

Microsoft Cognitive Services

George Spyrou , Software Engineer

Agenda

- What are Cognitive Services APIs
- Types of Microsoft Cognitive Services
- Development setup
- Demo for Speech API

Cognitive Services APIs

Microsoft Cognitive Services (formerly Project Oxford) are a set of APIs, SDKs and services available to developers to make their applications more intelligent, engaging and discoverable.

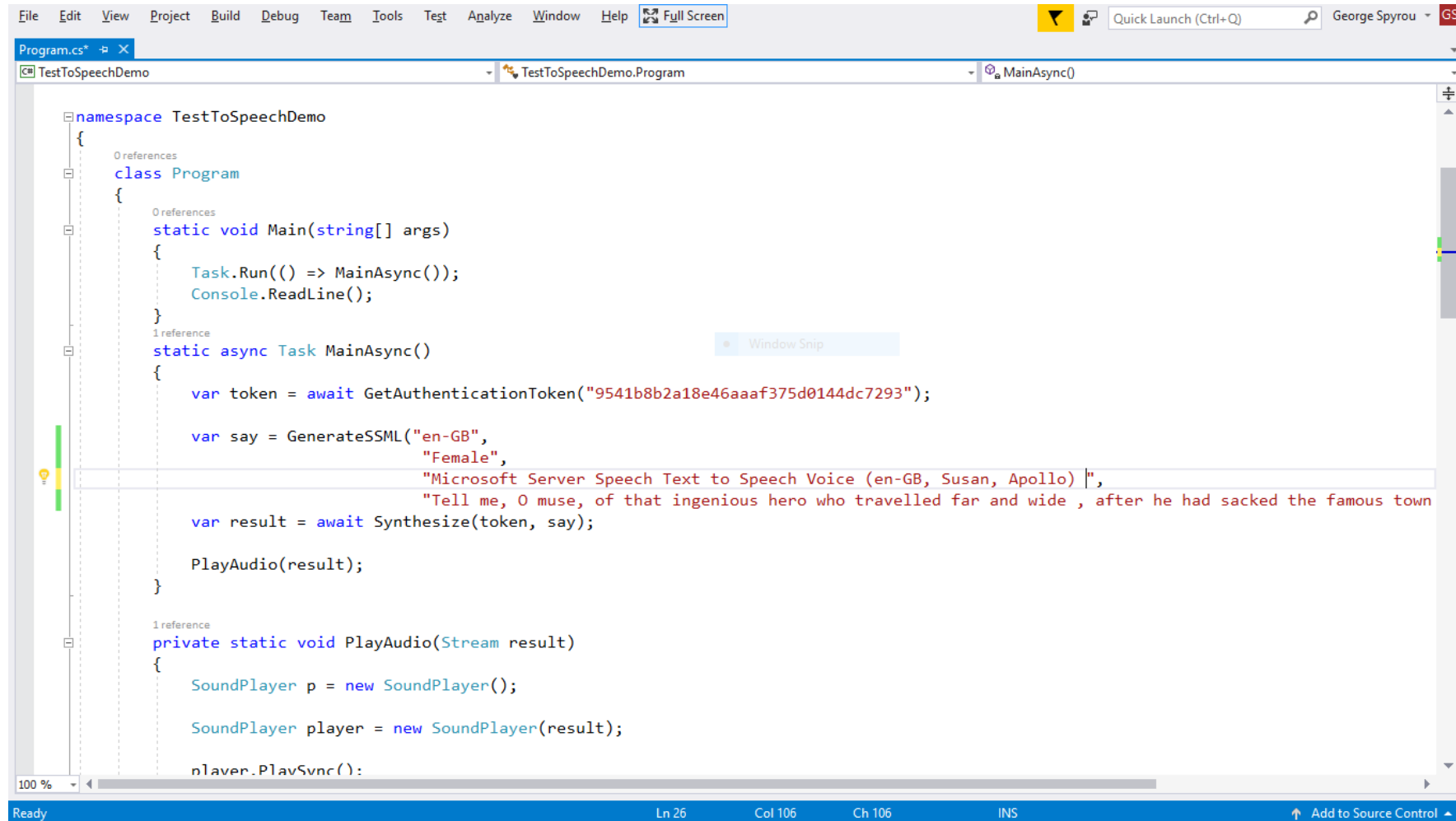
Types of Cognitive Service APIs

Vision	Speech	Language	Knowledge	Search
Computer Vision	Bing Speech	Bing Spell Check	Academic	Bing Autosuggest
Content Moderator	Custom Speech Service	Language Understanding	Entity Linking	Bing Image Search
Emotion	Speaker Recognition	Linguistic Analysis	Knowledge Exploration	Bing News Search
Face		Text Analytics	QnA Maker	Bing Video Search
Video		Translator	Recommendations	Bing Web Search
		WebLM		

