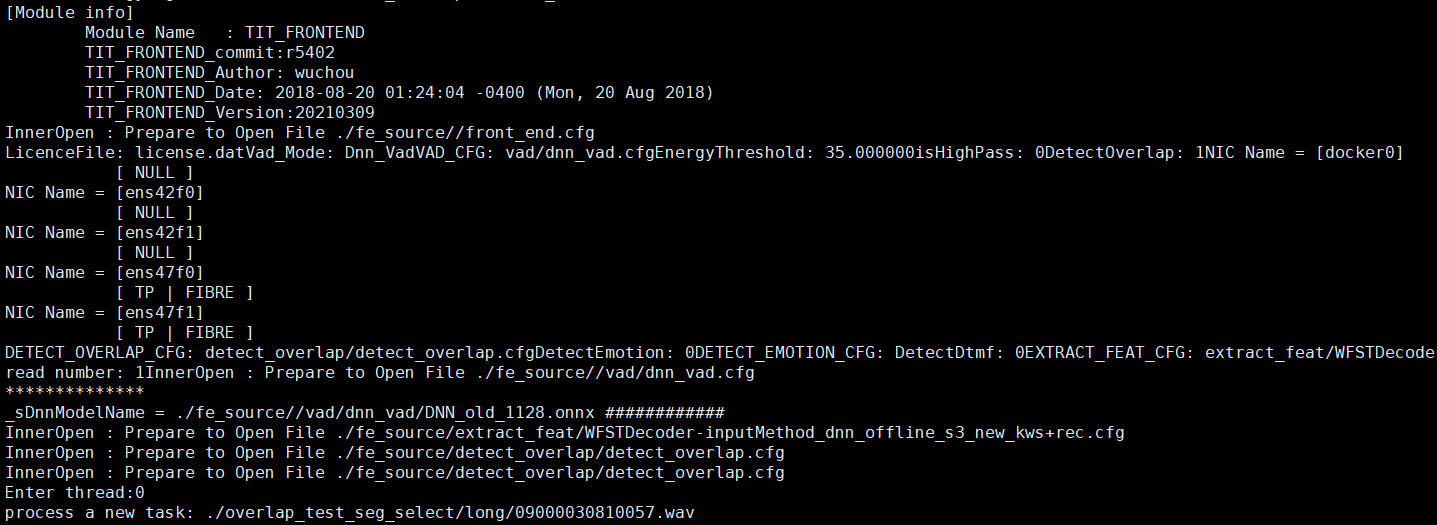


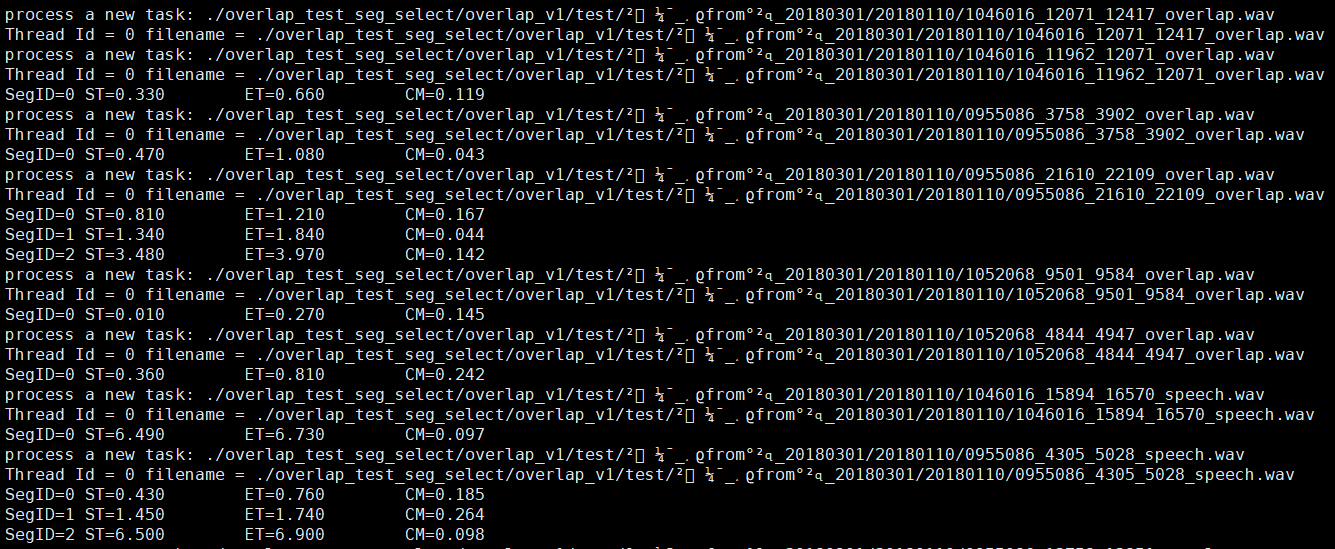
图 3：run.sh文件配置图

1. ./run.sh启动软件；
2. 运行过程；

软件启动后，首先进行初始化，



然后进行叠音段的起始点和尾点检测



### 4.1.3测试环境配置文件

**设置配置文件detect\_overlap.cfg**

WAVFORMAT---输入语音格式，目前支持8k16bit pcm

ANNMODEL ---叠音模型名称

FRMDIM---语音特征维数

LEFTCONT---神经网络模型左拼帧数

RIGHTCONT---神经网络模型左拼帧数

Online---缓存帧数

MSBias---叠音概率增益

SSBias---单人语音概率增益

MSpeechMiniLen---连续语音帧数阈值

MNoneSpeechMiniLen---连续叠音帧数阈值

MOverLapSpeechLen---叠音片段最大长度