# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>What Is Amazon Polly?</td>
<td>1</td>
</tr>
<tr>
<td>Are You a First-time User of Amazon Polly?</td>
<td>2</td>
</tr>
<tr>
<td>How It Works</td>
<td>3</td>
</tr>
<tr>
<td>What's Next?</td>
<td>3</td>
</tr>
<tr>
<td>Getting Started</td>
<td>4</td>
</tr>
<tr>
<td>Step 1: Set Up an Account &amp; User</td>
<td>4</td>
</tr>
<tr>
<td>Step 1.1: Sign up for AWS</td>
<td>4</td>
</tr>
<tr>
<td>Step 1.2: Create an IAM User</td>
<td>5</td>
</tr>
<tr>
<td>Next Step</td>
<td>5</td>
</tr>
<tr>
<td>Step 2: Getting Started (Console)</td>
<td>5</td>
</tr>
<tr>
<td>Exercise 1: Synthesizing Speech Quick Start (Console)</td>
<td>6</td>
</tr>
<tr>
<td>Exercise 2: Synthesizing Speech with Plain Text Input (Console)</td>
<td>6</td>
</tr>
<tr>
<td>Next Step</td>
<td>7</td>
</tr>
<tr>
<td>Step 3: Getting Started (AWS CLI)</td>
<td>7</td>
</tr>
<tr>
<td>Step 3.1: Set Up the AWS CLI</td>
<td>7</td>
</tr>
<tr>
<td>Step 3.2: Getting Started Exercise</td>
<td>8</td>
</tr>
<tr>
<td>Python Examples</td>
<td>10</td>
</tr>
<tr>
<td>Set Up Python and Test an Example (SDK)</td>
<td>10</td>
</tr>
<tr>
<td>Voices in Amazon Polly</td>
<td>12</td>
</tr>
<tr>
<td>Available Voices</td>
<td>12</td>
</tr>
<tr>
<td>Bilingual Voices</td>
<td>15</td>
</tr>
<tr>
<td>Accented bilingual voices</td>
<td>15</td>
</tr>
<tr>
<td>Fully bilingual voices</td>
<td>15</td>
</tr>
<tr>
<td>Listening to Voices</td>
<td>16</td>
</tr>
<tr>
<td>Voice Speed</td>
<td>16</td>
</tr>
<tr>
<td>Changing Your Voice Speed</td>
<td>17</td>
</tr>
<tr>
<td>Languages Supported by Amazon Polly</td>
<td>18</td>
</tr>
<tr>
<td>Phoneme and Viseme Tables for Supported Languages</td>
<td>19</td>
</tr>
<tr>
<td>Neural TTS</td>
<td>93</td>
</tr>
<tr>
<td>Feature and Region Compatibility</td>
<td>93</td>
</tr>
<tr>
<td>The Voice Engine</td>
<td>94</td>
</tr>
<tr>
<td>Choosing the Voice Engine (Console)</td>
<td>94</td>
</tr>
<tr>
<td>Choosing the Voice Engine (CLI)</td>
<td>95</td>
</tr>
<tr>
<td>Neural Voices</td>
<td>96</td>
</tr>
<tr>
<td>NTTTS Newscaster Speaking Style</td>
<td>97</td>
</tr>
<tr>
<td>Speech Marks</td>
<td>99</td>
</tr>
<tr>
<td>Speech Mark Types</td>
<td>99</td>
</tr>
<tr>
<td>Visemes and Amazon Polly</td>
<td>99</td>
</tr>
<tr>
<td>Using Speech Marks</td>
<td>100</td>
</tr>
<tr>
<td>Requesting Speech Marks</td>
<td>100</td>
</tr>
<tr>
<td>Speech Mark Output</td>
<td>101</td>
</tr>
<tr>
<td>Speech Mark Examples</td>
<td>102</td>
</tr>
<tr>
<td>Requesting Speech Marks (Console)</td>
<td>103</td>
</tr>
<tr>
<td>Using SSML</td>
<td>105</td>
</tr>
<tr>
<td>Reserved Characters</td>
<td>105</td>
</tr>
<tr>
<td>Using SSML in the Console</td>
<td>107</td>
</tr>
<tr>
<td>Using SSML in the AWS CLI</td>
<td>108</td>
</tr>
<tr>
<td>Using SSML With the Synthesize-Speech Command</td>
<td>108</td>
</tr>
<tr>
<td>Synthesizing an SSML-enhanced Document</td>
<td>109</td>
</tr>
<tr>
<td>Using SSML for Common Amazon Polly Tasks</td>
<td>110</td>
</tr>
<tr>
<td>Supported SSML Tags</td>
<td>112</td>
</tr>
<tr>
<td>Identifying SSML-Enhanced Text</td>
<td>113</td>
</tr>
<tr>
<td>Adding a Pause</td>
<td>113</td>
</tr>
<tr>
<td>Emphasizing Words</td>
<td>114</td>
</tr>
</tbody>
</table>
What Is Amazon Polly?

Amazon Polly is a cloud service that converts text into lifelike speech. You can use Amazon Polly to develop applications that increase engagement and accessibility. Amazon Polly supports multiple languages and includes a variety of lifelike voices, so you can build speech-enabled applications that work in multiple locations and use the ideal voice for your customers. With Amazon Polly, you only pay for the text you synthesize. You can also cache and replay Amazon Polly's generated speech at no additional cost.

Additionally, Amazon Polly includes a number of Neural Text-to-Speech (NTTS) voices, delivering ground-breaking improvements in speech quality through a new machine learning approach, thereby offering to customers the most natural and human-like text-to-speech voices possible. Neural TTS technology also supports a Newscaster speaking style that is tailored to news narration use cases.

Common use cases for Amazon Polly include, but are not limited to, mobile applications such as newsreaders, games, eLearning platforms, accessibility applications for visually impaired people, and the rapidly growing segment of Internet of Things (IoT).

Amazon Polly is certified for use with regulated workloads for HIPAA (the Health Insurance Portability and Accountability Act of 1996), and Payment Card Industry Data Security Standard (PCI DSS).

Some of the benefits of using Amazon Polly include:

- **High quality** – Amazon Polly offers both new neural TTS and best-in-class standard TTS technology to synthesize the superior natural speech with high pronunciation accuracy (including abbreviations, acronym expansions, date/time interpretations, and homograph disambiguation).

- **Low latency** – Amazon Polly ensures fast responses, which make it a viable option for low-latency use cases such as dialog systems.

- **Support for a large portfolio of languages and voices** – Amazon Polly supports dozens of voices languages, offering male and female voice options for most languages. Neural TTS currently supports three British English voices and eight US English voices. This number will continue to increase as we bring more neural voices online. US English voices Matthew and Joanna can also use the Neural Newscaster speaking style, similar to what you might hear from a professional news anchor.

- **Cost-effective** – Amazon Polly's pay-per-use model means there are no setup costs. You can start small and scale up as your application grows.

- **Cloud-based solution** – On-device TTS solutions require significant computing resources, notably CPU power, RAM, and disk space. These can result in higher development costs and higher power consumption on devices such as tablets, smart phones, and so on. In contrast, TTS conversion done in the AWS Cloud dramatically reduces local resource requirements. This enables support of all the available languages and voices at the best possible quality. Moreover, speech improvements are instantly available to all end-users and do not require additional updates for devices.
Are You a First-time User of Amazon Polly?

If you are a first-time user of Amazon Polly, we recommend that you read the following sections in the listed order:

1. How Amazon Polly Works (p. 3) – This section introduces various Amazon Polly inputs and options that you can work with in order to create an end-to-end experience.
2. Getting Started with Amazon Polly (p. 4) – In this section, you set up your account and test Amazon Polly speech synthesis.
3. Example Applications (p. 194) – This section provides additional examples that you can use to explore Amazon Polly.
How Amazon Polly Works

Amazon Polly converts input text into life-like speech. You call one of the speech synthesis methods, provide the text that you want to synthesize, choose one of the Neural Text-to-Speech (NTTS) or Standard Text-to-Speech (TTS) voices, and specify an audio output format. Amazon Polly then synthesizes the provided text into a high-quality speech audio stream.

- **Input text** – Provide the text that you want to synthesize, and Amazon Polly returns an audio stream. You can provide the input as plain text or in Speech Synthesis Markup Language (SSML) format. With SSML you can control various aspects of speech, such as pronunciation, volume, pitch, and speech rate. For more information, see Generating Speech from SSML Documents (p. 105).

- **Available voices** – Amazon Polly provides a portfolio of languages and a variety of voices, including a bilingual voice (for both English and Hindi). For most languages you can choose from several voices, both male and female. When launching a speech synthesis task, you specify the voice ID, and then Amazon Polly uses this voice to convert the text to speech. Amazon Polly is not a translation service—the synthesized speech is in the same language as the text. However, if the text is in a different language than designated for the voice, numbers represented as digits (for example, 53, not fifty-three) are synthesized in the language of the voice and not the text. For more information, see Voices in Amazon Polly.

- **Output format** – Amazon Polly can deliver the synthesized speech in multiple formats. You can select the audio format that suits your needs. For example, you might request the speech in the MP3 or Ogg Vorbis format for consumption by web and mobile applications. Or, you might request the PCM output format for consumption by AWS IoT devices and telephony solutions.

What's Next?

If you are new to Amazon Polly, we recommend that you to read the following topics in order:

- Getting Started with Amazon Polly (p. 4)
- Example Applications (p. 194)
- Quotas in Amazon Polly (p. 225)
Getting Started with Amazon Polly

Amazon Polly provides simple API operations that you can easily integrate with your existing applications. For a list of supported operations, see Actions (p. 248). You can use either of the following options:

- **AWS SDKs** – When using the SDKs, your requests to Amazon Polly are automatically signed and authenticated using the credentials you provide. This is the recommended choice for building your applications.
- **AWS CLI** – You can use the AWS CLI to access any of Amazon Polly functionality without having to write any code.

The following sections describe how to get set up and provide an introductory exercise.

