

中科院声学所在线识别引擎

详细设计说明

版 本 历 史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本/状态 | 作者 | 参与者 | 起止日期 | 备注 |
| V1.0 | 杜一广 |  | 2021-11-10 | 创建文件 |
|  |  |  |  |  |

目录

[1.概述 3](#_Toc29714)

[1.1编写目的 3](#_Toc25122)

[1.2 读者对象 3](#_Toc31562)

[1.3术语定义 4](#_Toc14655)

[2.总体设计 4](#_Toc2247)

[2.1 总体需求 5](#_Toc19689)

[2.2 运行环境 5](#_Toc15235)

[2.2.1 软件环境 5](#_Toc1320)

[2.2.2 硬件环境 5](#_Toc21646)

[2.3在线识别服务总体结构设计 6](#_Toc23371)

[2.3.1在线识别服务总体时序流程 6](#_Toc14968)

[3.能力接口模块设计 8](#_Toc3126)

[4.数据设计 10](#_Toc11483)

[4.1 WebSocket接口请求 10](#_Toc7740)

[4.1.1 开始帧 10](#_Toc20829)

[4.1.2 音频数据帧 11](#_Toc24220)

[4.1.3 结束帧 11](#_Toc19880)

[4.1.4 接收帧 12](#_Toc23581)

