

内网里独立部署的 录音文件自动语音识别 （ASR，语音转文字） 解决方案



北京力盈力泰科技有限公司

2025



目录

一、背景.....	3
二、目标.....	3
三、实现方法和网络结构图.....	4
四、优势.....	5
五、关键参数.....	5

一、背景

在很多行业，对电话、无线电中的双方通话以及密闭空间里的会议、竞标、宣判等场景的说话，都会录音，存档做记录或证据。

在需要获取并了解录音的内容时，就需要人工去回放倾听，有些场合还需要将内容敲打成文字，记录下来。在录音文件比较多时，就需要花费很多人力，而且不可避免可能会出现差错。

在计算机科技飞速发展的当下，比较合理的解决方法是用自动语音识别（ASR）引擎对这些录音文件进行自动文字识别并保存下来。

然而目前的市场现状是可购买的独立部署的识别引擎价格昂贵且安装繁琐，需要厂家到现场去安装。而用公网上的云平台识别引擎开发的系统，按次数或者年收费，虽然价格不算高，但是不保密，数据不安全；录音文件上传到云平台上进行识别，相当于数据发给了对方。对一些比较敏感的行业来讲，这是没法接受的。

在这种情况下，从保密性、成本以及使用便捷这几个方面兼顾作为出发点，力盈力泰推出了自己的内网里部署的录音文件自动语音识别（ASR，文字转语音）解决方案。

二、目标

给本公司或第三方的录音系统生成的录音文件，提供一个安全、便捷、可靠、低成本的自动语音识别（语音转文字）解决方案。

这有助于最终用户能够快速掌握海量录音文件的内容，或者从中定位一些关键词，做后续处理。

三、实现方法和网络结构图

在内网部署 1 至多台力盈力泰语音识别服务器；找一台 windows 电脑，运行力盈力泰录音文件提交识别软件。

力盈力泰录音文件提交识别软件对接录音系统的数据库，从中获取到录音文件，自动提交给内网的力盈力泰语音识别服务器，在得到识别结果的文字内容后，写回到数据库对应的记录中。

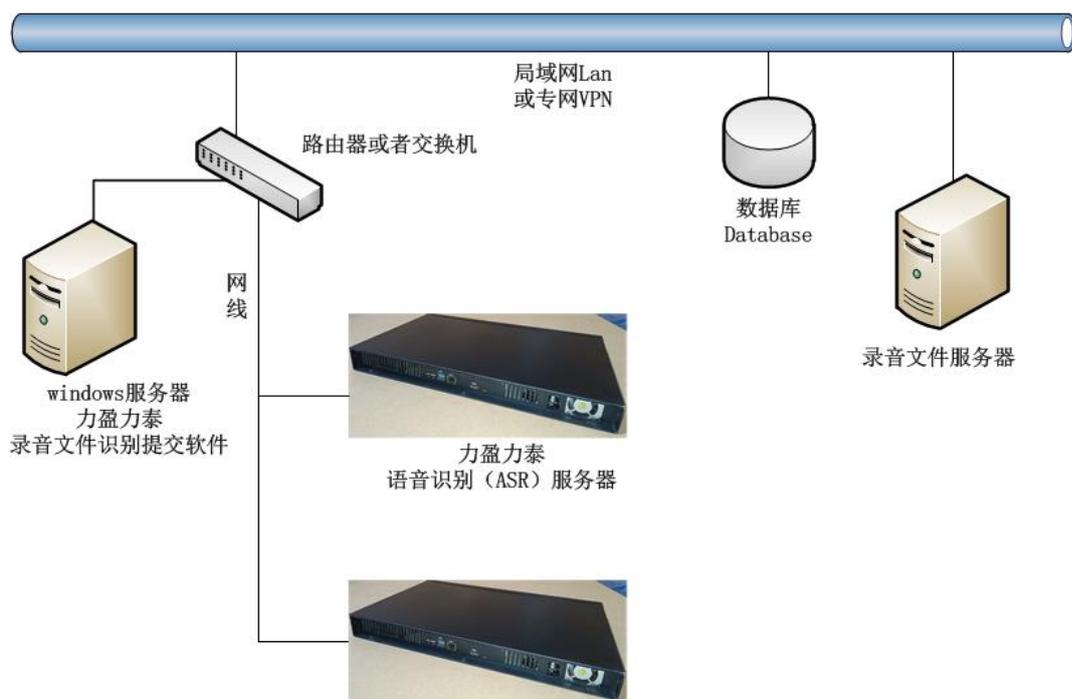


图 1.1 力盈力泰录音文件自动语音识别（ASR，文字转语音）网络结构图

四、优势

1. **保密性高。**和其他使用云服务器上的识别引擎方案相比，本系统在内网里运行，录音文件的数据是提交到内网里所部署的 ASR 服务器进行识别。无需连接到外网，不用担心网络风险；录音文件的语音数据不用上传到外网识别，不用担心数据被截留、复制而泄密。
2. **录音识别准确率高。**采用优秀的开源识别引擎，在录音没有太大噪音的情况下，普通话常用语的识别准确率能在 90% 以上。
3. **安装简单。**将力盈力泰 ASR 服务器寄到现场后，只需配置其 IP 地址就能使用。力盈力泰录音文件提交识别软件是绿色免安装版，只需拷贝到 windows 电脑上，配置几个参数，对接录音系统数据库后就能正常运行。
4. **成本可控。**和外网云平台识别方案一般按次数或者每年按套餐收费不同，本系统一次性采购后，后续没有额外费用（设备损坏需要维修或者更换除外），不会因录音文件的数量增多或者年限而增加费用。
5. **支持算力扩展。**系统支持多台 SR 服务器并发运行。后期如果增加了更多的录音通道，有更多的录音文件需要识别，到时可以根据需要，再增加 SR 服务器的数量。
6. **可对接大部分录音系统。**采用数据库对接方式，可配置对接的数据库字段，除了支持本公司的录音系统之外，还支持其它采用数据库方式的第三方录音系统。

五、关键参数

1. 支持对普通话和英文的录音文件进行识别。
2. 支持常见的录音文件格式，包括 wav 和 mp3 等。
3. 单个录音文件最大可支持 80M 字节。

4. 识别得到的文字结果支持多语言，可以保存成简体中文、繁体中文和英文。
5. 识别得到的文字结果支持采用 UTF-8 和 GBK 这两种字符集。