“英语模考软件”个人报告

**余常宏**

**12307130226**

这次《多媒体技术》课的Project我跟另外三个小伙伴一块儿做一个英语模考软件，我负责的是语音识别的模块，也是相对比较新颖的内容。

我们考虑到作为模考软件，想要对口语进行判分是不太现实的事，因为计算机没有一套合理的打分标准，无法针对每个人说的内容直接给出一个分数，那样不客观也没有意义，而且世界上无论是中高口译还是托福雅思，它们的口语都是由人来判分的，所以我们并未把口语纳入到打分中来，而是仅仅利用语音识别技术来对应试者的口语进行一个粗略的评价，即能顺利读出所给英文并被语音所正确识别的，我们根据识别后返回到本地的文字准确度对应试者的口语人为地作一个评价。

在组长完成了软件大体的框架构建之后，我开始着手干我的speaking模块。对于这块内容，我并不陌生，因为之前对于语音识别我还是有许多关注的，从苹果设备上的Siri到google语音我都亲自体验过，也大致对现代语音识别的发展水平有了一个了解。当然，作为一个小软件，我们肯定是要想办法利用google语音识别的API来帮助自己完成语音和文本的转换，因为世界上开源的语音识别技术中只有google的语音识别最成熟也最多人测试过。我们也尝试过百度的语音，但由于百度去年才对外公开百度语音接口，网上对于这方面的资料并不多，我们也只得放弃。

VS2013的新版本里面功能很全，这极大地方便了软件的设计，很多东西我们不再需要自己编程，而只需要引用自带的库文件和拖放各种控件，再分别对各种小控件进行功能细分。我这边主要就是有两个文字框，上面的“Please read”是要把默认文件夹目录下的txt文本中的文字显示在文字框里，这就是应试者所需要读的英文，然后下边的“Your result”是上传到谷歌语音以后返回的文字显示框。

所以，在各部分逻辑理清楚以后，就得开始编程了，这又是整个模块的重点部分。由于本人编程能力有限，这种程序还是第一次接触，所以不得不多番向不同的同学请教C#的许多语法和用法，还需要上网找找很多范例，来提高自己的编程能力，这又是一个漫长的过程。我需要先实现在文本框中显示txt中的文字，这就要求指针能链接到文件所在的子目录下（这边设定的文字目录为Bin->Debug->Audio\_essay），并将Microsoft的vs2013与其进行关联，这样程序才能成功通过调试，为此我还是费了好大一番功夫。这个完成之后，就得实现录音功能，我这边用的是windows自带的录音功能，并保存为wav格式（Bin->Debug->audio目录下），由于这个网上有比较详细的范例，我只需要照着进行一定的修改即可。最后就是核心部分了，将保存好应试者的语音上传至：<http://www.google.com/speech-api/v1/recognize?xjerr=1&client=chromium&lang=zh-CN>，即谷歌语音的api接口，然后返回相应的文字至本地，并且显示在“Your result”的文字框内，这一块的请求程序也能够从网上得到一些提示，但是调试起来并不是那么容易。

程序主要就是直接通过http请求，然后获取返回的识别结果，不过google默认的是audio/x-flac的，支持flac（无损音频压缩编码）格式，要想支持wav、mp3等格式，在网上查过之后说只能通过格式转换，如果是直接提交wav文件是无法识别的；后来受到某些网友的启发，他们说可以这样写：audio/L16，然后我就尝试着把google的也改成这样，不出所料可以支持wav格式了，这只说明google是有支持的，可能只是有些参数由于google对API的封闭，我们不清楚而已。尽管如此，这只解决了格式问题。后来大陆可能在6月份对google进行了更为全面的封杀，我的工作进展也受到了不小的影响。为此，我不得不求助于小组的其他成员们，后来又去找了计算机学院的同学，利用他的VPN继续帮我测试程序，可奇怪的是后来的返回文字老是出错或者出现各种乱码，我们也讨论过很多次，对于可能的解决方案一一进行了试验，也没能得出个所以然，无法确定究竟是自己的程序出了问题还是google那边出了状况。后又因为时间上的限制，我的语音识别最终也只能做到这里了，不得不说是个遗憾吧。

总之，这次的PJ总的来说还是学到了很多新的东西，自己的动手能力也有了一定的提高，同时也认识到自己不会的还有很多，在学习的道路上永远没有止境。