# 1.概述

1

1.1编写目的

本文档主要用于为实现系统功能而进行的系统数据设计说明，具体描述了系统包含的的软件模块的实现流程、功能、接口、数据结构等内容，供项目组开发人员和软件维护人员阅读。

## 1.2 读者对象

本文档的读者对象为在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。

本说明给出在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）。

# 2.总体设计

2

## 2.1 总体需求

功能需求：满足8k16bit pcm、16k16bit pcm语音转文字功能；

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 |

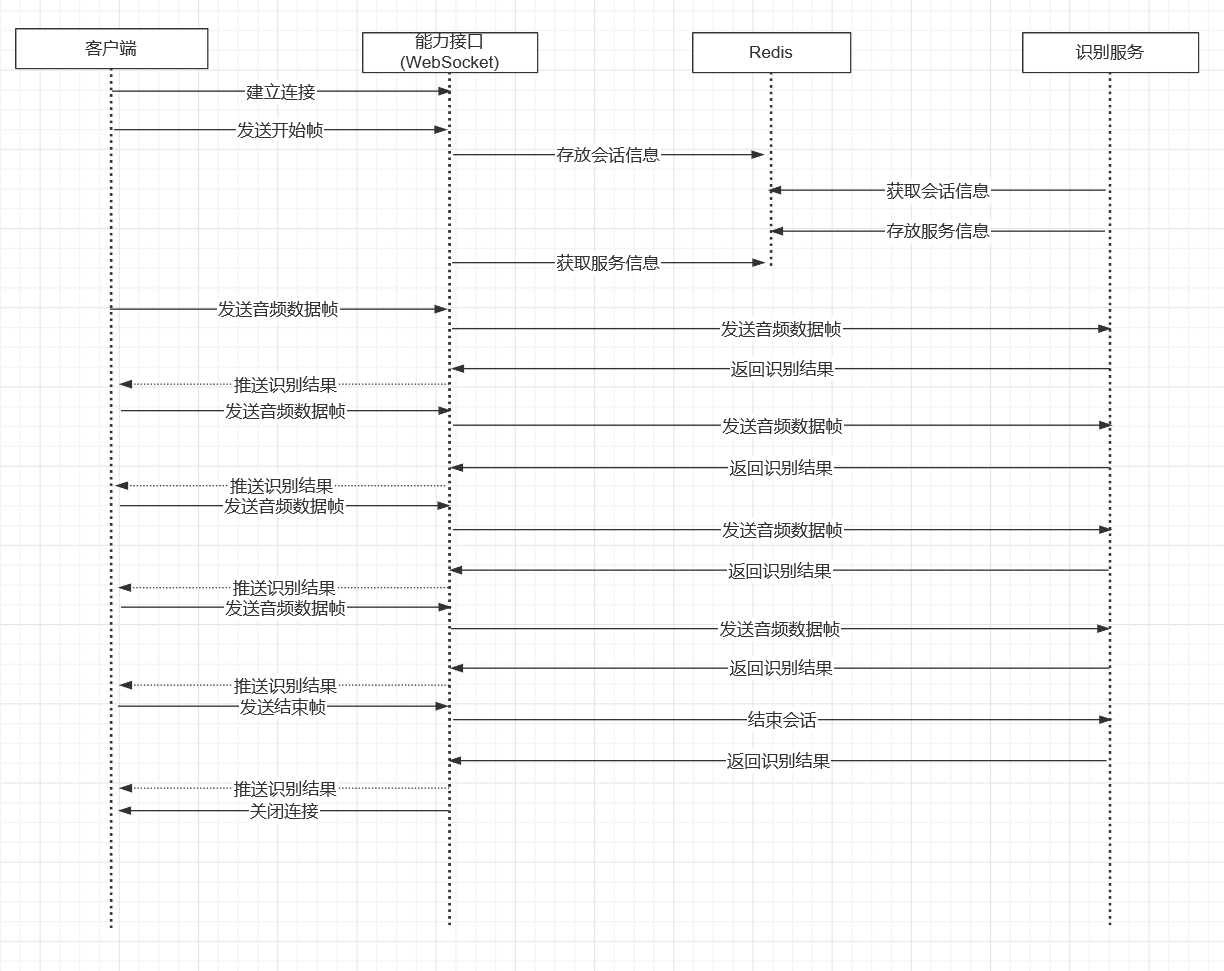
### 2.2.2 硬件环境

|  |  |  |
| --- | --- | --- |
| 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在线识别服务总体结构设计