**Topics**
- Step 1: Set Up an AWS Account and Create a User (p. 4)
- Step 2: Getting Started (Console) (p. 5)
- Step 3: Getting Started (AWS CLI) (p. 7)
- Python Examples (p. 10)

**Step 1: Set Up an AWS Account and Create a User**

Before you use Amazon Polly for the first time, complete the following tasks:

1. Step 1.1: Sign up for AWS (p. 4)
2. Step 1.2: Create an IAM User (p. 5)

**Step 1.1: Sign up for AWS**

When you sign up for Amazon Web Services (AWS), your AWS account is automatically signed up for all services in AWS, including Amazon Polly. You are charged only for the services that you use.

With Amazon Polly, you pay only for the resources you use. If you are a new AWS customer, you can get started with Amazon Polly for free. For more information, see AWS Free Usage Tier.

If you already have an AWS account, skip to the next step. If you don't have an AWS account, perform the steps in the following procedure to create one.

**To create an AWS account**

2. Follow the online instructions.

   Part of the sign-up procedure involves receiving a phone call and entering a verification code on the phone keypad.

Note your AWS account ID because you'll need it for the next step.
Step 1.2: Create an IAM User

Services in AWS, such as Amazon Polly, require that you provide credentials when you access them so that the service can determine whether you have permissions to access the resources owned by that service. The console requires your password. You can create access keys for your AWS account to access the AWS CLI or API. However, we don't recommend that you access AWS using the credentials for your AWS account. Instead, we recommend that you use AWS Identity and Access Management (IAM). Create an IAM user, add the user to an IAM group with administrative permissions, and then grant administrative permissions to the IAM user that you created. You can then access AWS using a special URL and that IAM user's credentials.

If you signed up for AWS, but you haven't created an IAM user for yourself, you can create one using the IAM console.

The exercises in this guide assume that you have a user (adminuser) with administrator privileges. Follow the procedure to create adminuser in your account.

To create an administrator user and sign in to the console

1. Create an administrator user called adminuser in your AWS account. For instructions, see Creating Your First IAM User and Administrators Group in the IAM User Guide.
2. A user can sign in to the AWS Management Console using a special URL. For more information, How Users Sign In to Your Account in the IAM User Guide.

Important
The Getting Started exercises use the adminuser credentials. For added security, when building and testing production application we recommend you create a service-specific administrator user who has permissions for only the Amazon Polly actions. For an example policy that grants Amazon Polly specific permissions, see Example 1: Allow All Amazon Polly Actions (p. 236).

For more information about IAM, see the following:

- AWS Identity and Access Management (IAM)
- Getting started
- IAM User Guide

Next Step

Step 2: Getting Started (Console) (p. 5)
Exercise 1: Synthesizing Speech Quick Start (Console)

The Quick Start walks you through the fastest way to test the Amazon Polly speech synthesis for speech quality. When you select the Text-to-Speech tab, the text field for entering your text is pre-loaded with example text so you can quickly try out Amazon Polly.

**To quickly test Amazon Polly (Console)**

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Text-to-Speech tab.
3. Turn off SSML.
4. Under Engine, choose Standard or Neural.
5. Choose a language and AWS Region, then choose a voice. If you choose Neural for Engine, only the languages and voices that support NTTS are available. All Standard voices are disabled.
6. Choose Listen.

For more in-depth testing, see the following topics:

- Exercise 2: Synthesizing Speech with Plain Text Input (Console) (p. 6)
- Using SSML (Console) (p. 107)
- Applying Lexicons Using the Console (Synthesize Speech) (p. 135)

Exercise 2: Synthesizing Speech with Plain Text Input (Console)

The following procedure synthesizes speech using plain text input. Note how "W3C" and the date "10/3" (October 3rd) are synthesized.

**To synthesize speech using plain text input (console)**

1. After logging on to the Amazon Polly console, choose Try Amazon Polly, and then choose the Text-to-Speech tab.
2. Turn off SSML.
3. Type or paste this text into the input box.

   ```plaintext
   He was caught up in the game.
   In the middle of the 10/3/2014 W3C meeting he shouted, "Score!" quite loudly.
   ```

4. For Engine, choose Standard or Neural.
5. Choose a language and AWS Region, then choose a voice. If you choose Neural for Engine, only the languages and voices that support NTTS are available. All Standard voices are disabled.
6. To listen to the speech immediately, choose Listen.
7. To save the speech to a file, do one of the following:
   a. Choose Download.
   b. To change to a different file format, expand Additional settings, turn on Speech file format settings, choose the file format that you want, and then choose Download.

For more in-depth examples, see the following topics:
• Applying Lexicons Using the Console (Synthesize Speech) (p. 135)
• Using SSML (Console) (p. 107)

Next Step

Step 3: Getting Started (AWS CLI) (p. 7)

Step 3: Getting Started (AWS CLI)

You can perform almost all of the Amazon Polly operations that you can perform using the Amazon Polly console using the AWS Command Line Interface (AWS CLI). You can’t listen to synthesized speech using the AWS CLI. Instead, you must save it to a file and then open the file in an application that can play it.

Topics
• Step 3.1: Set Up the AWS Command Line Interface (AWS CLI) (p. 7)
• Step 3.2: Getting Started Exercise Using the AWS CLI (p. 8)

Step 3.1: Set Up the AWS Command Line Interface (AWS CLI)

Follow the steps to download and configure the AWS CLI.

Important
You don’t need the AWS CLI to perform the steps in this exercise. However, some of the exercises in this guide use the AWS CLI. You can skip this step and go to Step 3.2: Getting Started Exercise Using the AWS CLI (p. 8), and then set up the AWS CLI later when you need it.

To set up the AWS CLI

1. Download and configure the AWS CLI. For instructions, see the following topics in the AWS Command Line Interface User Guide:
   • Getting Set Up with the AWS Command Line Interface
   • Configuring the AWS Command Line Interface
2. Add a named profile for the administrator user in the AWS CLI config file. You use this profile when running the AWS CLI commands. For more information about named profiles, see Named Profiles in the AWS Command Line Interface User Guide.

   [profile adminuser]
   aws_access_key_id = adminuser access key ID
   aws_secret_access_key = adminuser secret access key
   region = aws-region

For a list of available AWS Regions and those supported by Amazon Polly, see Regions and Endpoints in the Amazon Web Services General Reference.

Note
If you’re using the Region supported by Amazon Polly that you specified when you configured the AWS CLI, omit the following line from the AWS CLI code examples.
--region aws-region

3. Verify the setup by typing the following help command at the command prompt.

```
aws help
```

A list of valid AWS commands should appear in the AWS CLI window.

**To enable Amazon Polly in the AWS CLI (optional)**

If you have previously downloaded and configured the AWS CLI, Amazon Polly might not be available unless you reconfigure the AWS CLI. This procedure checks to see if this is necessary and provides instructions if Amazon Polly is not automatically available.

1. Verify the availability of Amazon Polly by typing the following help command at the AWS CLI command prompt.

```
aws polly help
```

If a description of Amazon Polly and a list of valid commands appears in the AWS CLI window, Amazon Polly is available in the AWS CLI and can be used immediately. In this case, you can skip the rest of this procedure. If this is not displayed, continue with Step 2.

2. Use one of the two following options to enable Amazon Polly:
   a. Uninstall and reinstall the AWS CLI.
      For instructions, see [Installing the AWS Command Line Interface](#) in the *AWS Command Line Interface User Guide*.
   
   or
   b. Download the file `service-2.json`.
      At the command prompt, run the following command.

```
aws configure add-model --service-model file://service-2.json --service-name polly
```

3. Reverify the availability of Amazon Polly.

```
aws polly help
```

The description of Amazon Polly should be visible.

**Next Step**

Step 3.2: Getting Started Exercise Using the AWS CLI (p. 8)

**Step 3.2: Getting Started Exercise Using the AWS CLI**

Now you can test the speech synthesis offered by Amazon Polly. In this exercise, you call the `SynthesizeSpeech` operation by passing in sample text. You can save the resulting audio as a file and verify its content.
1. Run the `synthesize-speech` AWS CLI command to synthesize sample text to an audio file (`hello.mp3`).

   The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (`\`) Unix continuation character at the end of each line with a caret (`^`) and use full quotation marks (`"`) around the input text with single quotes (`'`) for interior tags.

   ```
   aws polly synthesize-speech \
   --output-format mp3 \
   --voice-id Joanna \
   --text 'Hello, my name is Joanna. I learned about the W3C on 10/3 of last year.' \
   hello.mp3
   ```

   In the call to `synthesize-speech`, you provide sample text for the synthesis, the voice to use (by providing a voice ID, explained in the following step 3), and the output format. The command saves the resulting audio to the `hello.mp3` file.

   In addition to the MP3 file, the operation sends the following output to the console.

   ```
   {
     "ContentType": "audio/mpeg",
     "RequestCharacters": "71"
   }
   ```

2. Play the resulting `hello.mp3` file to verify the synthesized speech.

3. Get the list of available voices by using the `DescribeVoices` operation. Run the following `describe-voices` AWS CLI command.

   ```
   aws polly describe-voices
   ```

   In response, Amazon Polly returns the list of all available voices. For each voice, the response provides the following metadata: voice ID, language code, language name, and the gender of the voice. The following is a sample response.

   ```
   {
     "Voices": [
       {
         "Gender": "Female",
         "Name": "Salli",
         "LanguageName": "US English",
         "Id": "Salli",
         "LanguageCode": "en-US"
       },
       {
         "Gender": "Female",
         "Name": "Joanna",
         "LanguageName": "US English",
         "Id": "Joanna",
         "LanguageCode": "en-US"
       }
     ]
   }
   ```

   Optionally, you can specify the language code to find the available voices for a specific language. Amazon Polly supports dozens of voices. The following example lists all the voices for Brazilian Portuguese.

   ```
   aws polly describe-voices \ 
   ```
For a list of language codes, see Languages Supported by Amazon Polly (p. 18). These language codes are W3C language identification tags (ISO 639 code for the language name-ISO 3166 country code). For example, en-US (US English), en-GB (British English), and es-ES (Spanish), etc.