Development setup

- Microsoft account
- Github
- LinkedIn

Demo for Speech API



The screenshot shows the Visual Studio IDE with a C# project named 'TestToSpeechDemo'. The code is written in Program.cs and includes a namespace 'TestToSpeechDemo' with a 'Program' class. The 'Main' method calls 'MainAsync', which uses the Speech API to generate and play audio. A 'Window Snip' tooltip is visible over the 'MainAsync' method. The status bar at the bottom indicates 'Ready' and provides line/col/char/line info (Ln 26, Col 106, Ch 106, INS).

```
File Edit View Project Build Debug Team Tools Test Analyze Window Help Full Screen
TestToSpeechDemo TestToSpeechDemo.Program MainAsync()

namespace TestToSpeechDemo
{
    0 references
    class Program
    {
        0 references
        static void Main(string[] args)
        {
            Task.Run(() => MainAsync());
            Console.ReadLine();
        }
        1 reference
        static async Task MainAsync()
        {
            var token = await GetAuthenticationToken("9541b8b2a18e46aaaf375d0144dc7293");

            var say = GenerateSSML("en-GB",
                                   "Female",
                                   "Microsoft Server Speech Text to Speech Voice (en-GB, Susan, Apollo) |",
                                   "Tell me, O muse, of that ingenious hero who travelled far and wide , after he had sacked the famous town
            var result = await Synthesize(token, say);

            PlayAudio(result);
        }
        1 reference
        private static void PlayAudio(Stream result)
        {
            SoundPlayer p = new SoundPlayer();

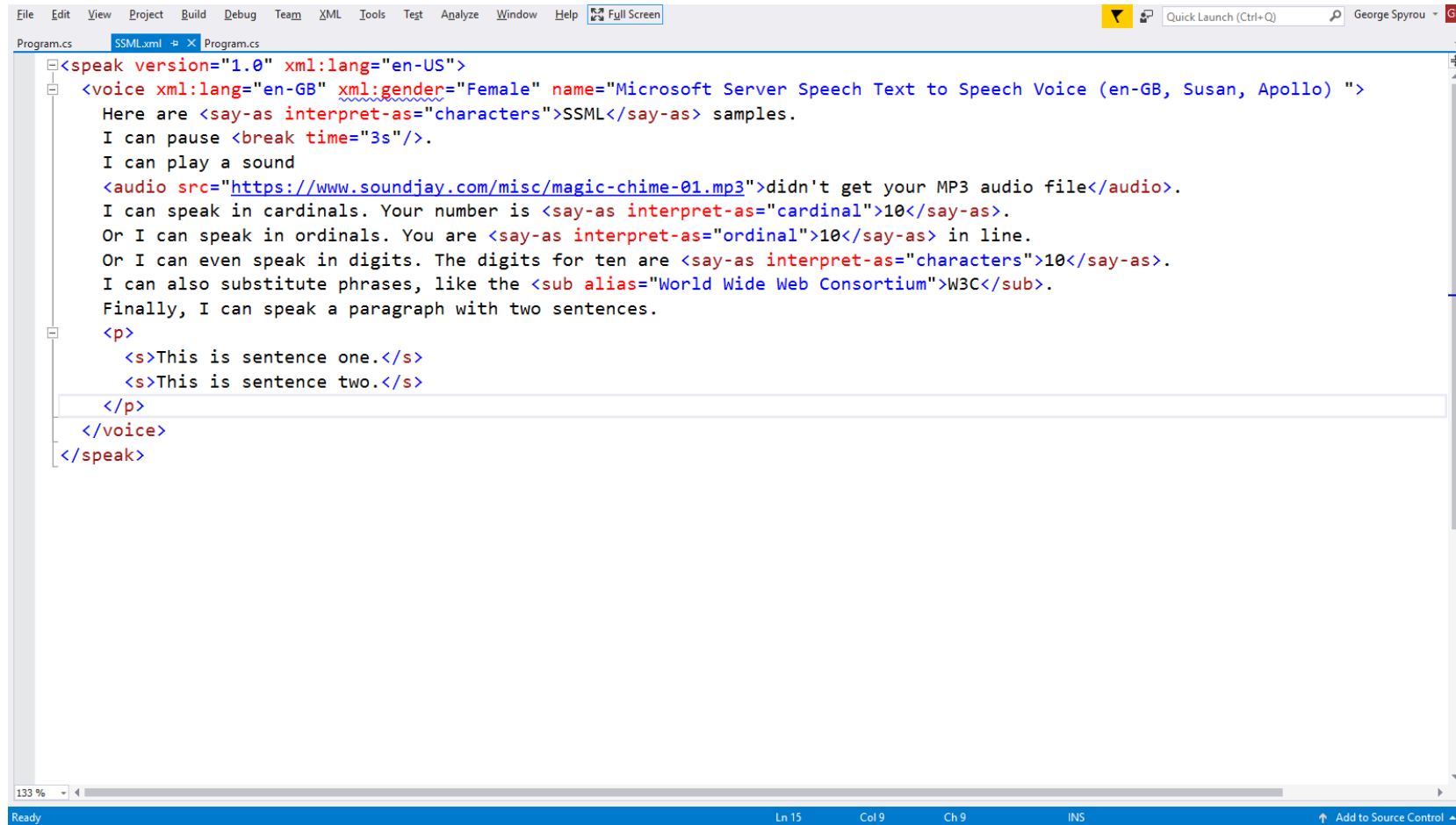
            SoundPlayer player = new SoundPlayer(result);

            player.PlaySync();
        }
    }
}
```

