

## 开发环境

电脑环境：Windows 8.1 64 位，IDE：Microsoft Visual Studio 2015

### Windows 版本

Windows 8.1 专业版

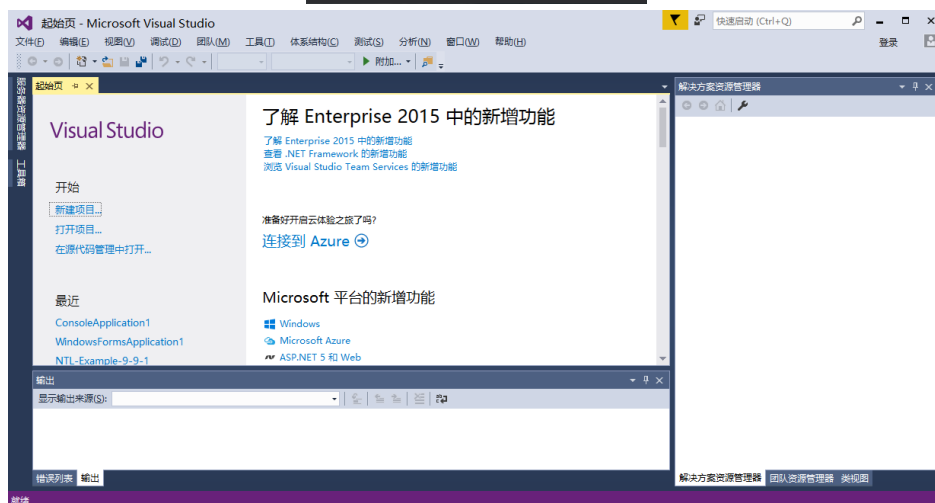
© 2013 Microsoft Corporation。保留  
所有权利。

[获取新版本的 Windows 的更多功能](#)



### 系统

处理器:	Intel(R) Core(TM) i5-4210U CPU @ 1.70GHz 2.40 GHz
安装内存(RAM):	8.00 GB (7.88 GB 可用)
系统类型:	64 位操作系统, 基于 x64 的处理器
笔和触摸:	没有可用于此显示器的笔或触控输入



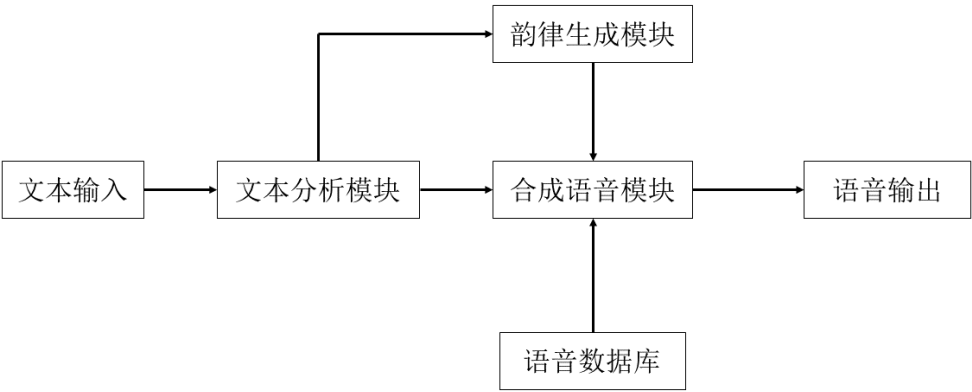
## “语音合成”技术学习小结

体验过讯飞输入法、讯飞语记，对讯飞的语音识别、语音合成功能印象深刻。语音合成，除了改进的算法之外，还需要分析海量语料库。科大讯飞在语音产业是领先的，尽管有时用讯飞语记听小说时，发音有些生硬。

在讯飞开放平台，体验了各类发言人线上 demo。语音合成发音有 14 种，涵盖了不同性别、不同方言、不同年龄、不同语言的发音；特色合成的发音就更多了，不同情感、不同播报人、不同情景(小说、故事)、不同娱乐角色的发音。能在此次项目中，接触最前沿的人工智能与智能语音技术，对于还是在校大学生的我，深感荣幸。

语音合成，又称文语转换(Text to Speech, TTS)技术，涉及声学、语言学、数字信号处理、计算机科学等多个学科技术，它能够将计算机中任意出现的文字，转换成自然流畅的语音输出。它使得计算机不仅能够处理数据，显示图像和文字，还能像人一样的说话，从而使得计算机变得更为亲切、自然。

语音合成系统主要包括三部分：文本分析模块、韵律生成模块和合成语音模块，结构见下图。其中，合成语音模块是最基本、最重要的模块。



文本信息首先由文本输入模块读入；接着文本分析模块即文本的预处理模块，主要完成自动分词、多音字处理、特殊符号的转换、文本的切分等，然后将处理好的数据送入韵律生成模块和合成语音模块；韵律生成模块根据各项合成规则规划出目标高音、音长、音强、停顿及语调等，将规划的结果参数送入合成语音模块；合成语音模块利用合成算法合成出满足目标要求的音节波形数据，将其拼接成语音流数据送入语音输出模块，最后由语音输出模块输出语音。

语音合成技术经历了一个逐步发展的过程，从参数合成到拼接合成再到两者的逐步结合，其不断发展的动力是人们认知水平和需求的提高。目前，常用的语音合成技术主要有共振峰合成、发音器官参数合成、LPC 合成、PSOLA 技术和 LMA 声道模型技术。它们各有优缺点，人们在应用过程中往往将多种技术有机结合，或将一种技术的优点运用到另一种技术上，以克服单一技术的不足。

语音合成是对人类发音的模拟，能用到语音的地方都是语音合成技术的用武之地，其具有广阔的发展前景。不同语言之间的交流在信息社会和网络时代十分重要，多语种的语音合成有着独特的应用价值。面对人机对话问题，不同说话风格(如评书、新闻、娱乐等)的合成技术显得非常必要。情感语音合成是目前国内外的研究热点，现有的研究多以韵律和声学特征为指导因素。另外，进一步提高语音合成的自然度、可视语音等都是语音合成技术的发展方向。