You can also use the help option in the AWS CLI to get the list of language codes:

```bash
aws polly describe-voices help
```

## Python Examples

This guide provides additional examples, some of which are Python code examples that use AWS SDK for Python (Boto) to make API calls to Amazon Polly. We recommend that you set up Python and test the example code provided in the following section. For additional examples, see Example Applications (p. 194).

### Set Up Python and Test an Example (SDK)

To test the Python example code, you need the AWS SDK for Python (Boto). For instruction, see AWS SDK for Python (Boto3).

**To test the example Python code**

The following Python code example performs the following actions:

- Uses the AWS SDK for Python (Boto) to send a `SynthesizeSpeech` request to Amazon Polly (by providing simple text as input).
- Accesses the resulting audio stream in the response and saves the audio to a file (`speech.mp3`) on your local disk.
- Plays the audio file with the default audio player for your local system.

Save the code to a file (example.py) and run it.

```python
# Getting Started Example for Python 2.7+/3.3+
from boto3 import Session
from botocore.exceptions import BotoCoreError, ClientError
from contextlib import closing
import os
import sys
import subprocess
from tempfile import gettempdir

# Create a client using the credentials and region defined in the [adminuser] section of the AWS credentials file (~/.aws/credentials).
session = Session(profile_name="adminuser")
polly = session.client("polly")

try:
    # Request speech synthesis
    response = polly.synthesize_speech(Text="Hello world!", OutputFormat="mp3", VoiceId="Joanna")
except (BotoCoreError, ClientError) as error:
    # The service returned an error, exit gracefully
```

### Other Python Examples

This section provides additional examples for using Amazon Polly with Python. Each example is designed to demonstrate different aspects of using the Polly API.

- **Get Text to Synthesize**
  - Use the `Text` parameter to specify the text to be synthesized.

- **Set Voice and Output Format**
  - Customize the voice and output format as needed.

- **Voice Variants**
  - Explore different voice variants for customization.

- **Stream Audio**
  - Stream audio directly to a file or to an S3 bucket.

- **Text to Speech Streams**
  - Use the `TextToSpeechStream` class for streaming text to speech operations.

- **Language Support**
  - Support for multiple languages, including regional variants.

By exploring these examples, developers can gain a deeper understanding of Amazon Polly's capabilities and how to integrate it into their applications.
print(error)
sys.exit(-1)

# Access the audio stream from the response
if "AudioStream" in response:
    # Note: Closing the stream is important because the service throttles on the
    # number of parallel connections. Here we are using contextlib.closing to
    # ensure the close method of the stream object will be called automatically
    # at the end of the with statement's scope.
    with closing(response["AudioStream"]) as stream:
        output = os.path.join(gettempdir(), "speech.mp3")

        try:
            # Open a file for writing the output as a binary stream
            with open(output, "wb") as file:
                file.write(stream.read())
        except IOError as error:
            # Could not write to file, exit gracefully
            print(error)
            sys.exit(-1)

else:
    # The response didn't contain audio data, exit gracefully
    print("Could not stream audio")
    sys.exit(-1)

# Play the audio using the platform's default player
if sys.platform == "win32":
    os.startfile(output)
else:
    # The following works on macOS and Linux. (Darwin = mac, xdg-open = linux).
    opener = "open" if sys.platform == "darwin" else "xdg-open"
    subprocess.call([opener, output])

For additional examples including an example application, see Example Applications (p. 194).
Voices in Amazon Polly

Amazon Polly provides a number of different voices for you to use. To hear example voices, see the Amazon Polly product overview. To hear a specific voice speak a sample that you provide, you can use the Amazon Polly console. For instructions, see Listening to the Voices (p. 16).

Available Voices

Amazon Polly provides a variety of different voices in multiple languages for synthesizing speech from text.

<table>
<thead>
<tr>
<th>Language</th>
<th>Name/ID</th>
<th>Gender</th>
<th>Neural Voice</th>
<th>Standard Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic (arb)</td>
<td>Zeina</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Chinese, Mandarin (cmn-CN)</td>
<td>Zhiyu</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Danish (da-DK)</td>
<td>Naja</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Mads</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Dutch (nl-NL)</td>
<td>Lotte</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Ruben</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>English (Australian) (en-AU)</td>
<td>Nicole</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Olivia</td>
<td>Female</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Russell</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>English (British) (en-GB)</td>
<td>Amy**</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Emma</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Brian</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>English (Indian) (en-IN)</td>
<td>Aditi*</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Raveena</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>English (New Zealand) (en-NZ)</td>
<td>Aria</td>
<td>Female</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>English (South African) (en-ZA)</td>
<td>Ayanda</td>
<td>Female</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>English (US) (en-US)</td>
<td>Ivy</td>
<td>Female (child)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Joanna**</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
## Available Voices

<table>
<thead>
<tr>
<th>Language</th>
<th>Name/ID</th>
<th>Gender</th>
<th>Neural Voice</th>
<th>Standard Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Kendra</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Kimberly</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Salli</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Joey</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Justin</td>
<td>Male (child)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Kevin</td>
<td>Male (child)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Matthew**</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(Welsh) (en-GB-WLS)</td>
<td>Geraint</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>French (fr-FR)</td>
<td>Céline/Celine</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Léa</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Mathieu</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>French (Canadian) (fr-CA)</td>
<td>Chantal</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Gabrielle</td>
<td>Female</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>German (de-DE)</td>
<td>Marlene</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Vicki</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Hans</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hindi (hi-IN)</td>
<td>Aditi*</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Icelandic (is-IS)</td>
<td>Dóra/Dora</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Karl</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Italian (it-IT)</td>
<td>Carla</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Bianca</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Giorgio</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Japanese (ja-JP)</td>
<td>Mizuki</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Takumi</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Korean (ko-KR)</td>
<td>Seoyeon</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Norwegian (nb-NO)</td>
<td>Liv</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
### Available Voices

<table>
<thead>
<tr>
<th>Language</th>
<th>Name/ID</th>
<th>Gender</th>
<th>Neural Voice</th>
<th>Standard Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polish (pl-PL)</td>
<td>Ewa</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Maja</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Jacek</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Jan</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Portuguese (Brazilian) (pt-BR)</td>
<td>Camila</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Vitória/Vitoria</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Ricardo</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Portuguese (European) (pt-PT)</td>
<td>Inês/Ines</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Cristiano</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Romanian (ro-RO)</td>
<td>Carmen</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Russian (ru-RU)</td>
<td>Tatyana</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Maxim</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Spanish (European) (es-ES)</td>
<td>Conchita</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Lucia</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Enrique</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Spanish (Mexican) (es-MX)</td>
<td>Mia</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Spanish (US) (es-US)</td>
<td>Lupe**</td>
<td>Female</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Penélope/Penelope</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Miguel</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Swedish (sv-SE)</td>
<td>Astrid</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Turkish (tr-TR)</td>
<td>Filiz</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Welsh (cy-GB)</td>
<td>Gwyneth</td>
<td>Female</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* This voice is bilingual and can speak both English and Hindi. For more information, see [Bilingual Voices (p. 15)](#).

** These voices can be used with Newscaster speaking styles when used with the Neural format. For more information, see [NTTS Newscaster Speaking Style (p. 97)](#).
In addition to the above voices, Amazon Polly can build you a custom Brand Voice that reflects your brand persona, providing you the opportunity to offer unique and exclusive NTTS voices to your customers. To learn more about Amazon Polly Brand Voices, please see Brand Voice.

Bilingual Voices

Amazon Polly has two ways of producing bilingual voices:

- Accented bilingual voices (p. 15)
- Fully bilingual voices (p. 15)

Accented bilingual voices

Accented bilingual voices can be created using any Amazon Polly voice, but only when using SSML tags. Normally, all words in the input text are spoken in the default language of the voice specified you're using.

For example, if you're using the voice of Joanna (who speaks US English), Amazon Polly speaks the following in the Joanna voice without a French accent:

```xml
<speak>
  Why didn't she just say, 'Je ne parle pas français?'
</speak>
```

In this case, the words *Je ne parle pas français* are spoken as they would be if they were English.

However, if you use the Joanna voice with the `<lang>` tag, Amazon Polly speaks the sentence in the Joanna voice in American-accented French:

```xml
<speak>
  Why didn't she just say, <lang xml:lang="fr-FR">'Je ne parle pas français?'</lang>.
</speak>
```

Because Joanna is not a native French voice, pronunciation is based on her native language, US English. For instance, although perfect French pronunciation features an uvular trill /ʁ/ in the word *français*, Joanna's US English voice pronounces this phoneme as the corresponding sound /r/.

If you use the voice of Giorgio, who speaks Italian, with the following text, Amazon Polly speaks the sentence in Giorgio's voice with an Italian pronunciation:

```xml
<speak>
  Mi piace Bruce Springsteen.
</speak>
```

Fully bilingual voices

A fully bilingual voice like Aditi (Indian English and Hindi) can speak two languages fluently. This gives you the ability to use words and phrases from both languages in a single text using the same voice.

Currently, Aditi is the only fully bilingual voice available.

Using a Bilingual Voice (Aditi)

Aditi speaks both Indian English (en-IN) and Hindi (hi-IN) fluently. You can synthesize speech in both English and Hindi, and the voice can switch between the two languages even within the same sentence.
Hindi can be used in two different forms:

- Devanagari: “उसेन कहँा, खेल तोह अब शुूर होगा”
- Romanagari (using the Latin alphabet): "Usne kahan, khel toh ab shuru hoga"

Additionally, it's possible to mix English and Hindi of either or both forms within a single sentence:

- Devanagari + English: "This is the song कभी कभी अदिति"
- Romanagari + English: "This is the song from the movie Jaane Tu Ya Jaane Na."
- Devanagari + Romanagari + English: "This is the song कभी कभी अदिति from the movie Jaane Tu Ya Jaane Na."

Because Aditi is a bilingual voice, text in all of these cases will be read correctly, as Amazon Polly can differentiate between the languages and scripts.

