AIT_MachineLearning_Usage_1_微软认知服务使用-语音配置

微软认知服务实训-语音配置

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通过种类的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
```

全部支持语音参考：https://aka.ms/csspeech/voicenames

中文支持如下几种

中文（粤语，繁体)	`zh-HK`	**Female**	`zh-HK-HiuGaaiNeural`	**常规**
中文（粤语，繁体)	`zh-HK`	Female	`zh-HK-HiuMaanNeural`	常规
中文(粤语，繁体)	`zh-HK`	男	`zh-HK-WanLungNeural`	常规
中文（普通话，简体)	`zh-CN`	女	`zh-CN-XiaohanNeural`	常规，使用 SSML 提供多种风格
中文（普通话，简体)	`zh-CN`	女	`zh-CN-XiaomoNeural`	常规，使用 SSML 提供多种角色扮演和风格
中文（普通话，简体)	`zh-CN`	女	`zh-CN-XiaoruiNeural`	高级语音，使用 SSML 提供多种风格
中文（普通话，简体)	`zh-CN`	女	`zh-CN-XiaoxiaoNeural`	常规，使用 SSML 提供多种语音风格
中文（普通话，简体)	`zh-CN`	女	`zh-CN-XiaoxuanNeural`	常规，使用 SSML 提供多种角色扮演和风格
中文（普通话，简体)	`zh-CN`	女	`zh-CN-XiaoyouNeural`	儿童语音，针对讲故事进行了优化
中文（普通话，简体)	`zh-CN`	男	`zh-CN-YunxiNeural`	常规，使用 SSML 提供多种风格
中文（普通话，简体)	`zh-CN`	男	`zh-CN-YunyangNeural`	针对新闻阅读进行了优化，使用 SSML 提供多种语音风格
中文（普通话，简	`zh-CN`	男	`zh-CN-YunyeNeural`	针对讲故事进行了优化

中文（粤语，繁体)	**zh-HK**	**Female**	**zh-HK-HiuGaaiNeural**	**常规**
体)				
中文(台湾普通话)	zh-TW	Female	zh-TW-HsiaoChenNeural	常规
中文(台湾普通话)	zh-TW	Female	zh-TW-HsiaoYuNeural	常规
中文(台湾普通话)	zh-TW	男	zh-TW-YunJheNeural	常规

另外，如果官方提供的声音都不能满足，你可以自己定义声音的

详细参看：https://aka.ms/customvoice

```
speech_config = speechsdk.SpeechConfig(subscription=speech_key,
region=service_region)
speech_config.endpoint_id = "YourEndpointId"
speech_config.speech_synthesis_voice_name = "YourVoiceName"
```

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>
```

最基本的一个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

```
#SSML文字转语音
txt = "<speak version=\"1.0\"
xmlns=\"http://www.w3.org/2001/10/synthesis\" xml:lang=\"zh-
CN\">"
txt += " <voice name=\"zh-CN-XiaomoNeural\">"
txt += "   你好，欢迎是用微软认知服务"
txt += " </voice>"
txt += " <voice name=\"zh-CN-YunxiNeural\">"
txt += "   你好，欢迎是用微软认知服务"
txt += " </voice>"
txt += "</speak>"
result = speech_synthesizer.speak_ssml_async(txt).get()
```

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`	指定讲话角色扮演。 语音将充当不同的年龄和性别，但语音名称不会更改。	可选（`role` 仅支持 zh-CN-XiaomoNeural 和 zh-CN-XiaoxuanNeural。）

Style 属性

`zh-CN-XiaomoNeural`	`style="calm"`	**以沉着冷静的态度说话。 语气、音调、韵律与其他语音类型相比要统一得多。**
	`style="cheerful"`	以较高的音调和音量表达欢快、热情的语气
	`style="angry"`	以较低的音调、较高的强度和较高的音量来表达恼怒的语气。 说话者处于愤怒、生气和被冒犯的状态。
	`style="fearful"`	以较高的音调、较高的音量和较快的语速来表达恐惧、紧张

`zh-CN-XiaomoNeural`	`style="calm"`	**以沉着冷静的态度说话。 语气、音调、韵律与其他语音类型相比要统一得多。**
		的语气。 说话者处于紧张和不安的状态。
	`style="disgruntled"`	表达轻蔑和抱怨的语气。 这种情绪的语音表现出不悦和蔑视。
	`style="serious"`	表达严肃和命令的语气。 说话者的声音通常比较僵硬，节奏也不那么轻松。
	`style="depressed"`	调低音调和音量来表达忧郁、沮丧的语气
	`style="gentle"`	以较低的音调和音量表达温和、礼貌和愉快的语气

