

认知服务调用

该课程主要为大家讲授如下的内容：

- Azure认知服务简介
- 认知服务调用方法
- 练习
- 实战

1. Azure认知服务简介

1. 微软认知服务分类

微软认知服务服务主要包括Vision（影像）、Speech（语音）、Language（语言）以及Decision（决策）四大类。

影像服务：让应用能够理解所看到的图片和视频。

语音服务：让应用可以倾听，和客户交流。

自然语言：提供基本的沟通能力，通过用户日常对话的表达方式，让应用了解他想要做什么。

决策：在纷繁的数据中，帮用户提炼出一些关键的特征，辅助做出更睿智的决定。

2. 微软认知服务特点

微软认知服务具有简单、灵活、成熟的特点。



为什么选择微软认知服务？

简单	灵活	成熟
提供通用易集成的REST APIs； 几行代码就能轻松开发；	支持主流平台和开发语言，按你所需自主选择； 丰富的服务为你的应用提供最佳API；	服务由来自Microsoft Research, Bing, and Azure Machine Learning的专家为你构建； 高质量的文档、例子代码和社区支持；

GET A KEY BUILD

python nodejs

GitHub stackoverflow msdn uservoice

2. 认知服务调用方法

第一步，在Azure云上创建认知服务。

第二步，安装相应SDK，或采用REST API的调用方式。

第三步，在应用中利用SDK或REST API调用所需得认知服务。



3. 练习

接下来，用语音的服务进行实际练习，在使用TTS服务的时候需要对这个服务进行一些配置。

1. 基本配置

首先，作为访问语音服务需要有一个语音服务实例的密钥Key，以及这个语音实例在Azure的哪个地区，就是region这个参数。

其次，语音设置使用中文zh-CN。

最后，指定用哪种声音来表达。

```
1 基本配置项
1.1 语音服务的配置

speech_key, speech_region = "[access key]", "[region]"
speech_config = speechsdk.SpeechConfig(subscription=speech_key, region=speech_region)

1.2 语言的配置

language = "zh-CN" #default en-US

#转换成语音时候，选择需要转换的语言
speech_config.speech_synthesis_language = language

目前微软认知服务支持119种语言，具体参考 支持语言list

1.3 语音的配置
对于不同的语言会存在不同种类的voice 供调用者使用，可以根据需求任意切换自己需要的声音

#voice配置
voice = "Microsoft Server Speech Text to Speech Voice (zh-CN, YunxiNeural)"
speech_config.speech_synthesis_voice_name = voice
```

2. 高级配置

voice语音在基础配置里同样有，但在SSML的加持下，可以很容易在一段文字里进行语音切换。

2 高级配置：SSML（语音合成标记语言）使用

这里是一些高级选项，需要使用SSML来转换语音了，SSML结构如下

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xmlns:mstts="https://www.w3.org/2001/mstts" xml:lang="zh-CN">
...
</speak>
```

最基本的SSML结构体，lang用来指定语言

2.1 语音 voice

voice参数的name就是用来制定讲话语音的，简单列举一些常用的

- zh-CN-XiaohanNeural
- zh-CN-XiaomoNeural
- zh-CN-XiaoxuanNeural
- zh-CN-XiaoruiNeural
- en-US-AriaNeural
- en-US-JennyNeural

使用语音合成标记语言SSML

style属性，指讲话风格。

styledegree可以配置讲话风格的强度。

2.2 讲话风格配置 -express-as

为了更进一步描述讲话风格，微软定义了新标记mstts:express-as，可以用来更精确的表达语言的轻重、甚至扮演不同年龄的角色等。

```
<voice name="zh-CN-XiaomoNeural">
  <mstts:express-as style="cheerful" role="Girl" styledegree="1">
    你好，欢迎您使用微软认知服务
  </mstts:express-as>
</voice>
```

注意使用express-as标签需要引入命名空间 xmlns:mstts

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xmlns:mstts="https://www.w3.org/2001/mstts">
...
</speak>
```

属性	说明	必需/可选
style	指定讲话风格。目前，讲话风格特定于语音。	如果调整神经语音的讲话风格，则此属性是必需的。如果使用mstts:express-as，则必须提供风格。如果提供无效的值，将忽略此元素。
styledegree	指定说话风格的强度。接受的值：0.01 到 2（含边界值）。默认值为 1，表示预定义的风格强度。最小单位为 0.01，表示略倾向于目标风格。值为 2 表示是默认风格强度的两倍。	可选（目前，styledegree 仅支持中文（普通话，简体）神经语音。）

Role属性，用来模拟男孩、女孩、老人等风格。

Role 属性

角色	说明
role="Girl"	该语音模拟女孩。
role="Boy"	该语音模拟男孩。
role="YoungAdultFemale"	该语音模拟年轻成年女性。
role="YoungAdultMale"	该语音模拟年轻成年男性。
role="OlderAdultFemale"	该语音模拟年长的成年女性。
role="OlderAdultMale"	该语音模拟年长的成年男性。
role="SeniorFemale"	该语音模拟老年女性。
role="SeniorMale"	该语音模拟老年男性。