Amazon Polly also supports numbers, dates, times, and currency expansion in both English (Arabic numerals) and Hindi (Devanagari numerals). By default, Arabic numerals are read in Indian English. To make Amazon Polly read them in Hindi, you must use the `hi-IN` language code parameter.

Listening to the Voices

You can use the Amazon Polly console to hear a sample from any of the voices available in Amazon Polly.

To listen to a voice in Amazon Polly

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Text-to-Speech tab.
3. For Engine, choose Standard or Neural.
4. Choose a language and a Region, then choose a voice.
5. Enter text for the voice to speak or use the default phrase, and then choose Listen.

You can choose any of the languages offered by Amazon Polly and the console will display the voices available for that language. In most cases, there will be at least one male and one female voice, often more than one of each. A few only have a single voice. For a complete list, see Voices in Amazon Polly (p. 12)

Note
The inventory of voices and the number of languages included is continually being updated to include additional choices. To suggest a new language or voice, feel free to provide feedback on this page. Unfortunately, we are not able to comment on plans for specific new languages before they are released.

Each voice is created using native language speakers, so there are variations from voice to voice, even within the same language. When selecting a voice for your project, you should test each of the possible voices with a passage of text to see which best suits your needs.

Voice Speed

Because of the natural variation between voices, each available voice will speak the text at slightly different speeds. For instance, with US English voices, Ivy and Joanna are slightly faster than Matthew when saying "Mary had a little lamb," and considerably faster than Joey.
Since there is so much variation between voices, and the degree of that variation can depend on the
text being spoken, no standard speed (words per minute) is available for Amazon Polly voices. However,
you can find how long it takes for your voice to say the selected text using SpeechMarks. For more
information on using speechmarks in Amazon Polly, see Using Speech Marks (p. 100)

To see approximately how long it takes to speak a text passage

1. Open the AWS CLI.
2. Run the following code, filling in as needed

   ```bash
   aws polly synthesize-speech \
   --language-code [optional language code if needed] \
   --output-format json \
   --voice-id [name of desired voice] \
   --text ' [desired text] ' \
   --speech-mark-types=' ["viseme"] ' \
   LengthOfText.txt
   ```
3. Open LengthOfText.txt

If the text were "Mary had a little lamb," the last few lines returned by Amazon Polly would be:

   ```json
   {"time":882,"type":"viseme","value":"t"}
   {"time":964,"type":"viseme","value":"a"}
   {"time":1082,"type":"viseme","value":"p"}
   ```

The last viseme, essentially the sound for the final letters in "lamb" starts 1082 milliseconds after the
beginning of the speech. While this is not exactly the length of the audio, it's close and can serve as the
basis for comparison between voices.

**Changing Your Voice Speed**

For certain applications, you may find that you'd prefer the voice you like be slowed down, or speeded
up. If the speed of the voice is a concern, Amazon Polly provides the ability to modify this using SSML
tags.

For example:

Your organization is making an application that reads books to immigrant audiences. The audience
speaks English, but their fluency is limited. In this case, you might consider slowing the rate of speech to
give your audience a little more time for comprehension while the application is speaking.

Amazon Polly helps you slow down the rate of speech using the SSML `<prosody>` tag, as in:

   ```xml
   <speak>
   In some cases, it might help your audience to <prosody rate="85%">slow
   the speaking rate slightly to aid in comprehension.</prosody>
   </speak>
   ```

or

   ```xml
   <speak>
   In some cases, it might help your audience to <prosody rate="slow">slow
   the speaking rate slightly to aid in comprehension.</prosody>
   </speak>
   ```

Two speed options are available to you when using SSML with Amazon Polly:
• Preset speeds: x-slow, slow, medium, fast, and x-fast. In these cases, the speed of each option is approximate, depending on your preferred voice. The medium option is the normal speed of the voice.

• n% of speech rate: any percentage of the speech rate, between 20% and 200% can be used. In these cases, you can choose exactly the speed you want. However, the actual speed of the voice is approximate, depending on the voice you've chosen. 100% is considered to be the normal speed of the voice.

Because the speed of each option is approximate and depends on the voice you choose, we recommend that you test your selected voice at various speeds to see what exactly meets your needs.

For more information on using the prosody tag to best effect, see Controlling Volume, Speaking Rate, and Pitch (p. 117)

Languages Supported by Amazon Polly

The following languages are supported by Amazon Polly and can be used to synthesize speech. With each language is the language code. These language codes are W3C language identification tags (ISO 639-3 for the language name and ISO 3166 for the country code).

For in-depth tables showing the phonemes and visemes associated with each language, choose the link on each language in the table below.

<table>
<thead>
<tr>
<th>Language</th>
<th>Language Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic (p. 20)</td>
<td>arb</td>
</tr>
<tr>
<td>Chinese, Mandarin (p. 23)</td>
<td>cmn-CN</td>
</tr>
<tr>
<td>Danish (p. 25)</td>
<td>da-DK</td>
</tr>
<tr>
<td>Dutch (p. 28)</td>
<td>nl-NL</td>
</tr>
<tr>
<td>English, Australian (p. 30)</td>
<td>en-AU</td>
</tr>
<tr>
<td>English, British (p. 35)</td>
<td>en-GB</td>
</tr>
<tr>
<td>English, Indian (p. 37)</td>
<td>en-IN</td>
</tr>
<tr>
<td>English, New Zealand (p. 40)</td>
<td>en-NZ</td>
</tr>
<tr>
<td>English, South African (p. 44)</td>
<td>en-ZA</td>
</tr>
<tr>
<td>English, US (p. 33)</td>
<td>en-US</td>
</tr>
<tr>
<td>English, Welsh (p. 46)</td>
<td>en-GB-WLS</td>
</tr>
<tr>
<td>French (p. 49)</td>
<td>fr-FR</td>
</tr>
<tr>
<td>French, Canadian (p. 51)</td>
<td>fr-CA</td>
</tr>
<tr>
<td>Hindi (p. 56)</td>
<td>hi-IN</td>
</tr>
<tr>
<td>German (p. 53)</td>
<td>de-DE</td>
</tr>
<tr>
<td>Icelandic (p. 58)</td>
<td>is-IS</td>
</tr>
<tr>
<td>Italian (p. 61)</td>
<td>it-IT</td>
</tr>
</tbody>
</table>
The following tables list the phonemes for the languages supported by Amazon Polly, along with examples and the corresponding visemes.

**Topics**
- Arabic (arb) (p. 20)
- Chinese, Mandarin (cmn-CN) (p. 23)
- Danish (da-DK) (p. 25)
- Dutch (nl-NL) (p. 28)
- English, Australian (en-AU) (p. 30)
- English, American (en-US) (p. 33)
- English, British (en-GB) (p. 35)
- English, Indian (en-IN) (p. 37)
- English, New Zealand (en-NZ) (p. 40)
- English, South African (en-ZA) (p. 44)
- English, Welsh (en-GB-WSL) (p. 46)
- French (fr-FR) (p. 49)
- French, Canadian (fr-CA) (p. 51)
• German (de-DE) (p. 53)
• Hindi (hi-IN) (p. 56)
• Icelandic (is-IS) (p. 58)
• Italian (it-IT) (p. 61)
• Japanese (ja-JP) (p. 63)
• Korean (ko-KR) (p. 65)
• Norwegian (nb-NO) (p. 66)
• Polish (pl-PL) (p. 69)
• Portuguese (pt-PT) (p. 71)
• Portuguese, Brazilian (pt-BR) (p. 73)
• Romanian (ro-RO) (p. 75)
• Russian (ru-RU) (p. 77)
• Spanish (es-ES) (p. 79)
• Spanish, Mexican (es-MX) (p. 81)
• Spanish, US (es-US) (p. 83)
• Swedish (sv-SE) (p. 85)
• Turkish (tr-TR) (p. 87)
• Welsh (cy-GB) (p. 90)

Arabic (arb)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Arabic voice of Zeina that is supported by Amazon Polly.

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʔ</td>
<td>?</td>
<td>glottal stop</td>
<td>أَنا</td>
<td></td>
</tr>
<tr>
<td>ء</td>
<td>?\</td>
<td>voiced pharyngeal fricative</td>
<td>عُمر</td>
<td>k</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>بَلَد</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>دَارِي</td>
<td>t</td>
</tr>
<tr>
<td>dˤ</td>
<td>d_?\</td>
<td>emphatic voiced alveolar plosive</td>
<td>ضَّوۡء</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>جَمِيل</td>
<td>S</td>
</tr>
<tr>
<td>δ</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>دَلَلَة</td>
<td>T</td>
</tr>
<tr>
<td>δˤ</td>
<td>D_?\</td>
<td>emphatic voiced dental fricative</td>
<td>طَلَام</td>
<td>T</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>---------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>قُصُول</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>إنجلترا</td>
<td>k</td>
</tr>
<tr>
<td>ɣ</td>
<td>G</td>
<td>voiced velar fricative</td>
<td>غَرب</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>هذا</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>يَمَشِي</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>لَبِيب</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>لاقِي</td>
<td>t</td>
</tr>
<tr>
<td>tˤ</td>
<td>t_ˤ</td>
<td>emphatic voiceless alveolar plosive</td>
<td>الله</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>مَاذا</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>نُور</td>
<td>t</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>جِبَس</td>
<td>p</td>
</tr>
<tr>
<td>q</td>
<td>q</td>
<td>voiceless uvular plosive</td>
<td>قَرِيب</td>
<td>k</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>رَوْل</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>سَوَال</td>
<td>s</td>
</tr>
<tr>
<td>sˤ</td>
<td>s_ˤ\</td>
<td>emphatic voiceless alveolar fricative</td>
<td>صِيَاح</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>شَلَّر</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>سَمَر</td>
<td>t</td>
</tr>
<tr>
<td>tˤ</td>
<td>t_ˤ\</td>
<td>emphatic voiceless alveolar plosive</td>
<td>طالِب</td>
<td>t</td>
</tr>
<tr>
<td>θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>سَلَات</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>في ساميَّين</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labio-velar approximant</td>
<td>وَلَدَ</td>
<td>u</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>غَوْف</td>
<td>k</td>
</tr>
</tbody>
</table>
### IPA | X-SAMPA | Description | Example | Viseme
---|---|---|---|---

| h | X\ | voiceless pharyngeal fricative | حَولَ | k |
| z | z | voiced alveolar fricative | زُور | s |

