

中科院声学所离线场景分割引擎

概要设计说明

版 本 历 史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本/状态 | 作者 | 参与者 | 起止日期 | 备注 |
| V1.0 | 王超锋 |  | 2021-11-02 | 创建文件 |
|  |  |  |  |  |

目录

[1.概述 3](#_Toc4919)

[1.1简介 3](#_Toc30468)

[1.2 读者对象 3](#_Toc23684)

[1.3术语定义 4](#_Toc4477)

[2.总体设计 5](#_Toc14748)

[2.1 总体需求 5](#_Toc17869)

[2.2 运行环境 6](#_Toc18041)

[2.2.1 软件环境 6](#_Toc31717)

[2.2.2 硬件环境 6](#_Toc32706)

[2.3系统总体结构设计 7](#_Toc26746)

[3.模块设计 9](#_Toc28502)

[3.1任务处理 10](#_Toc32231)

[3.1.1角色区分 10](#_Toc30262)

[3.2 结果发送 13](#_Toc14577)

[3.3 任务接收 15](#_Toc26292)

[4.配置及测试设计 16](#_Toc20673)

[4.1 配置文件设计 16](#_Toc30986)

[4.2 测试功能点 17](#_Toc19345)

[5.可靠性设计 18](#_Toc13772)

[6.扩展性设计 19](#_Toc17884)

[7.维护性设计 20](#_Toc3284)

[8.易用性设计 21](#_Toc26655)

# 1.概述

1

**1.1简介**

角色区分模块主要是对语音说话人角色进行判断（坐席/客户）。

## 1.2 读者对象

本文档的读者对象为离线识别系统的开发人员、测试人员、系统维护人员及接入识别系统的第三方业务人员，通过本文档能够从总体上了解识别系统的架构形式及数据流向。

本说明给出离线识别系统的设计说明，包括最终实现的系统必须满足的功能、性能、接口、附属测试工具程序及设计约束等。

目的在于：

* 为开发人员提供依据；
* 为代码修改、维护提供条件；

## 1.3术语定义

语音识别（Automatic Speech Recognition），简称（ASR）。

# 2.总体设计

2

## 2.1 总体需求

功能需求：对语音进行聚类和角色区分；

性能需求：根据机器性能开不同线数；

接口需求：接口参数简单明了，接口中的每个参数都要有实际意义，保证接口调用流程清晰。

系统验证工具：提供系统完整性，功能正确性的验证工具；

## 2.2 运行环境

### 2.2.1 软件环境

|  |  |  |
| --- | --- | --- |
| 分类 | 名称 | 版本 |
| 操作系统 | Centos | 7.0以上 |
| 数据库 | Redis | 3.2.0 |
| 能力接口 | Tomcat | 9.0.35 |

### 2.2.2 硬件环境

|  |  |  |
| --- | --- | --- |
| 服务器 | 最低配置 | 推荐配置 |
| redis服务器 | CPU：1  内存：256M | CPU：1  内存：1G |
| 识别服务器 | CPU：4  内存：10G  磁盘：20G | CPU：10  内存：15G  磁盘：30G |

## 2.3系统总体结构设计

角色区分部分主要包括一个任务接收线程，一个结果发送线程，一个任务处理线程组成。



任务接收线程：完成接收总控端发送的任务，存入任务队列。

任务处理线程：从任务队列中取出待处理任务，经过调用聚类相关函数，生成AB中间结果文件，在调用角色区分tomcat服务生成role中间结果文件。处理结果存入结果队列。

任务发送线程:从结果队列中取已经处理完的任务，发送到总控端。

# 3.模块设计

3

## 3.1任务处理



### 3.1.1角色区分

主要函数介绍

函数原型：

int InitClustering(const char\* kpcConfigFile = “clustering.cfg”)

参数说明：

const char\* kpcConfigFile[in] 聚类配置文件

返回值：

int类型。取值及说明如下：

0 ：成功

-1 ：失败

功能：

聚类引擎初始化。

说明：

InitClustering()函数为聚类引擎接口函数，在场景分割服务端主函数调用。

函数原型：

int ClusteringOutput( const char\* kpcWavFile, const bool kbOutputWav,const char\* outDir);

参数说明：

const char\* kpcWavFile[in] 待聚类的语音路径

const bool kbOutputWav[in] 是否输出聚类后的语音\*\_1.wav和\*\_2.wav

char\* outDir[in] 聚类结果文件存放的绝对路径

返回值：

int类型。取值及说明如下：

0 ：成功

-1 ：失败

功能：

对语音做聚类，生成\*\_1.txt和\*\_2.txt的聚类结果文件。

函数原型：

bool scene\_segmentation( string wavfile, string sentRstFile,string wordsRsttFile,string ABRstFile,string ABRstFileFormat,string delRstSpace);