### 2.3.1在线识别服务总体时序流程

在线识别服务总体时序流程如下所示：

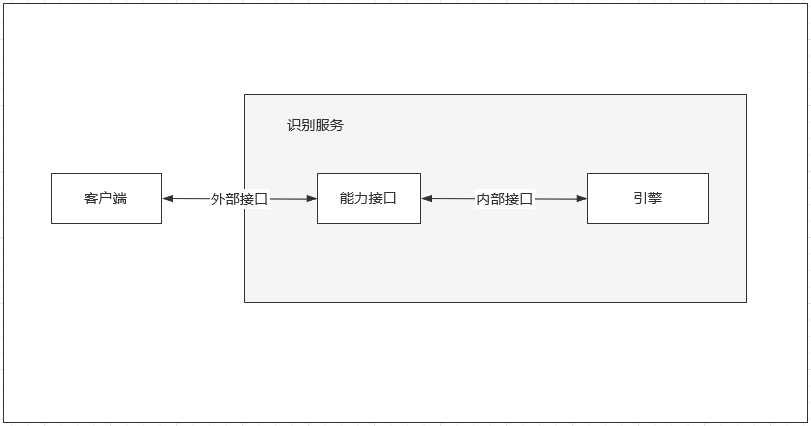


# 3.能力接口模块设计

3

在线识别能力接口通过外部接口与客户端交互，通过内部接口与引擎进行交互。

模块间关系如下：



外部接口相关函数介绍：

该接口主要通过WebSocket协议与客户端进行网络通信的数据交互，接收客户端发送的请求数据，并校验引擎是否返回识别结果，当有识别结果时，将结果推送给客户端。

**public** **void** **onOpen** (Session session);

建立连接：客户端与服务端进行握手操作。

**public** **void** **onClose**(Session session,String sid);

关闭连接：服务端主动与客户端断开连接，一般发生在结束帧之后。

**public** **void** **onMessage**(**byte**[] bytes,Session session);

音频数据帧：客户端发送音频数据至服务端，服务端通过内部接口(TSocket)与引擎交互，获取结果，校验组装后通过外部接口(sendMessage)至客户端。

**public** **void** **onMessage**(String msg,Session session);

开始帧：客户端发送开始帧到服务端，服务端通过内部接口(Redis)与引擎交互，建立会话。

结束帧：客户端发送结束帧到服务端，服务端通过内部接口(TSocket)与引擎交互，并调用外部接口(onClose)主动与客户端断开连接。

**public** **void** **onError**(Session session, Throwable error);

异常：若连接发生异常，将通过外部接口(onError)通知服务端。

**public** **void** **sendMessage**(String message,Session session,String sid);

推送数据：服务端通过内部接口(TSocket)与引擎交互后，若有识别结果，通过此接口将数据推送至客户端。

内部接口相关介绍：

能力接口将客户端发送的请求数据经过处理，发送至在线识别引擎,并解析在线识别结果后，组装成相应的格式返回给客户端。

Redis接口：

建立会话，将会话信息按照下列格式组装成json字符串，放入redis的List队列，由有空闲线程的引擎取走。

|  |  |  |
| --- | --- | --- |
| 接口参数 | 数据格式 | 参数说明 |
| request | JSON字符串 |  |
| cmd | 字符串 | 命令类型 |
| auf | 字符串 | 音频采样 |
| aue | 字符串 | 音频格式 |
| sid | 字符串 | 请求的唯一标识号 |

TSocket接口：

负责与引擎交互，数据帧发送数据/获取结果，结束帧关闭会话。

|  |  |  |
| --- | --- | --- |
| **数据帧请求** | | |
| 接口参数 | 数据格式 | 参数说明 |
| request | JSON字符串 |  |
| cmd | 字符串 | 命令类型 |
| syncid | 字符串 | 同步标识 |
| sid | 字符串 | 请求的唯一标识号 |
| data | 字符串 | 音频数据 |

|  |  |  |
| --- | --- | --- |
| **结束帧请求** | | |
| 接口参数 | 数据格式 | 参数说明 |
| request | JSON字符串 |  |
| cmd | 字符串 | 命令类型 |
| sid | 字符串 | 请求的唯一标识号 |

# 4.数据设计

4

## 4.1 WebSocket接口请求

4.1.1 开始帧

WebSocket连接成功后，发送开始帧开始语音识别，帧的类型（Opcode） 是Text， 使用json序列化。

开始帧参数如下：

| 字段名 | 数据类型 | 必选项 | 描述 |
| --- | --- | --- | --- |
| action | 字符串 | 是 | START，开始帧的类型 |
| type | 字符串 | 是 | 识别语音类型  1:普通话 |
| auf | 字符串 | 是 | 音频采样，支持如下格式：  audio/L16;rate=8000  audio/L16;rate=16000 |

4.1.2 音频数据帧

开始帧发送完成后，发送音频数据帧，帧的类型（Opcode）是Binary，内容是二进制的音频内容。除最后一个音频数据帧，每个帧的音频数 据长度建议为100ms，音频数据采用网络字节序发送。

注意：

1.建议音频流每100ms发送1600（8k16bit）/3200(16k16bit)字节；

4.1.3 结束帧

最后一个音频数据帧发送结束后，发送结束帧， 帧的类型（Opcode） 是Text， 使用json序列化。

结束帧参数如下：

| 字段名 | 数据类型 | 必选项 | 描述 |
| --- | --- | --- | --- |
| action | 字符串 | 是 | STOP，结束帧的类型 |

4.1.4 接收帧

语音识别服务对音频数据帧进行实时识别，返回文本识别结果，帧的类型（Opcode）是Text， 使用json序列化。

接收结果帧参数如下（成功）：

| **参数** | **类型** | **说明** |
| --- | --- | --- |
| action | string | 结果标识，started:开始帧，result:结果，error:异常 |
| code | string | 结果码(具体见[错误码](https://www.xfyun.cn/doc/asr/rtasr/API.html" \l "%E9%94%99%E8%AF%AF%E7%A0%81)) |
| data | string | 结果数据 |
| desc | string | 描述 |
| sid | string | 会话ID |

其中sid字段主要用于DEBUG追查问题，如果出现问题，可以提供sid帮助确认问题。

结果数据（data）

| **字段** | **含义** | **描述** |
| --- | --- | --- |
| bg | 句子开始时间 | 前端点帧数 |
| ed | 句子结束时间 | 后端点帧数 |
| engine\_name | 备用字段 | 备用字段 |
| epStatus | 引擎端点状态 | 1：表示已找到前端点2：表示已找到后端点 |
| pags | 识别结果状态 | 0：表示中间结果1：表示最终结果 |
| phoneme | 分词 | 分词 |
| time | 分词时间 | 分词时间 |
| recStatus | 引擎识别结果状态 | 0:中间识别结果；5-最终识别结果 |
| result | 识别结果 | 识别结果 |
| status | 引擎状态备用字段 | 目前同recstatus |
| syncid | 同步标识 | 包序 |