#### Vowels

| a | a | open front unrounded vowel | بَرد | a |
| aː | aː | long open front unrounded vowel | دَار | a |
| ɑˤ | A_?\ | emphatic open back unrounded vowel | طَبئَل | a |
| ɑˤː | A_?ː | emphatic long open back unrounded vowel | تَالِم | a |
| u | u | close back rounded vowel | شَرِب | u |
| uː | uː | long close back rounded vowel | سُور | u |
| uˤ | u_?\ | emphatic close back rounded vowel | طَول | u |
| uˤː | u_?ː | emphatic long close back rounded vowel | طَول | u |
| i | i | close front unrounded vowel | بَينِ | i |
| iː | iː | long close front unrounded vowel | جَزيئ | i |
| iˤ | i_?\ | emphatic close front unrounded vowel | ضَدَّ | i |
| iˤː | i_?ː | emphatic long close front unrounded vowel | مَاضِي | i |
| e | e | close-mid front unrounded vowel | مَاركَت | e |
| eː | eː | long close-mid front unrounded vowel | مَودِيِل | e |
| ɔ | O | open-mid back rounded vowel | تَكنُولوژِي | O |
| ɔː | Oː | long open-mid back rounded vowel | سَليڤرزيون | O |
### Chinese, Mandarin (cmn-CN)

The following table lists the Pinyin and International Phonetic Alphabet (IPA) phonemes for the Mandarin Chinese voice that is supported by Amazon Polly. Pinyin is the international standard for Standard Chinese romanization. IPA and X-SAMPA are not commonly used but are available for English support. The IPA and X-SAMPA symbols in the table are for reference only and should not be used for Chinese transcription. Pinyin examples and the corresponding visemes are also shown.

To make Amazon Polly use phonetic pronunciation in Pinyin, use the `phoneme alphabet="x-amazon-phonetic standard used"` tag.

The following examples show this with each standard.

#### Pinyin:

```xml
<speak>
  ## <phoneme alphabet="x-amazon-pinyin" ph="bo2">#</phoneme>#
  ## <phoneme alphabet="x-amazon-pinyin" ph="bao2">#</phoneme>#
</speak>
```

#### IPA:

```xml
<speak>
  ## <phoneme alphabet="ipa" ph="p##k##n">pecan</phoneme>#
  ## <phoneme alphabet="ipa" ph="#pi.kæn">pecan</phoneme>#
</speak>
```

#### X-SAMPA:

```xml
<speak>
  ## <phoneme alphabet='x-sampa' ph='pI"kA:n'>pecan</phoneme>#
  ## <phoneme alphabet='x-sampa' ph='"pi.k{n'>pecan</phoneme>#
</speak>
```

**Note**

Amazon Polly accepts Mandarin Chinese input encoded in UTF-8 only. The GB 18030 encoding standard is not currently supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>Pinyin</th>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Pinyin Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>发, fa1</td>
<td>f</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>和, he2</td>
<td>k</td>
</tr>
<tr>
<td>g</td>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>古, gu3</td>
<td>k</td>
</tr>
<tr>
<td>k</td>
<td>kʰ</td>
<td>kʰ</td>
<td>aspirated voiceless velar plosive</td>
<td>苦, ku3</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>拉, la1</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>骂, ma4</td>
<td>p</td>
</tr>
</tbody>
</table>
### Phoneme and Viseme Tables for Supported Languages

<table>
<thead>
<tr>
<th>Pinyin</th>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Pinyin Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>那, na4</td>
<td>t</td>
</tr>
<tr>
<td>ng</td>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>正, zheng4</td>
<td>k</td>
</tr>
<tr>
<td>b</td>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>爸, ba4</td>
<td>p</td>
</tr>
<tr>
<td>p</td>
<td>pʰ</td>
<td>p_h</td>
<td>aspirated voiceless bilabial plosive</td>
<td>怕, pa4</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>四, si4</td>
<td>s</td>
</tr>
<tr>
<td>x</td>
<td>ɕ</td>
<td>s\</td>
<td>voiceless alveolo-palatal fricative</td>
<td>西, xi4</td>
<td>J</td>
</tr>
<tr>
<td>sh</td>
<td>ş</td>
<td>s`</td>
<td>voiceless retroflex fricative</td>
<td>是, shi4</td>
<td>S</td>
</tr>
<tr>
<td>d</td>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>打, da3</td>
<td>t</td>
</tr>
<tr>
<td>t</td>
<td>tʰ</td>
<td>t_h</td>
<td>aspirated voiceless alveolar plosive</td>
<td>他, ta1</td>
<td>t</td>
</tr>
<tr>
<td>zh</td>
<td>ʈʂ</td>
<td>t<code>s</code></td>
<td>voiceless retroflex affricate</td>
<td>之, zhi1</td>
<td>S</td>
</tr>
<tr>
<td>ch</td>
<td>ʈʂʰ</td>
<td>t<code>s</code>_h</td>
<td>aspirated voiceless retroflex affricate</td>
<td>吃, chi1</td>
<td>S</td>
</tr>
<tr>
<td>s</td>
<td>ʂ</td>
<td>ts</td>
<td>voiceless alveolar affricate</td>
<td>字, zi4</td>
<td>s</td>
</tr>
<tr>
<td>j</td>
<td>ʈɕ</td>
<td>ts\</td>
<td>voiceless alveolo-palatal affricate</td>
<td>鸡, ji1</td>
<td>J</td>
</tr>
<tr>
<td>q</td>
<td>ʈʂʰ</td>
<td>ts_h</td>
<td>aspirated voiceless alveolo-palatal affricate</td>
<td>七, qi1</td>
<td>J</td>
</tr>
<tr>
<td>c</td>
<td>ʈɕʰ</td>
<td>ts_h</td>
<td>aspirated voiceless alveolar affricate</td>
<td>次, ci4</td>
<td>s</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>w</td>
<td>labio-velar approximant</td>
<td>我, wo3</td>
<td>u</td>
</tr>
<tr>
<td>r</td>
<td>ʐ</td>
<td>z`</td>
<td>voiced retroflex fricative</td>
<td>日, ri4</td>
<td>S</td>
</tr>
</tbody>
</table>

"er" and "r" colored syllables

<table>
<thead>
<tr>
<th>er</th>
<th>ɚ</th>
<th>r-coloured mid central vowel</th>
<th>二, er4</th>
<th>@</th>
</tr>
</thead>
<tbody>
<tr>
<td>-r</td>
<td></td>
<td>r-colored syllable</td>
<td>馅儿, xianr4</td>
<td>@</td>
</tr>
</tbody>
</table>

### Vowels

<table>
<thead>
<tr>
<th>e</th>
<th>ɤ</th>
<th>close-mid back unrounded vowel</th>
<th>恶, e4</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>a</td>
<td>mid central vowel</td>
<td>恩, en1</td>
<td>@</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>安, an1</td>
<td>a</td>
</tr>
<tr>
<td>ai</td>
<td>a1</td>
<td>diphthong</td>
<td>爱, ai4</td>
<td>a</td>
</tr>
</tbody>
</table>
Danish (da-DK)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Danish voices that are supported by Amazon Polly.

### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bat</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>da</td>
<td>t</td>
</tr>
<tr>
<td>Ø</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>mad, thriller</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fat</td>
<td>f</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------------------------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>gat</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>hat</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>jo</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>kat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>ladt</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mat</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nay</td>
<td>t</td>
</tr>
<tr>
<td>ñ</td>
<td>N</td>
<td>velar nasal</td>
<td>lang</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pande</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>thriller, story</td>
<td>r</td>
</tr>
<tr>
<td>ð</td>
<td>R</td>
<td>voiced uvular fricative</td>
<td>rat</td>
<td>k</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>sat</td>
<td>s</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tal</td>
<td>t</td>
</tr>
<tr>
<td>û</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vat</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>hav, weekend</td>
<td>u</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| ð | 2 | close-mid front rounded vowel | øst | o       |
| ð: | 2: | long close-mid front rounded vowel | øse | o       |
| ð | 6 | near-open central vowel | mor | a       |
| ð | 9 | open-mid front rounded vowel | skøn, grønt | O |
| ð: | 9: | long open-mid front rounded vowel | høne, gøre | O |
| ð | @ | mid central vowel | ane | @       |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>æː</td>
<td>{</td>
<td>long near-open front unrounded vowel</td>
<td>male</td>
<td>a</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>man</td>
<td>a</td>
</tr>
<tr>
<td>æ</td>
<td>{</td>
<td>near-open front unrounded vowel</td>
<td>adresse</td>
<td>a</td>
</tr>
<tr>
<td>o</td>
<td>A</td>
<td>open back unrounded vowel</td>
<td>lak, tak</td>
<td>a</td>
</tr>
<tr>
<td>oː</td>
<td>Aː</td>
<td>long open back unrounded vowel</td>
<td>rase</td>
<td>a</td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>close-mid front unrounded vowel</td>
<td>midt</td>
<td>e</td>
</tr>
<tr>
<td>eː</td>
<td>eː</td>
<td>long close-mid front unrounded vowel</td>
<td>mele</td>
<td>e</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>mæt</td>
<td>E</td>
</tr>
<tr>
<td>ɛː</td>
<td>Eː</td>
<td>long open-mid front unrounded vowel</td>
<td>mæle</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>mit</td>
<td>i</td>
</tr>
<tr>
<td>iː</td>
<td>iː</td>
<td>long close front unrounded vowel</td>
<td>mile</td>
<td>i</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>foto</td>
<td>o</td>
</tr>
<tr>
<td>oː</td>
<td>oː</td>
<td>long close-mid back rounded vowel</td>
<td>mole</td>
<td>o</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>mund</td>
<td>O</td>
</tr>
<tr>
<td>ɔː</td>
<td>Oː</td>
<td>long open-mid back rounded vowel</td>
<td>måle</td>
<td>O</td>
</tr>
<tr>
<td>ɔː̃</td>
<td>Oː</td>
<td>long open back rounded vowel</td>
<td>morse</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>lusk</td>
<td>u</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close back rounded vowel</td>
<td>mule</td>
<td>u</td>
</tr>
<tr>
<td>ʌ</td>
<td>V</td>
<td>open-mid back unrounded vowel</td>
<td>kører</td>
<td>E</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>close front rounded vowel</td>
<td>yt</td>
<td>u</td>
</tr>
</tbody>
</table>
### IPA, X-SAMPA, Description, Example, Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>y:</td>
<td>y:</td>
<td>long close front rounded vowel</td>
<td>hyle</td>
<td>u</td>
</tr>
</tbody>
</table>