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。"
txt += " </mstts:express-as>"
txt += "</voice>"
txt += "</speak>"
result = speech_synthesizer.speak_ssml_async(txt).get()
```

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>
```

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=\"en-US\">"
txt += "    <voice name=\"en-US-JennyMultilingualNeural\">"
txt += "        I am looking forward to the exciting things."
txt += "        <lang xml:lang=\"zh-CN\">"
txt += "            你好，欢迎您使用微软认知服务"
txt += "        </lang>"
txt += "        <lang xml:lang=\"ja-JP\">"
txt += "            你好，欢迎您使用微软认知服务"
```

```
txt += "        </lang>"
txt += "    </voice>"
txt += "</speak>"
result = speech_synthesizer.speak_ssml_async(txt).get()
```

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	必须
`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	必须

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\"/>"
    txt += "    我们快点走吧，看天气要下雨了。"
    txt += "    不行，我必须这些东西打包好，"
    txt += "    那好吧，我等你一起"
    txt += "</voice>"
    txt += "</speak>"
    result = speech_synthesizer.speak_ssml_async(txt).get()
```

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

rate 参数

rate 指示文本的讲出速率。 可将 rate 表述为：

- 以充当默认值倍数的数字表示的相对值。 例如，如果值为 *1*，则速率不会变化。 如果值为 *0.5*，则速率会减慢一半。 如果值为 *3*，则速率为三倍。
- 常量值：
  - x-slow
  - slow
  - medium
  - fast

- x-fast
- default

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 += "        女儿看见父亲走了进来，问道："
    txt += "        <mstts:express-as role=\"YoungAdultFemale\"
style=\"cheerful\">"
    txt += "            您来的挺快的，怎么过来的？"
    txt += "        </mstts:express-as>"
    txt += "        父亲放下手提包，说："
    txt += "    </voice>"
    txt += "    <voice name=\"zh-CN-YunxiNeural\">"
    txt += "       <prosody rate=\"0.8\" pitch=\"low\">"
    txt += "        <mstts:express-as style=\"sad\"
styledegree=\"2\"  >"
    txt += "           刚打车过来的，一路还算顺利。"
    txt += "        </mstts:express-as>"
    txt += "     </prosody>"
    txt += "    </voice>"
    txt += "</speak>"

    result = speech_synthesizer.speak_ssml_async(txt).get()
```

官方提供了一个用来测试的UI Demo，在线测试

课堂练习：

练习1 TTS基础

要求：
个人练习，通过TTS将音频保存成3个不同的wav文件

1）参照TTS例子代码，调整语言、语音等参数；
2）尽可能多的使用我们学习到的参数配置；

练习2 TTS高级

要求：
个人练习：使用恰当的配置来朗读如下一段童话故事,将音频保存成1个wav文件

严冬时节，鹅毛一样的大雪片在天空中到处飞舞着，有一个王后坐在王宫里的一扇窗子边，正在为她的女儿做针线活儿，寒风卷着雪片飘进了窗子，乌木窗台上飘落了不少雪花。 她抬头向窗外望去，一不留神，针刺进了她的手指，红红的鲜血从针口流了出来，有三点血滴落在飘进窗子的雪花上。 她若有所思地凝视着点缀在白雪上的鲜红血滴，又看了看乌木窗台，说道："但愿我小女儿的皮肤长得白里透红，看起来就像这洁白的雪和鲜红的血一样，那么艳丽，那么骄嫩，头发长得就像这窗子的乌木一般又黑又亮！"
她的小女儿渐渐长大了，小姑娘长得水灵灵的，真是人见人爱，美丽动人。 她的皮肤真的就像雪一样的白嫩，又透着血一样的红润，头发像乌木一样的黑亮。 所以王后给她取了个名字，叫白雪公主。

练习3 TTS高级 - 情景对话

要求：
小组练习：使用SSML语言来编写一个情景对话，通过TTS将对话保存成1个wav文件

1）尽可能多的使用我们学习到的标记

2）对话内容控制在8句话以内

3）对话内容取材，自由发挥。