100 %

Ready Ln 26 Col 106 Ch 106 INS Add to Source Control

Speech Synthesis Markup Language



The screenshot shows a code editor window with a menu bar (File, Edit, View, Project, Build, Debug, Team, XML, Tools, Test, Analyze, Window, Help) and a toolbar with icons for Full Screen, Quick Launch (Ctrl+Q), and a user profile (George Spyrou). The editor has two tabs: 'SSML.xml' and 'Program.cs'. The 'SSML.xml' tab is active, displaying the following XML code:

```
<?xml version="1.0" encoding="UTF-8" ?>
<speak version="1.0" xml:lang="en-US">
  <voice xml:lang="en-GB" xml:gender="Female" name="Microsoft Server Speech Text to Speech Voice (en-GB, Susan, Apollo) ">
    Here are <say-as interpret-as="characters">SSML</say-as> samples.
    I can pause <break time="3s"/>.
    I can play a sound
    <audio src="https://www.soundjay.com/misc/magic-chime-01.mp3">didn't get your MP3 audio file</audio>.
    I can speak in cardinals. Your number is <say-as interpret-as="cardinal">10</say-as>.
    Or I can speak in ordinals. You are <say-as interpret-as="ordinal">10</say-as> in line.
    Or I can even speak in digits. The digits for ten are <say-as interpret-as="characters">10</say-as>.
    I can also substitute phrases, like the <sub alias="World Wide Web Consortium">W3C</sub>.
    Finally, I can speak a paragraph with two sentences.
    <p>
      <s>This is sentence one.</s>
      <s>This is sentence two.</s>
    </p>
  </voice>
</speak>
```

The status bar at the bottom shows 'Ready', 'Ln 15', 'Col 9', 'Ch 9', 'INS', and 'Add to Source Control'.

Resources

Cognitive Services APIs

<https://www.microsoft.com/cognitive-services/>

Speech Synthesis Markup Language (SSML)

<https://www.w3.org/TR/speech-synthesis/>

Microsoft Bing Speech API: Text-to-Speech Samples

<https://github.com/Azure-Samples/Cognitive-Speech-TTS>

Thank you :)

George Spyrou

george@plusapps.eu

@gspyrou

<http://www.linkedin.com/in/gspyrou>