

HTML5语音合成Speech Synthesis API简介

这篇文章发布于 2017年01月13日, 星期五, 02:20, 归类于 [JS相关](#)。阅读 34042 次, 今日 21 次 [21 条评论](#)

by zhangxinxu from <http://www.zhangxinxu.com/wordpress/?p=5865>

本文可全文转载, 但需得到原作者书面许可, 同时保留原作者和出处, 摘要引流则随意。

一、关于HTML5语音Web Speech API

HTML5中和Web Speech相关的API实际上有两类, 一类是“语音识别(Speech Recognition)”, 另外一个就是“语音合成(Speech Synthesis)”, 这两个名词听上去很高大上, 实际上指的分别是“语音转文字”, 和“文字变语音”。

而本文要介绍的就是这里的“语音合成-文字变语音”。为什么称为“合成”呢? 比方说你Siri发音“你好, 世界!” 实际上是把“你”、“好”、“世”、“界”这4个字的读音给合并在一起, 因此, 称为“语音合成”。

“语音识别”和“语音合成”看上去像是正反两方面, 应该带有镜面气质, 实际上, 至少从兼容性来看, 两者并无法直接对等。

“语音识别(Speech Recognition)”目前的就Chrome浏览器和死忠小弟Opera浏览器默认支持, 并且需要 `webkit` 私有前缀:

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android *	Chrome for Android
			49 51 54					4.3 4.4 4.4.4	
11	14	50	55	10	42	10.1	all	53	55
	15	51 52 53	56 57 58	TP	43 44				

张 <http://www.zhangxinxu.com>
鑫旭-鑫空间-鑫生活

但是, “语音合成(Speech Synthesis)”的兼容性要好上太多了, 如下图:

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android *	Chrome for Android
			49 51 54					4.3 4.4 4.4.4	
11	14	50	55	10	42	10.1	all	53	55
	15	51 52 53	56 57 58	TP	43 44				

张 <http://www.zhangxinxu.com>
鑫旭-鑫空间-鑫生活

所以, 本文主要介绍下理论上更为适用的HTML5 Speech Synthesis API。不过在此之前, Speech Recogniti

on API还是简单提一下。

Speech Recognition API也就是语音识别功能，需要麦克风等音频输入设备的支持，在Chrome浏览器下，实际上可以添加简单的属性就可以让部分控件有语音识别功能，一行JS都不需要写，这个我之前有专门写文章介绍过：“[渐进使用HTML5语言识别, so easy!](#)”

就是在输入框上加一个 `x-webkit-speech` 属性即可，例如：

```
<input x-webkit-speech />
```

然而，我刚刚打开demo页面一测试，发现原来有的麦克风（下图为以前截图）居然不见了。。。看来已经被chrome无情抛弃了！



好吧，就当我不上面的内容，什么都没说过。不过有一点是可以肯定的，就是原来输入框上的那个语音识别底层就是使用的Speech Recognition API，因此存在一定的共性，比方说文字内容识别需要google服务器返回，因此功能就与网络环境有很大关系，比方说google被墙了，或者网速很慢了，都有可能导致识别出现异常。

使用的基本套路如下：

1. 创建SpeechRecognition的新实例。由于到目前为止，浏览器还没有广泛支持，所以需要 `webkit` 的前缀：

```
var newRecognition = webkitSpeechRecognition();
```

2. 设置是持续听还是听到声音之后就关闭接收。通过设置 `continuous` 属性值实现。一般聊天沟通使用 `false` 属性值，如果是写文章写公众号之类的则可以设置为 `true`，如下示意：

```
newRecognition.continuous = true;
```

3. 控制语音识别的开启和停止，可以使用start()和stop()方法：

```
// 开启
newRecognition.start();

// 停止
newRecognition.stop();
```

4. 对识别到的结果进行处理，可以使用一些事件方法，比方说 `onresult`：

```
newRecognition.onresult = function(event) {
    console.log(event);
}
```

`event` 是个对象，我家里电脑不知道什么原因，无法成功返回识别内容，显示网络错误，可能因为墙的缘故：

```
> var recognition = new webkitSpeechRecognition();
  recognition.onresult = function(event) {
    console.log(event)
  }
  recognition.onerror = function(event) {
    console.log(event)
  }
  recognition.start();
< undefined

▼ SpeechRecognitionError VM328:6
  bubbles: false
  cancelBubble: false
  cancelable: false
  composed: false
  ▶ currentTarget: SpeechRecognition
  defaultPrevented: false
  error: "network"
  eventPhase: 0
  message: ""
  ▶ path: Array[0]
  returnValue: true
  ▶ srcElement: SpeechRecognition
  ▶ target: SpeechRecognition
  timeStamp: 25678.655000000002
  type: "error"
  ▶ __proto__: SpeechRecognitionError
```