例子：

```
txt = "<speak version='1.0' xmlns='http://www.w3.org/2001/10/synthesis'"
txt += " xmlns:mstts='https://www.w3.org/2001/mstts' xml:lang='zh-CN'"
txt += "<voice name='zh-CN-XiaoxiaoNeural'"
txt += " <mstts:express-as style='sad' styledegree='2'"
txt += " 快走吧，路上一定要注意安全，早去早回。"
```

语言参数，在基础配置也有，这里可以方便进行语言切换。

2.3 调整语言-lang

SSML结构中使用lang标签

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
  xmlns:mstts="https://www.w3.org/2001/mstts" xml:lang="en-US">
  <voice name="en-US-JennyMultilingualNeural">
    I am looking forward to the exciting things.
    <lang xml:lang="zh-CN">
      你好, 欢迎您使用微软认知服务
    </lang>
    <lang xml:lang="en-US">
      Welcome to use azure cognitive services
    </lang>
  </voice>
</speak>
```

Silence是静音参数，控制不再发声。

2.4 静音处理-silence

标签mstts:silence 用来控制静音的

```
<mstts:silence type="Sentenceboundary" value="200ms"/>
```

属性使用

属性	说明	必需/可选
type	指定添加静音的位置: Leading - 在文本的开头Tailing - 在文本的结尾Sentenceboundary - 在相邻句子之间	必需
Value	指定暂停的绝对持续时间, 以秒或毫秒为单位; 该值应设为小于 5000 毫秒。例如, 2s 和 500ms 是有效值	必需

具体用例

```
txt = "<speak version='1.0' xmlns='http://www.w3.org/2001/10/synthesis'"
txt += "      xmlns:mstts='https://www.w3.org/2001/mstts' xml:lang='zh-CN'"
txt += "<voice name='zh-CN-XiaoxiaoNeural'"
txt += "  <mstts:silence type='Sentenceboundary' value='5s'/"
```

Prosody参数控制音调和语速调节，它通过两个参数实现。

Pitch设置音调，可以用预定义的常量，也可以直接使用频率赫兹值。

Rate设置语速，来控制说话的快慢。

2.5 语速音量的调整-prosody

prosody 用来调整速率，音量，持续时间等声音参数

```
<prosody pitch="value" rate="value" ...>
</prosody>
```

pitch 参数

指示文本的基线音节。可将音调表述为

- 以某个数字后接"Hz" (赫兹) 表示的绝对值。例如 <prosody pitch="600Hz">some text</prosody>。
- 以前面带有"+"或"-"的数字，后接"Hz"或"st" (用于指定音节的变化量) 表示的相对值。例如 <prosody pitch="+80Hz">some text</prosody> 或 <prosody pitch="-2st">some text</prosody>。"st"表示变化单位为半音，即，标准全音阶中的半调 (半步)。
- 常量值：
 - x-low
 - low
 - medium
 - high
 - x-high
 - default

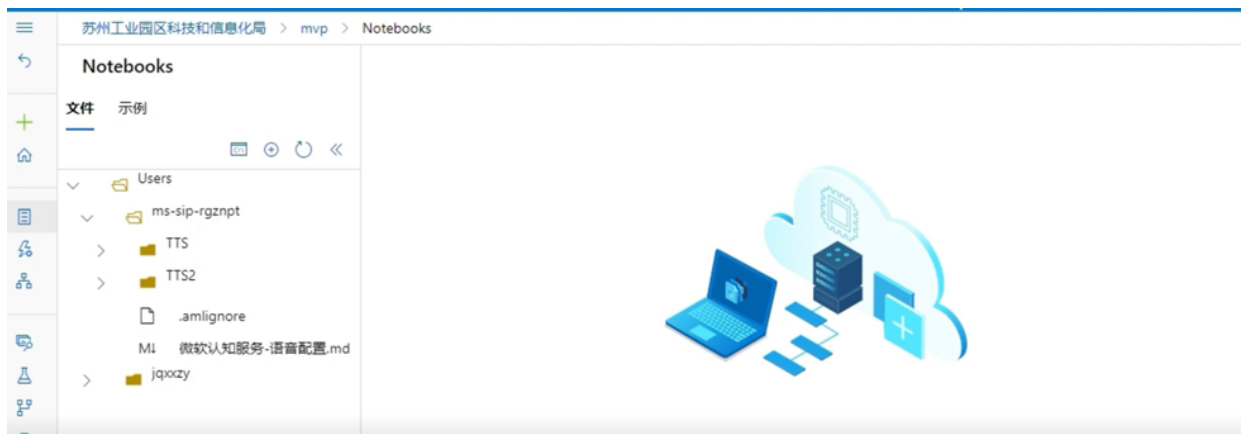
4. 实战

在浏览器上打开Azure Machine Learning Studio，网址

<https://studio.ml.azure.cn/>。



点击Notebooks菜单项进入相应页面，两个例子代码文件分别在TTS和TTS2目录下。



点击tts.ipynb，进入代码编辑窗，点击运行，就可以查看代码运行结果。

如果要下载文件，右键点击output.wav，出现的弹出菜单里选择下载。