参数说明：

string wavfile[in] 语音绝对路径

string sentRstFile[in] 识别结果\*\_sent文件绝对路径

string wordsRsttFile[in] 分词聚类结果\* \_words\_result.txt绝对路径

string ABRstFile[in] \*\_text\_AB\_result.txt绝对路径

string ABRstFileFormat[in] \*\_text\_AB\_result.txt文件格式

string delRstSpace[in] 是否删除\*\_text\_AB\_result.txt文件中转写内容的空格

返回值：

bool类型。取值及说明如下：

true：成功。

false：失败。

功能：

根据\*\_sent.txt及聚类结果\*\_1.txt、\*\_2.txt，生成\*\_text\_AB\_result.txt和\* \_words\_result.txt中间结果文件。适用于单声道语音。

函数原型：

bool Text\_AB\_Result\_File(string wavfile,string sentfile, string ABResultFile,string ABFormat,string delSpace)

参数说明：

string wavfile[in] 语音绝对路径

string sentfile [in] 识别结果\*\_sent文件绝对路径

string ABResultFile [in] \*\_text\_AB\_result.txt绝对路径

string ABFormat [in] \*\_text\_AB\_result.txt文件格式

string delSpace [in] 是否删除\*\_text\_AB\_result.txt文件中转写内容的空格

返回值：

bool类型。取值及说明如下：

true：成功。

false：失败。

功能：

双声道语音，根据\*\_sent.txt文件，生成 \*\_text\_AB\_result.txt；同时整理双声道语音\*\_sent.txt文件格式与单声道语音sent文件格式一样。

函数原型：

int role( string ABResultFile,string roleResultFile )

参数说明：

string ABResultFile[in]，\*\_text\_AB\_result.txt绝对路径

string roleResultFile[in]，\*\_text\_role\_result.txt绝对路径

返回值：

int类型。0：成功。非0：失败。取值及说明如下：

0：成功。

1：失败，connection error.

2：失败，Invalid response.

3: 失败，Response returned with status\_code=200

4: 处理失败。

## 3.2 结果发送

结果发送线程主要是从结果队列中获取任务后，发送处理结果给总控端。

结果发送线程对应线程函数为RestaskSendThreadProc(),流程图如下：



RestaskSendThreadProc()函数流程图

## 3.3 任务接收

任务接收线程主要是接收总控端发送的任务，并将接收到的内容放入任务队列。

任务接收线程对应线程函数为TaskInsertThreadProc(),流程图如下：



TaskInsertThreadProc()函数流程图

# 4.配置及测试设计

4

## 4.1 配置文件设计

设置配置文件，方便使用不同控制逻辑，支持不同的功能及结果输出，场景分割配置文件实现如下目的：

* 通过参数控制支持特殊功能
* 日志输出
* 引擎资源加载

配置文件主要参数如下表：

|  |  |  |
| --- | --- | --- |
| 参数 | 取值 | 说明 |
| ipAddr | 字符串 | 启动IP地址 |
| ports | 字符串 | 启动端口 |
| roleServerPort | 字符串 | Tomcat角色区分端口 |
| logFile | 字符串 | 日志文件 |
| nThread | int | 启动线数 |

## 4.2 测试功能点

根据设计需求，对服务必须满足的功能点进行验证测试。

|  |  |
| --- | --- |
| 测试功能 | 期望结果 |
| 服务启动验证 | 正确配置服务参数，启动服务，CPU、内存满足情况下，服务能够正常启动。 |
| 输出日志 | 打开日志参数，正确输出日志。 |

# 5.可靠性设计

5

离线系统角色区分部分设计考虑了实际应用场景的复杂性和输入的多样性的情况，内部加入了大部分实际情况中会出现的异常情况的处理方案，socket连接异常断开重连、语音文件有效性检查等异常事件的处理；并且整套系统测试按照严格的7\*24小时压力测试标准进行测试，同时也针对所有异常数据进行异常情况测试，保证引擎的稳定性和可靠性。

# 6.扩展性设计

6

服务端处理引擎支持多路并发，保证了不同cpu数和内存大小的机器拓展需求；系统本身遵循模块化设计，可以随时添加定制化功能，满足不同应用场景的需要。

# 7.维护性设计

7

离线系统为了方便后期的引擎维护，采用标准c/c++函数同时在较低编译器上进行编译。

角色区分模块可配置configure.cfg文件部分参数，保证了引擎的可配性和可维护性。

# 8.易用性设计

8

系统安装部署时有一键安装部署脚本，只需要执行脚本即可安装引擎转写系统及所需的中间件（如ffmpeg、sox、redis等），不需要安装人员逐个安装，减少操作量。系统启停均有对应的脚本，实现自动启动、停止服务。