所以，我从网上找了下大致数据结构：

```
{
  ..
  results: {
    0: {
      0: {
        confidence: 0.695017397403717,
        transcript: "你好，世界"
      },
      isFinal:true,
      length:1
    },
    length:1
  },
  ..
}
```

除了 `result` 事件外，还有其他一些事件，例如， `soundstart`、`speechstart`、`error` 等。

二、关于语音合成Speech Synthesis API

先从最简单的例子说起，如果想让浏览器读出“你好，世界！”的声音，可以下面的JS代码：

```
var utterThis = new window.SpeechSynthesisUtterance('你好，世界! ');
window.speechSynthesis.speak(utterThis);
```

没错，只需要这么一点代码就足够了，大家可以在自己浏览器的控制台里面运行上面两行代码，看看有没有读出声音。

上面代码出现了两个长长的对象，`SpeechSynthesisUtterance` 和 `speechSynthesis`，就是语音合成Speech Synthesis API的核心。

首先是 `SpeechSynthesisUtterance` 对象，主要用来构建语音合成实例，例如上面代码中的实例对象 `utterThis`。我们可以直接在构建的时候就把要读的文字内容写进去：

```
var utterThis = new window.SpeechSynthesisUtterance('你好, 世界!');
```

又或者是使用实例对象的一些属性，包括：

- `text` – 要合成的文字内容，字符串。
- `lang` – 使用的语言，字符串，例如：`"zh-cn"`
- `voiceURI` – 指定希望使用的声音和服务，字符串。
- `volume` – 声音的音量，区间范围是 `0` 到 `1`，默认是 `1`。
- `rate` – 语速，数值，默认值是 `1`，范围是 `0.1` 到 `10`，表示语速的倍数，例如 `2` 表示正常语速的两倍。
- `pitch` – 表示说话的音高，数值，范围从 `0`（最小）到 `2`（最大）。默认值为 `1`。

因此上面的代码也可以写作：

```
var utterThis = new window.SpeechSynthesisUtterance();  
utterThis.text = '你好, 世界!';
```

不仅如此，该实例对象还暴露了一些方法：

- `onstart` – 语音合成开始时候的回调。
- `onpause` – 语音合成暂停时候的回调。
- `onresume` – 语音合成重新开始时候的回调。
- `onend` – 语音合成结束时候的回调。

接下来是 `speechSynthesis` 对象，主要作用是触发行为，例如读，停，还原等：

- `speak()` – 只能接收 `SpeechSynthesisUtterance` 作为唯一的参数，作用是读合成的话语。
- `stop()` – 立即终止合成过程。
- `pause()` – 暂停合成过程。
- `resume()` – 重新开始合成过程。
- `getVoices` – 此方法不接受任何参数，用来返回浏览器支持的语音包列表，是个数组，例如，在我的电脑下，Firefox浏览器返回的语言包是两个：

而在chrome浏览器下，数量就很惊人了：

虽然数量很多，是有种给人中看不中用的感觉，为什么这么说呢！在我的chrome浏览器下，不知道为什么，不会读任何声音，但是同样的demo见面，公司的电脑就可以，我后来仔细查了一下，有可能(20%可能性)是我家里的电脑win7版本是阉割版，没有安装或配置TTS引擎。

手机Safari浏览器也不会读。

其中，17是普通话大陆：

另外，`getVoices` 的获取是个异步的过程，因此，你可以直接在控制台输入，`speechSynthesis.getVoices()` 返回的是一个空数组，没关系，多试几次，或者搞个定时器之类的。

三、语音合成Speech Synthesis API有什么用

对于盲人或弱视用户，往往会借助一些辅助设备或者软件访问我们的网页，其原理是通过触摸或定位某些元素发出声音，来让用户感知内容。

有了语音合成Speech Synthesis API，对于这类用户，以及开发人员自己，都会带来一定的便利性。首先对于视觉有障碍的用户，他们就不需要再安装其他软件或购买其他设备，就可以无障碍地访问我们的产品。对于开发人员自己，我们的无障碍建设可以变得更加灵活，不一定要百分百契合ARIA无障碍规范（可参见我之前文章“[WAI-ARIA无障碍网页应用属性完全展示](#)”），因为我们可以直接让浏览器合成我想要的语音内容，例如，VoiceOver在对一些标签读取的时候，总会附带一句“标志性内容”，就算对于我这样专业从业者，这个词也是有点生涩难懂的，我们其实可以把这个语音合成交给自己，使用更通俗易懂的词反馈给用户，我想体验上应该是更好的。

