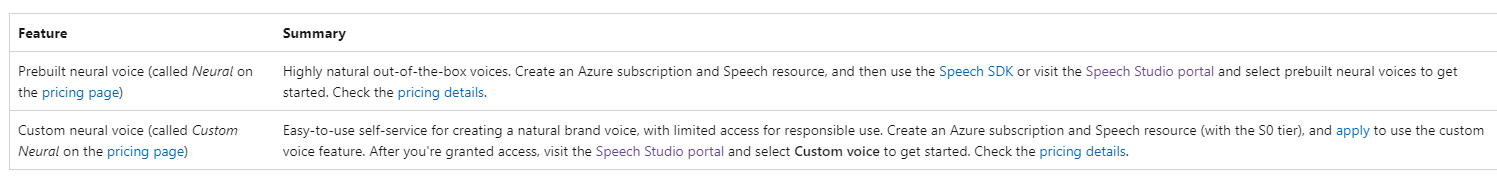
**Implement text-to-speech**

Text to speech enables your applications, tools, or devices to convert text into human like synthesized speech. The text to speech capability is also known as speech synthesis. Use human like prebuilt neural voices out of the box or create a custom neural voice that's unique to your product or brand.

**Core features**

Text to speech includes the following features:



**Billable characters**

[Speech service pricing](https://azure.microsoft.com/pricing/details/cognitive-services/speech-services/).

When you use the text to speech feature, you're billed for each character that's converted to speech, including punctuation. Although the SSML document itself isn't billable, optional elements that are used to adjust how the text is converted to speech, like phonemes and pitch, are counted as billable characters. Here's a list of what's billable:

* Text passed to the text to speech feature in the SSML body of the request
* All markup within the text field of the request body in the SSML format, except for <speak> and <voice> tags
* Letters, punctuation, spaces, tabs, markup, and all white-space characters
* Every code point defined in Unicode

**Important**

* Each Chinese character is counted as two characters for billing, including kanji used in Japanese, hanja used in Korean, or hanzi used in other languages.

**Model training and hosting time for custom neural voice**

Custom neural voice training and hosting are both calculated by hour and billed per second. For the billing unit price.

Custom neural voice (CNV) training time is measured by ‘compute hour’ (a unit to measure machine running time). Typically, when training a voice model, two computing tasks are running in parallel. So, the calculated compute hours are longer than the actual training time. On average, it takes less than one compute hour to train a CNV Lite voice; while for CNV Pro, it usually takes 20 to 40 compute hours to train a single-style voice, and around 90 compute hours to train a multi-style voice. The CNV training time is billed with a cap of 96 compute hours. So in the case that a voice model is trained in 98 compute hours, you'll only be charged with 96 compute hours.

Custom neural voice (CNV) endpoint hosting is measured by the actual time (hour). The hosting time (hours) for each endpoint is calculated at 00:00 UTC every day for the previous 24 hours. For example, if the endpoint has been active for 24 hours on day one, it's billed for 24 hours at 00:00 UTC the second day. If the endpoint is newly created or suspended during the day, it's billed for its accumulated running time until 00:00 UTC the second day. If the endpoint isn't currently hosted, it isn't billed. In addition to the daily calculation at 00:00 UTC each day, the billing is also triggered immediately when an endpoint is deleted or suspended. For example, for an endpoint created at 08:00 UTC on December 1, the hosting hour will be calculated to 16 hours at 00:00 UTC on December 2 and 24 hours at 00:00 UTC on December 3. If the user suspends hosting the endpoint at 16:30 UTC on December 3, the duration (16.5 hours) from 00:00 to 16:30 UTC on December 3 will be calculated for billing.

**Personal voice**

When you use the personal voice feature, you're billed for both profile storage and synthesis.

* **Profile storage**: After a personal voice profile is created, it will be billed until it is removed from the system. The billing unit is per voice per day. If voice storage lasts for a period of less than 24 hours, it will be billed as one full day.
* **Synthesis**: Billed per character. For details on billable characters, see the above [billable characters](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech#billable-characters).

**Text to speech avatar**

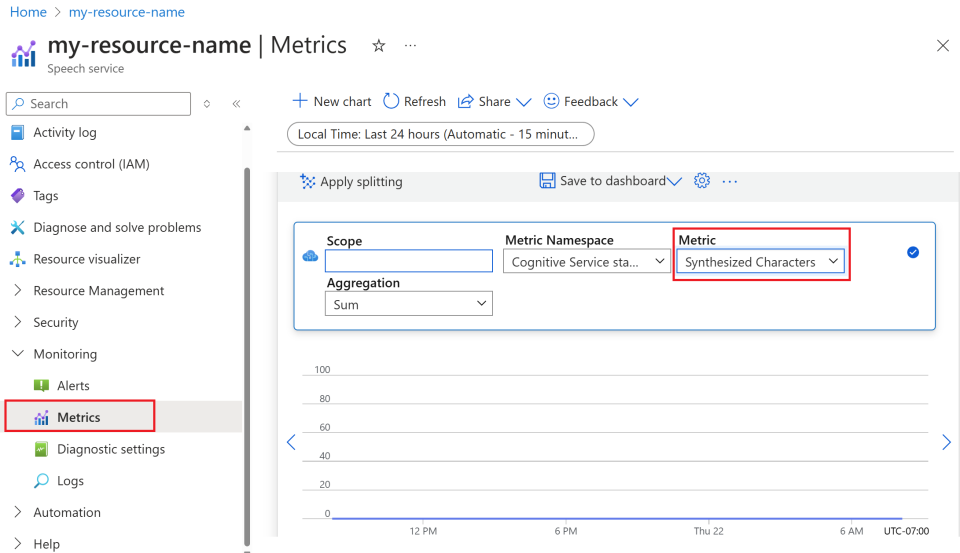
When using the text-to-speech avatar feature, charges will be incurred based on the length of video output and will be billed per second. However, for the real-time avatar, charges are based on the time when the avatar is active, regardless of whether it is speaking or remaining silent, and will also be billed per second. To optimize costs for real-time avatar usage, refer to the tips provided in the [sample code](https://github.com/Azure-Samples/cognitive-services-speech-sdk/tree/master/samples/js/browser/avatar#chat-sample) (search "Use Local Video for Idle"). Avatar hosting is billed per second per endpoint. You can suspend your endpoint to save costs. If you want to suspend your endpoint, you can delete it directly. To use it again, simply redeploy the endpoint.