### Additional Symbols

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>&quot;</td>
<td>primary stress</td>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>%</td>
<td>secondary stress</td>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
<td></td>
</tr>
</tbody>
</table>

### Dutch (nl-NL)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Dutch voices that are supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bak</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dak</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>manager</td>
<td>S</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fel</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>goal</td>
<td>k</td>
</tr>
<tr>
<td>γ</td>
<td>G</td>
<td>voiced velar fricative</td>
<td>hoed</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h\</td>
<td>voiced glottal fricative</td>
<td>hand</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>ja</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>kap</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>land</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>met</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>net</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>bang</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pak</td>
<td>p</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>rand</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>sein</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>show</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tak</td>
<td>t</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vel</td>
<td>f</td>
</tr>
<tr>
<td>ʌ</td>
<td>v\</td>
<td>labiodental approximant</td>
<td>wit</td>
<td>f</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>toch</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>ziin</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>bagage</td>
<td>S</td>
</tr>
<tr>
<td>øː</td>
<td>2ː</td>
<td>long close-mid front rounded vowel</td>
<td>neus</td>
<td>o</td>
</tr>
<tr>
<td>œy</td>
<td>9y</td>
<td>diphthong</td>
<td>buit</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>@</td>
<td>mid central vowel</td>
<td>de</td>
<td>@</td>
</tr>
<tr>
<td>aː</td>
<td>aː</td>
<td>long open front unrounded vowel</td>
<td>baad</td>
<td>a</td>
</tr>
<tr>
<td>ɑː</td>
<td>A</td>
<td>open back unrounded vowel</td>
<td>bad</td>
<td>a</td>
</tr>
<tr>
<td>eː</td>
<td>eː</td>
<td>long close-mid front unrounded vowel</td>
<td>beet</td>
<td>e</td>
</tr>
<tr>
<td>æː</td>
<td>3ː</td>
<td>long open-mid central unrounded vowel</td>
<td>barrière</td>
<td>E</td>
</tr>
<tr>
<td>e</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>bed</td>
<td>E</td>
</tr>
<tr>
<td>εi</td>
<td>Ei</td>
<td>diphthong</td>
<td>beet</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>vier</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>l</td>
<td>near-close near-front unrounded vowel</td>
<td>pit</td>
<td>i</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>o:</td>
<td>o:</td>
<td>long close-mid back</td>
<td>boot</td>
<td>o</td>
</tr>
<tr>
<td>o</td>
<td>O</td>
<td>open-mid back rounded</td>
<td>pot</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded</td>
<td>hoed</td>
<td>u</td>
</tr>
<tr>
<td>au</td>
<td>Vu</td>
<td>diphthong</td>
<td>fout</td>
<td>E</td>
</tr>
<tr>
<td>y:</td>
<td>y:</td>
<td>long close front</td>
<td>fuut</td>
<td>u</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>near-close near-front</td>
<td>hut</td>
<td>u</td>
</tr>
</tbody>
</table>

**Additional Symbols**

- `ˈ` primary stress Alabama
- `ˌ` secondary stress Alabama
- `.'` syllable boundary A.la.ba.ma

**English, Australian (en-AU)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Australian English voices that are supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>d̠</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>l=</td>
<td>l=</td>
<td>syllabic alveolar lateral approximant</td>
<td>battle</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>m=</td>
<td>m=</td>
<td>syllabic bilabial nasal</td>
<td>anthem</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>n=</td>
<td>n=</td>
<td>syllabic alveolar nasal</td>
<td>button</td>
<td>t</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pin</td>
<td>p</td>
</tr>
<tr>
<td>r\</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>s</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>task</td>
<td>t</td>
</tr>
<tr>
<td>tS</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>Z</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>vision</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<table>
<thead>
<tr>
<th>a</th>
<th>@</th>
<th>mid central vowel</th>
<th>arena</th>
<th>@</th>
</tr>
</thead>
<tbody>
<tr>
<td>æu</td>
<td>@U</td>
<td>diphthong</td>
<td>goat</td>
<td>@</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>æ</td>
<td>{</td>
<td>near open-front unrounded vowel</td>
<td>trap</td>
<td>a</td>
</tr>
<tr>
<td>aɪ</td>
<td>al</td>
<td>diphthong</td>
<td>price</td>
<td>a</td>
</tr>
<tr>
<td>au</td>
<td>aU</td>
<td>diphthong</td>
<td>mouth</td>
<td>a</td>
</tr>
<tr>
<td>σ:</td>
<td>A:</td>
<td>long open-back unrounded vowel</td>
<td>father</td>
<td>a</td>
</tr>
<tr>
<td>ɔɪ</td>
<td>el</td>
<td>diphthong</td>
<td>face</td>
<td>e</td>
</tr>
<tr>
<td>ɔː</td>
<td>3:</td>
<td>long open mid-central unrounded vowel</td>
<td>nurse</td>
<td>E</td>
</tr>
<tr>
<td>ᵇ</td>
<td>E</td>
<td>open mid-front unrounded vowel</td>
<td>dress</td>
<td>E</td>
</tr>
<tr>
<td>ᵇʊ</td>
<td>E@</td>
<td>diphthong</td>
<td>square</td>
<td>E</td>
</tr>
<tr>
<td>i:</td>
<td>i</td>
<td>long close front unrounded vowel</td>
<td>fleece</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>I</td>
<td>near-close near-front unrounded vowel</td>
<td>kit</td>
<td>i</td>
</tr>
<tr>
<td>ɪə</td>
<td>I@</td>
<td>diphthong</td>
<td>near</td>
<td>i</td>
</tr>
<tr>
<td>ɔː</td>
<td>OI</td>
<td>long open-mid back rounded vowel</td>
<td>thought</td>
<td>O</td>
</tr>
<tr>
<td>ɔɪ</td>
<td>OI</td>
<td>Diphthong</td>
<td>choice</td>
<td>O</td>
</tr>
<tr>
<td>ʊ</td>
<td>Q</td>
<td>open back rounded vowel</td>
<td>lot</td>
<td>O</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>foot</td>
<td>u</td>
</tr>
<tr>
<td>uə</td>
<td>U@</td>
<td>diphthong</td>
<td>cure</td>
<td>u</td>
</tr>
<tr>
<td>ʌ</td>
<td>V</td>
<td>Open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
</tbody>
</table>

**Additional Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><code> </code></td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td><code>, </code></td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td><code>. </code></td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>
# English, American (en-US)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the American English voices that are supported by Amazon Polly.

## Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>speak</td>
<td>p</td>
</tr>
<tr>
<td>r\</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>ŋ</td>
<td>ŋ</td>
<td>voiceless postalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>trap</td>
<td>t</td>
</tr>
</tbody>
</table>
### IPA X-SAMPA Description Example Viseme

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>tʃ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>vision</td>
<td>S</td>
</tr>
</tbody>
</table>

### Vowels

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ə</td>
<td>@</td>
<td>mid-central vowel</td>
<td>arena</td>
<td>@</td>
</tr>
<tr>
<td>Ʉ</td>
<td>@`</td>
<td>mid-central r-colored vowel</td>
<td>reader</td>
<td>@</td>
</tr>
<tr>
<td>æ</td>
<td>{</td>
<td>near open-front unrounded vowel</td>
<td>trap</td>
<td>a</td>
</tr>
<tr>
<td>aɪ</td>
<td>ai</td>
<td>diphthong</td>
<td>price</td>
<td>a</td>
</tr>
<tr>
<td>au</td>
<td>aU</td>
<td>diphthong</td>
<td>mouth</td>
<td>a</td>
</tr>
<tr>
<td>aɪ</td>
<td>əl</td>
<td>diphthong</td>
<td>face</td>
<td>e</td>
</tr>
<tr>
<td>ə</td>
<td>ə`</td>
<td>open mid-central unrounded r-colored vowel</td>
<td>nurse</td>
<td>E</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open mid-front unrounded vowel</td>
<td>dress</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>long close front unrounded vowel</td>
<td>fleece</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>ɪ</td>
<td>near-close near-front unrounded vowel</td>
<td>kit</td>
<td>i</td>
</tr>
<tr>
<td>oʊ</td>
<td>oU</td>
<td>diphthong</td>
<td>goat</td>
<td>o</td>
</tr>
<tr>
<td>o</td>
<td>O</td>
<td>long open mid-back rounded vowel</td>
<td>thought</td>
<td>O</td>
</tr>
<tr>
<td>ɔɪ</td>
<td>Oɪ</td>
<td>diphthong</td>
<td>choice</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
</tbody>
</table>
### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʊ</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>foot</td>
<td>u</td>
</tr>
<tr>
<td>θ</td>
<td>V</td>
<td>open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
</tbody>
</table>

#### Additional Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>,</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

### English, British (en-GB)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the British English voices that are supported by Amazon Polly.