至少我会在17年，尝试在一些产品上推动这方面工作的建设。

另外一个就是二维码识别，有时候肉眼看得眼睛疼，加个按钮让用户听。

等等~

最后，放上一张今天别墅home party团建照一张：



感谢阅读，欢迎交流！

《CSS世界》签名版独家发售，包邮，可指定寄语，点击显示购买码

（本篇完） // 想要打赏？点击[这里](#)。有话要说？点击[这里](#)。



« [web移动端浮层滚动阻止window窗体滚动JS/CSS处理](#)

[分享一个即插即用的私藏缓动动画JS小算法](#) »

猜你喜欢

- 渐进使用HTML5语言识别, so easy!
- Chrome下语音识别实现页面快速引导以及重要交互
- WAI-ARIA无障碍网页应用属性完全展示
- HTML可访问性fieldset, legend元素及CSS布局应用
- 基于VoiceOver的移动web站无障碍访问实战
- 翻译-盲人如何使用互联网的8个误区
- span与a元素的键盘聚焦性以及键盘点击性研究
- 找到适合自己的前端发展方向
- HTML <area><map>标签及在实际开发中的应用
- HTML tabindex属性与web网页键盘无障碍访问
- HTML5 number类型文本框step属性的验证机制

分享到： 1

发表评论（目前21条评论）

<input type="text"/>	名称 (必须)
<input type="text"/>	邮件地址(不会被公开) (必须)
<input type="text"/>	网站
<div></div>	
<input type="button" value="提交评论"/>	

1. **hsapple**说道:

2018年09月26日 15:38

1700, 播放出来是: 一千七

如果不增加空格, 不修改这个串本身情况下, 有什么办法能设置他朗读一七零零吗?

就是播放数字, 不不放数量。

[回复](#)



2. **xcsweb**说道:

2018年05月17日 12:41

chrome 已经不行了, 没有声音

[回复](#)

达到说道:

2018年07月4日 10:39

为何我试的可以呢? 你chrome太低了吧

[回复](#)



3. **张sir**说道:

2017年11月2日 14:17

手机safari不读, android机上的浏览器有的读有的不读, 兼容性太差了还是需要开启什么设置? 公司想做一个基于web端的叫号系统, 本来想用这个的, 兼容性问题放弃了, 有什么好的解决方案吗? 百度tts也是一样的问题

[回复](#)



4. **knife**说道:

2017年09月6日 16:23

看来楼主对tts知道的还是太少, 微软的语音能不能读出来首先取决于是否安装了语音引擎, 其次是语言, 英文引擎是无法朗读中文的, 目前针对Windows系统, Window XP 的tts语音包需要单独安装, Windows 7 自带了2个, Windows 10 的Cortana是目前市面上语音引擎里发音最好听的且接近人声的, 只可惜默认没有开启, 只能Cortana使用, 需要修改注册表相关配置才可以开启。



回复

5. knife说道:

2017年09月06日 16:13

回复的代码内容被屏蔽移除了，可能怀疑是注入代码吧

回复



6. knife说道:

2017年09月06日 16:11

你忘记说多音字了

```
var utterThis = new window.SpeechSynthesisUtterance("贾宝玉,商贾");  
window.speechSynthesis.speak(utterThis);
```

回复



7. colin说道:

2017年08月14日 18:20

不错鑫旭，哈哈，这么多年了，无意中看到了你的网站，跟sama到一个公司了啊。

回复



8. bestRenekton说道:

2017年02月13日 15:21

chrome.,刚开始可以读出来，， 怎么多刷新几次就不读了

回复



9. 熊猫桑说道:

2017年01月23日 18:36

Win10的chrome也是蜜汁静默.....后来发现只有[MDN的Demo](<http://mdn.github.io/web-speech-api/speak-easy-synthesis/>)是可以让chrome成功发声的。另外，Win10的Metro版控制面板能成功看到3个语音包，但旧式控制面板只能看到2个语音包（而且还是1中1英），且浏览器是以旧式控制面板为准的.....解决方法貌似只有改注册表，把另外俩语音包的信息加过去就行了。在MDN的Demo上，FireFox能自由选择用哪个语音包（跟旧式控制面板看到的是一致的）但Chrome就只有一个native（就是系统默认的）其余都是Google系的确有点中看不中用的感脚.....

回复



10. ###说道:

2017年01月20日 14:58

我的chrome（mac版），window.speechSynthesis.getVoices()返回82个语言包。。。

回复



11. 二少说道:

2017年01月16日 10:38

我打印出 new window.SpeechSynthesisUtterance() 对象后发现没有 voiceURI 属性，不过有 voice 属性，是用来改变声音的吗，假如我想用星爷的声音替换掉原来那个女声应该怎么做呢，还有语音识别这个测试起来好坑，周围的人看我都像智障

回复



12. 盖大楼说道:

2017年01月16日 09:48

```
newRecognition.onresult = function(event) {  
  console.log(event);  
}
```



```
}
```

onresult是只有 webkitSpeechRecognition()的实例才能用吗? 别的对象有onresult这个方法吗? 如果有是什么意思?

[回复](#)

13. xixi说道:

2017年01月15日 16:52

张大神, 其实关注你博客很久了, 我想说的是为嘛不弄个分类, 比如css单独分出来, 不然40多页的博客真的很难找啊。。。



[回复](#)

张 鑫旭说道:

2017年01月16日 21:27

右上方有, 自从上次有人提过这个问题我特意背景色加强了下, 难道还是不行吗~~



[回复](#)

阿牛说道:

2017年01月18日 17:00

确实不明显, 建议主导航上区分好些

[回复](#)



14. cshenger说道:

2017年01月13日 23:26

这个试了一下还是蛮好玩的, 但是不知道是什么限制吧, 我这里只有女声, 不知道有没有男声可选, 哈哈哈

[回复](#)



yue说道:

2017年07月13日 10:49

同问可不可以男女声切换, 谷歌的语音包里的en-us里都是女声, 如果我想切换男声该怎么操作呢?

[回复](#)



15. 在一起的青春说道:

2017年01月13日 15:22

去年11月份用了, 火狐低版本不支持~

[回复](#)



16. exoticknight说道:

2017年01月13日 12:53

要跟上国际水平, aria 方面不容忽视

[回复](#)



17. 渣布说道:

2017年01月13日 11:03

语音合成, chrome并没有读出来, 但是360极速模式读出来了。明明是同一种内核。

[回复](#)



最新文章

- » 常见的CSS图形绘制合集
- » 粉丝群第1期CSS小测点评与答疑
- » 分享三个纯CSS实现26个英文字母的案例
- » 小tips: 纯CSS实现打字动画效果
- » CSS/CSS3 box-decoration-break属性简介
- » CSS :placeholder-shown伪类实现Material Design占位符交互效果
- » 从天猫某活动视频不必要的3次请求说起
- » CSS vector-effect与SVG stroke描边缩放
- » CSS ::backdrop伪元素是干嘛用的?
- » 周知: CSS -webkit-伪元素选择器不再导致整行无效

今日热门

- » 常见的CSS图形绘制合集 ⁽¹⁹⁰⁾
- » 未来必热: SVG Sprite技术介绍 ⁽¹¹⁹⁾
- » 粉丝群第1期CSS小测点评与答疑 ⁽¹¹⁵⁾
- » HTML5终极备忘大全 (图片版+文字版) ⁽⁹³⁾
- » 让所有浏览器支持HTML5 video视频标签 ⁽⁸⁶⁾
- » Selectivizr-让IE6~8支持CSS3伪类和属性选择器 ⁽⁸²⁾
- » CSS3下的147个颜色名称及对应颜色值 ⁽⁷⁹⁾
- » 视区相关单位vw, vh..简介以及可实际应用场景 ⁽⁷⁶⁾
- » 写给自己看的display: flex布局教程 ⁽⁷⁶⁾
- » 小tips: 纯CSS实现打字动画效果 ⁽⁷⁶⁾

今年热议

- » 《CSS世界》女主角诚寻靠谱一起奋斗之人 ⁽⁷⁶⁾
- » 不借助Echarts等图形框架原生JS快速实现折线图效果 ⁽⁶⁴⁾
- » 看, for..in和for..of在那里吵架! ⁽⁶⁰⁾
- » 是时候好好安利下LuLu UI框架了! ⁽⁴⁷⁾
- » 原来浏览器原生支持JS Base64编码解码 ⁽³⁵⁾
- » 妙法攻略: 渐变虚框及边框滚动动画的纯CSS实现 ⁽³³⁾
- » 炫酷H5中序列图片视频化播放的高性能实现 ⁽³¹⁾
- » CSS scroll-behavior和JS scrollIntoView让页面滚动平滑 ⁽³⁰⁾
- » windows系统下批量删除OS X系统.DS_Store文件 ⁽²⁶⁾
- » 写给自己看的display: flex布局教程 ⁽²⁶⁾

猜你喜欢

- 渐进使用HTML5语言识别, so easy!
- Chrome下语音识别实现页面快速引导以及重要交互
- WAI-ARIA无障碍网页应用属性完全展示
- HTML可访问性fieldset, legend元素及CSS布局应用
- 基于VoiceOver的移动web站无障碍访问实战
- 翻译-盲人如何使用互联网的8个误区
- span与a元素的键盘聚焦性以及键盘点击性研究
- 找到适合自己的前端发展方向
- HTML <area><map>标签及在实际开发中的应用
- HTML tabindex属性与web网页键盘无障碍访问
- HTML5 number类型文本框step属性的验证机制