**Monitor Azure text to speech metrics**

**How to find usage information in the Azure portal**

To effectively manage your Azure resources, it's essential to access and review usage information regularly. Here's how to find the usage information:

1. Go to the [Azure portal](https://ms.portal.azure.com/) and sign in with your Azure account.
2. Navigate to **Resources** and select your resource you wish to monitor.
3. Select **Metrics** under **Monitoring** from the left-hand menu.



1. Customize metric views.

You can filter data by resource type, metric type, time range, and other parameters to create custom views that align with your monitoring needs. Additionally, you can save the metric view to dashboards by selecting **Save to dashboard** for easy access to frequently used metrics.

1. Set up alerts.

To manage usage more effectively, set up alerts by navigating to the **Alerts** tab under **Monitoring** from the left-hand menu. Alerts can notify you when your usage reaches specific thresholds, helping to prevent unexpected costs.

**Definition of metrics**

A screenshot of a computer

Description automatically generated

**using the client library and REST API**

**Create an Azure resource**

To use the code sample below, you'll need to deploy an Azure resource. This resource will contain a key and endpoint you'll use to authenticate the API calls you send to the Speech service.

1. Use the following link to [Create a Speech resource](https://portal.azure.com/#create/Microsoft.CognitiveServicesSpeechServices) using the Azure portal. You will need to sign in using your Azure subscription.
2. In the **Create Speech Services** screen, provide the following information:

A computer screen with a blue and white screen

Description automatically generated

A close up of a white card

Description automatically generated

1. Select **Review + Create** at the bottom of the page.
2. In the screen that appears, make sure the validation has passed, and that you entered your information correctly. Then select **Create**.

**Get your key and endpoint**

* Get the Speech resource key and region. After your Speech resource is deployed, select **Go to resource** to view and manage keys.

**Create a new .NET Core application**

Using the Visual Studio IDE, create a new .NET Core console app.

Install the client library by right-clicking on the solution in the Solution Explorer and selecting Manage NuGet Packages. In the package manager that opens select Browse and search for   
Microsoft.CognitiveServices.Speech.

**Code example**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer code

Description automatically generated**

To change the speech synthesis language, replace en-US-AvaMultilingualNeural with another [supported voice](https://learn.microsoft.com/en-us/azure/ai-services/speech-service/language-support#prebuilt-neural-voices).

All neural voices are multilingual and fluent in their own language and English. For example, if the input text in English is *I'm excited to try text to speech* and you set es-ES-ElviraNeural as the language, the text is spoken in English with a Spanish accent. If the voice doesn't speak the language of the input text, the Speech service doesn't output synthesized audio

**Output**

Enter some text that you want to speak. For example, type *I'm excited to try text to speech*. Select the Enter key to hear the synthesized speech.

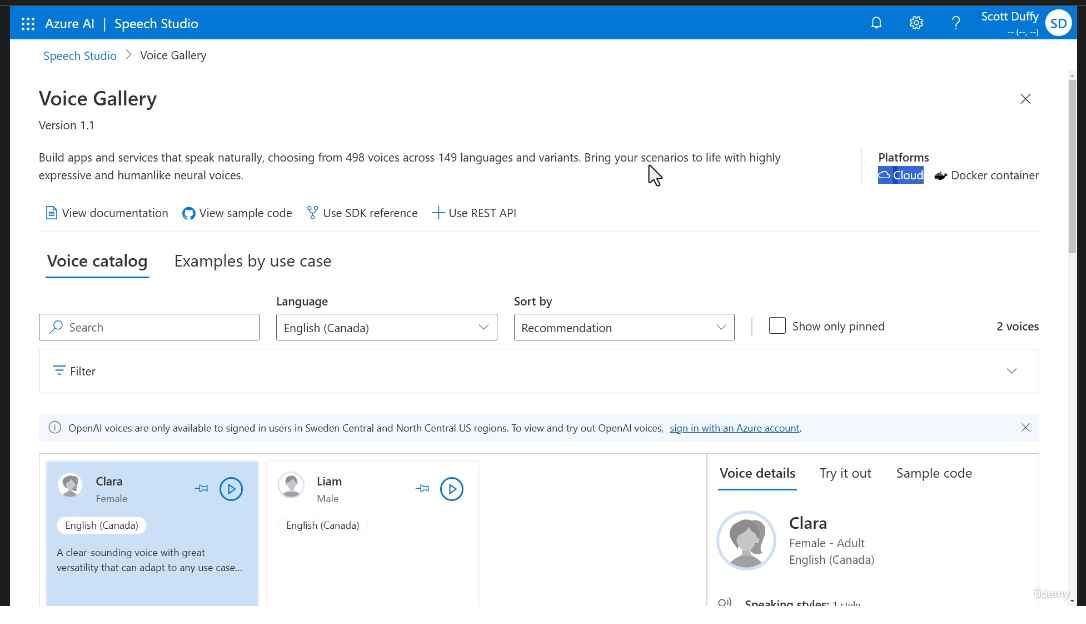
A white rectangular object with a black border

Description automatically generated

**Use Speech Studio with your own text**

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated