Text-to-Speech allows you to convert words and sentences into base64 encoded audio data of natural human speech. You can then convert the audio data into a playable audio file like an MP3 by decoding the base64 data. The Cloud Text-to-Speech API accepts input as raw text or Speech Synthesis Markup Language (SSML) (/text-to-speech/docs/ssml).

This document describes how to create an audio file from either text or SSML input using Text-to-Speech. You can also review the Text-to-Speech basics (/text-to-speech/docs/basics) article if you are unfamiliar with concepts like speech synthesis or SSML.

These samples require that you have set up `gcloud` and have created and activated a service account. For information about setting up `gcloud`, and also creating and activating a service account, see Quickstart:Text-to-Speech (/text-to-speech/docs/quickstart-protocol).

The following code samples demonstrate how to convert a string into audio data.

You can configure the output of speech synthesis in a variety of ways, including selecting a unique voice (/text-to-speech/docs/voices) or modulating the output in pitch, volume, speaking rate, and sample rate (/text-to-speech/docs/basics#audio-config).

Using SSML in your audio synthesis request can produce audio that is more similar to natural human speech. Specifically, SSML gives you finer-grain control over how the audio output represents pauses in the speech or how the audio pronounces dates, times, acronyms, and abbreviations.

For more details on the SSML elements supported by Cloud Text-to-Speech API, see the SSML reference (/text-to-speech/docs/ssml).