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>l̩</td>
<td>l=</td>
<td>syllabic alveolar lateral approximant</td>
<td>battle</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>----------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>m̩</td>
<td>m=</td>
<td>syllabic bilabial nasal</td>
<td>anthem</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>n=</td>
<td>syllabic alveolar nasal</td>
<td>button</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pin</td>
<td>p</td>
</tr>
<tr>
<td>r\</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>S</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>task</td>
<td>t</td>
</tr>
<tr>
<td>tS</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>Z</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>vision</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| a   | @       | mid central vowel               | arena   | @      |
| A:  | A:      | long open-back unrounded vowel   | father  | a      |
| a1  | aI      | diphthong                        | price   | a      |
| au  | aU      | diphthong                        | mouth   | a      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ɜː</td>
<td>ɜː</td>
<td>long open mid-central unrounded vowel</td>
<td>nurse</td>
<td>E</td>
</tr>
<tr>
<td>ε</td>
<td>E</td>
<td>open mid-front unrounded vowel</td>
<td>dress</td>
<td>E</td>
</tr>
<tr>
<td>ɛə</td>
<td>ɛ@</td>
<td>diphthong</td>
<td>square</td>
<td>E</td>
</tr>
<tr>
<td>i:</td>
<td>i</td>
<td>long close front unrounded vowel</td>
<td>fleece</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>I</td>
<td>near-close near-front unrounded vowel</td>
<td>kit</td>
<td>i</td>
</tr>
<tr>
<td>ɪə</td>
<td>ɪ@</td>
<td>diphthong</td>
<td>near</td>
<td>i</td>
</tr>
<tr>
<td>ɔː</td>
<td>ɔː</td>
<td>long open-mid back rounded vowel</td>
<td>thought</td>
<td>O</td>
</tr>
<tr>
<td>ɔɪ</td>
<td>ɔɪ</td>
<td>Diphthong</td>
<td>choice</td>
<td>O</td>
</tr>
<tr>
<td>ō</td>
<td>ɑː</td>
<td>open back rounded vowel</td>
<td>lot</td>
<td>O</td>
</tr>
<tr>
<td>ū</td>
<td>ʊː</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
<tr>
<td>ūə</td>
<td>ʊ@</td>
<td>diphthong</td>
<td>cure</td>
<td>u</td>
</tr>
<tr>
<td>ʌ</td>
<td>V</td>
<td>Open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
</tbody>
</table>

**Additional Symbols**

| ˈ  | "      | primary stress                  | Alabama  |
| ˌ  | %      | secondary stress                | Alabama  |
| ˌ  | .      | syllable boundary               | A.la.ba.ma |

**English, Indian (en-IN)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Indian English voice supported by Amazon Polly.

For additional phonemes used in conjunction with Indian English, see Hindi (hi-IN) (p. 56).

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
</table>

**Consonants**
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>l̩</td>
<td>l=</td>
<td>syllabic alveolar lateral approximant</td>
<td>battle</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>m̩</td>
<td>m=</td>
<td>syllabic bilabial nasal</td>
<td>anthem</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>n̩</td>
<td>n=</td>
<td>syllabic alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pin</td>
<td>p</td>
</tr>
<tr>
<td>ṛ</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>task</td>
<td>t</td>
</tr>
<tr>
<td>tʃ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>---------------------------------------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>Θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>vision</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| a   | @       | mid central vowel                    | arena    | @      |
| αu  | @U      | diphthong                             | goat     | @      |
| æ   | {       | near open-front unrounded vowel       | trap     | a      |
| aI  | al      | diphthong                             | price    | a      |
| au  | aU      | diphthong                             | mouth    | a      |
| oː  | Aː      | long open-back unrounded vowel        | father   | a      |
| eɪ  | el      | diphthong                             | face     | e      |
| ɔː  | ɔː      | long open mid-central unrounded vowel | nurse    | E      |
| e   | E       | open mid-front unrounded vowel        | dress    | E      |
| eə  | E@      | diphthong                             | square   | E      |
| iː  | i       | long close front unrounded vowel      | fleece   | i      |
| ɪ   | l       | near-close near-front unrounded vowel | kit      | i      |
| ɪə  | l@      | diphthong                             | near     | i      |
| ɔː  | Oː      | long open-mid back rounded vowel      | thought  | O      |
| ɔɪ  | Oɪ      | Diphthong                             | choice   | O      |
| œ   | Q       | open back rounded vowel               | lot      | O      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>u:</td>
<td>u:</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>foot</td>
<td>u</td>
</tr>
<tr>
<td>uə</td>
<td>U@</td>
<td>diphthong</td>
<td>cure</td>
<td>u</td>
</tr>
<tr>
<td>ñ</td>
<td>V</td>
<td>Open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
</tbody>
</table>

**Additional Symbols**

- ' primary stress
- , secondary stress
- . syllable boundary

**English, New Zealand (en-NZ)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the New Zealand English voices that are supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>ðdʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>--------------------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>l</td>
<td>l=</td>
<td>syllabic alveolar lateral approximant</td>
<td>battle</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>m̩</td>
<td>m=</td>
<td>syllabic bilabial nasal</td>
<td>anthem</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>n̩</td>
<td>n=</td>
<td>syllabic alveolar nasal</td>
<td>button</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pin</td>
<td>p</td>
</tr>
<tr>
<td>r̩</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>task</td>
<td>t</td>
</tr>
<tr>
<td>ʒ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>vision</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| a    | @      | mid central vowel                          | arena   | @      |
| au   | @U     | diphthong                                  | goat    | @      |
| æ    | {}     | near open-front unrounded vowel            | trap    | a      |
| at   | al     | diphthong                                  | price   | a      |
| au   | aU     | diphthong                                  | mouth   | a      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>aː</td>
<td>Aː</td>
<td>long open-back unrounded vowel</td>
<td>father</td>
<td>a</td>
</tr>
<tr>
<td>eɪ</td>
<td>el</td>
<td>diphthong</td>
<td>face</td>
<td>e</td>
</tr>
<tr>
<td>ɔː</td>
<td>ɔː</td>
<td>long open mid-central unrounded vowel</td>
<td>nurse</td>
<td>E</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open mid-front unrounded vowel</td>
<td>dress</td>
<td>E</td>
</tr>
<tr>
<td>ɛɔ</td>
<td>E@</td>
<td>diphthong</td>
<td>square</td>
<td>E</td>
</tr>
<tr>
<td>iː</td>
<td>i</td>
<td>long close front unrounded vowel</td>
<td>fleece</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>l</td>
<td>near-close near-front unrounded vowel</td>
<td>kit</td>
<td>i</td>
</tr>
<tr>
<td>ɪə</td>
<td>l@</td>
<td>diphthong</td>
<td>near</td>
<td>i</td>
</tr>
<tr>
<td>ɔː</td>
<td>Oː</td>
<td>long open-mid back rounded vowel</td>
<td>thought</td>
<td>O</td>
</tr>
<tr>
<td>ɔɪ</td>
<td>Oɪ</td>
<td>Diphthong</td>
<td>choice</td>
<td>O</td>
</tr>
<tr>
<td>ə</td>
<td>Q</td>
<td>open back rounded vowel</td>
<td>lot</td>
<td>O</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>foot</td>
<td>u</td>
</tr>
<tr>
<td>uə</td>
<td>U@</td>
<td>diphthong</td>
<td>cure</td>
<td>u</td>
</tr>
<tr>
<td>θ</td>
<td>V</td>
<td>Open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
</tbody>
</table>

**Additional Symbols**

- ′ " primary stress
- ′ % secondary stress
- ′ . syllable boundary

The Aria voice speaks New Zealand English and offers limited support for Maori. It can pronounce the following Maori words and phrases. The Maori phrases are case-sensitive.

<table>
<thead>
<tr>
<th>English</th>
<th>Maori</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello/cheers</td>
<td>Kia ora</td>
</tr>
<tr>
<td>Welcome (to)</td>
<td>Nau mai (ki)</td>
</tr>
<tr>
<td>English</td>
<td>Maori</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Hello (one person)/thank you</td>
<td>Tēnā koe</td>
</tr>
<tr>
<td>Hello (three or more people)/thank you</td>
<td>Tēnā koutou</td>
</tr>
<tr>
<td>Good morning</td>
<td>Ata mārie</td>
</tr>
<tr>
<td>Good morning</td>
<td>Mörena</td>
</tr>
<tr>
<td>Thank you</td>
<td>Ngā mihi</td>
</tr>
<tr>
<td>Take care</td>
<td>Ngā manaakitanga</td>
</tr>
<tr>
<td>See you</td>
<td>Ka kite</td>
</tr>
<tr>
<td>See you later</td>
<td>Mā te wā</td>
</tr>
<tr>
<td>Have a good day</td>
<td>Kia pai tō rā</td>
</tr>
<tr>
<td>Merry Christmas</td>
<td>Meri Kirihimete</td>
</tr>
<tr>
<td>Maori</td>
<td>Māori</td>
</tr>
<tr>
<td>Maori language</td>
<td>te reo Māori</td>
</tr>
<tr>
<td>Maori language week</td>
<td>Te wiki o te reo Māori</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Aotearoa</td>
</tr>
<tr>
<td>Maori New Year</td>
<td>Mātariki</td>
</tr>
<tr>
<td>Town in New Zealand / Waitangi Day is the</td>
<td>Waitangi</td>
</tr>
<tr>
<td>national day of New Zealand</td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>tahi</td>
</tr>
<tr>
<td>Two</td>
<td>rua</td>
</tr>
<tr>
<td>Three</td>
<td>toru</td>
</tr>
<tr>
<td>Four</td>
<td>whā</td>
</tr>
<tr>
<td>Five</td>
<td>rima</td>
</tr>
<tr>
<td>Six</td>
<td>ono</td>
</tr>
<tr>
<td>Seven</td>
<td>whitu</td>
</tr>
<tr>
<td>Eight</td>
<td>waru</td>
</tr>
<tr>
<td>Nine</td>
<td>iwa</td>
</tr>
<tr>
<td>Ten</td>
<td>tekau</td>
</tr>
<tr>
<td>Twenty</td>
<td>rua tekau</td>
</tr>
<tr>
<td>Thirty</td>
<td>Toru tekau</td>
</tr>
</tbody>
</table>
English, South African (en-ZA)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the South African English voices that are supported by Amazon Polly.

Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>δ</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>l̩</td>
<td>l=</td>
<td>syllabic alveolar lateral approximant</td>
<td>battle</td>
<td>t</td>
</tr>
<tr>
<td>ŋ</td>
<td>K</td>
<td>voiceless lateral fricative</td>
<td>umhlanga</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>m̩</td>
<td>m=</td>
<td>syllabic bilabial nasal</td>
<td>anthem</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>n̩</td>
<td>n=</td>
<td>syllabic alveolar nasal</td>
<td>button</td>
<td>t</td>
</tr>
<tr>
<td>ŋ</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pin</td>
<td>p</td>
</tr>
<tr>
<td>ŋ</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>pareis</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voicelesspostalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>task</td>
<td>t</td>
</tr>
<tr>
<td>xʃ</td>
<td>tʃS</td>
<td>voicelesspostalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>Θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>gauteng</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>!</td>
<td>!\</td>
<td>post-alveolar click</td>
<td>qqeberha</td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>/\</td>
<td>dental click</td>
<td>ncube</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Vowels**

<p>| ə   | @       | mid central vowel                               | arena    | @      |
| ai  | @i      | diphthong                                       | neslpruit | i     |
| əu  | @U      | diphthong                                       | goat     | @      |
| æ   | {       | near open-front unrounded vowel                 | trap     | a      |
| aɪ  | al      | diphthong                                       | price    | a      |
| aʊ  | aU      | diphthong                                       | mouth    | a      |
| ɑː  | A:      | long open-back unrounded vowel                  | father   | a      |
| ɛɪ  | el      | diphthong                                       | face     | e      |
| ɜː  | ɜː      | long open mid-central unrounded vowel           | nurse    | E      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open mid-front unrounded vowel</td>
<td>dress</td>
<td>E</td>
</tr>
<tr>
<td>ɛ@</td>
<td>E@</td>
<td>diphthong</td>
<td>square</td>
<td>E</td>
</tr>
<tr>
<td>iː</td>
<td>i</td>
<td>long close front unrounded vowel</td>
<td>fleece</td>
<td>i</td>
</tr>
<tr>
<td>i@</td>
<td>I@</td>
<td>diphthong</td>
<td>du preez</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>I</td>
<td>near-close near-front unrounded vowel</td>
<td>kit</td>
<td>i</td>
</tr>
<tr>
<td>i@</td>
<td>I@</td>
<td>diphthong</td>
<td>near</td>
<td>i</td>
</tr>
<tr>
<td>oː</td>
<td>Oː</td>
<td>long open-mid back rounded vowel</td>
<td>thought</td>
<td>O</td>
</tr>
<tr>
<td>oː</td>
<td>Oː</td>
<td>Diphthong</td>
<td>choice</td>
<td>O</td>
</tr>
<tr>
<td>ɒ</td>
<td>Q</td>
<td>open back rounded vowel</td>
<td>lot</td>
<td>O</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>foot</td>
<td>u</td>
</tr>
<tr>
<td>u@</td>
<td>U@</td>
<td>diphthong</td>
<td>cure</td>
<td>u</td>
</tr>
<tr>
<td>ʌ</td>
<td>V</td>
<td>Open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>close front rounded vowel</td>
<td>van vuuren</td>
<td>u</td>
</tr>
</tbody>
</table>

### Additional Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>′</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>‐</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

---

**English, Welsh (en-GB-WSL)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Welsh English voice supported by Amazon Polly.

### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>ð3</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yes</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cat</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lay</td>
<td>t</td>
</tr>
<tr>
<td>l\</td>
<td>l=</td>
<td>syllabic alveolar lateral approximant</td>
<td>battle</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mouse</td>
<td>p</td>
</tr>
<tr>
<td>m=</td>
<td>m=</td>
<td>syllabic bilabial nasal</td>
<td>anthem</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>n=</td>
<td>n=</td>
<td>syllabic alveolar nasal</td>
<td>nap</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>thing</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pin</td>
<td>p</td>
</tr>
<tr>
<td>r\</td>
<td>r\</td>
<td>alveolar approximant</td>
<td>red</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>seem</td>
<td>s</td>
</tr>
<tr>
<td>s</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>ship</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>task</td>
<td>t</td>
</tr>
<tr>
<td>t\</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chart</td>
<td>S</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>Θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>thin</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vest</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>west</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>vision</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

| a   | @       | mid central vowel           | arena   | @      |
| aʊ  | @U      | diphthong                   | goat    | @      |
| æ   | {       | near open-front unrounded vowel | trap   | a      |
| aɪ  | aI      | diphthong                   | price   | a      |
| aʊ  | aU      | diphthong                   | mouth   | a      |
| əː  | Aː      | long open-back unrounded vowel | father | a      |
| eɪ  | eI      | diphthong                   | face    | e      |
| əː  | əI      | long open mid-central unrounded vowel | nurse | E      |
| e   | E       | open mid-front unrounded vowel | dress  | E      |
| eʊ  | E@      | diphthong                   | square  | E      |
| iː  | i       | long close front unrounded vowel | fleece | i      |
| ɪ   | I       | near-close near-front unrounded vowel | kit    | i      |
| ɪʊ  | l@      | diphthong                   | near    | i      |
| ɔː  | Oː      | long open-mid back rounded vowel | thought | O      |
| ɔɪ  | OI      | Diphthong                   | choice  | O      |
| ɔ   | Q       | open back rounded vowel     | lot     | O      |
### IPA, X-SAMPA, Description, Example, Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>u:</td>
<td>u:</td>
<td>long close-back rounded vowel</td>
<td>goose</td>
<td>u</td>
</tr>
<tr>
<td>ʊ</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>foot</td>
<td>u</td>
</tr>
<tr>
<td>ʊə</td>
<td>U@</td>
<td>diphthong</td>
<td>cure</td>
<td>u</td>
</tr>
<tr>
<td>Ꚅ&gt;V</td>
<td></td>
<td>Open-mid-back unrounded vowel</td>
<td>strut</td>
<td>E</td>
</tr>
</tbody>
</table>

#### Additional Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>,</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

### French (fr-FR)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the French voices that are supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>boire</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>madame</td>
<td>t</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>femme</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>grand</td>
<td>k</td>
</tr>
<tr>
<td>ɥ</td>
<td>H</td>
<td>labial-palatal approximant</td>
<td>bruit</td>
<td>u</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>meilleur</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>quatre</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>malade</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>maison</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>astronome</td>
<td>t</td>
</tr>
<tr>
<td>n</td>
<td>J</td>
<td>palatal nasal</td>
<td>baigner</td>
<td>J</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>ŋ</td>
<td>N</td>
<td>velar nasal</td>
<td>parking</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pomme</td>
<td>p</td>
</tr>
<tr>
<td>ʁ</td>
<td>R</td>
<td>voiced uvular fricative</td>
<td>amoureux</td>
<td>k</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>santé</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>chat</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>téléphone</td>
<td>t</td>
</tr>
<tr>
<td>ν</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vrai</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>soir</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>raison</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>aubergine</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| ø   | 2       | close-mid front rounded vowel | deux    | o      |
| œ   | 9       | open-mid front rounded vowel  | neuf    | O      |
| œ̃  | 9~      | nasal open-mid front rounded vowel | brun  | O      |
| a   | @       | mid central vowel             | je      | @      |
| a   | a       | open front unrounded vowel    | table   | a      |
| ë   | A~      | nasal open back unrounded vowel | camembert| a      |
| e   | e       | close-mid front unrounded vowel | marché  | e      |
| ë   | E       | open-mid front unrounded vowel | neige   | E      |
| ê   | E~      | nasal open-mid front unrounded vowel | sapin | E      |
| i   | i       | close front unrounded vowel   | mille   | i      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>hôpital</td>
<td>o</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>homme</td>
<td>O</td>
</tr>
<tr>
<td>õ</td>
<td>O~</td>
<td>nasal open-mid back rounded vowel</td>
<td>bon</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>sous</td>
<td>u</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>close front rounded vowel</td>
<td>dur</td>
<td>u</td>
</tr>
</tbody>
</table>

**Additional Symbols**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ˈ</td>
<td>&quot;</td>
<td>primary stress</td>
</tr>
<tr>
<td>ˌ</td>
<td>%</td>
<td>secondary stress</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
</tr>
</tbody>
</table>

**French, Canadian (fr-CA)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the French Canadian voice supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>boire</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>madame</td>
<td>t</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>femme</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>grand</td>
<td>k</td>
</tr>
<tr>
<td>ɥ</td>
<td>H</td>
<td>labial-palatal approximant</td>
<td>bruit</td>
<td>u</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>meilleur</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>quatre</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>malade</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>maison</td>
<td>p</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>--------</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>astronomome</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>J</td>
<td>palatal nasal</td>
<td>baigner</td>
<td>J</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>parking</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pomme</td>
<td>p</td>
</tr>
<tr>
<td>w</td>
<td>R</td>
<td>voiced uvular fricative</td>
<td>amoureux</td>
<td>k</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>santé</td>
<td>s</td>
</tr>
<tr>
<td>j</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>chat</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>téléphone</td>
<td>t</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vrai</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>soir</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>raison</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>aubergine</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

| ø   | 2       | close-mid front rounded vowel       | deux          | o      |
| œ   | 9       | open-mid front rounded vowel        | neuf          | O      |
| œ˜  | 9~      | nasal open-mid front rounded vowel  | brun          | O      |
| a   | @       | mid central vowel                   | je            | @      |
| a   | a       | open front unrounded vowel          | table         | a      |
| ö   | A~      | nasal open back unrounded vowel     | camembert     | a      |
| e   | e       | close-mid front unrounded vowel     | marché         | e      |
| e   | E       | open-mid front unrounded vowel      | neige         | E      |
### German (de-DE)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the German voices that are supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ĕ</td>
<td>E~</td>
<td>nasal open-mid front unrounded vowel</td>
<td>sapin</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>mille</td>
<td>i</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>hôpital</td>
<td>o</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>homme</td>
<td>O</td>
</tr>
<tr>
<td>õ</td>
<td>O~</td>
<td>nasal open-mid back rounded vowel</td>
<td>bon</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>sous</td>
<td>u</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>close front rounded vowel</td>
<td>dur</td>
<td>u</td>
</tr>
</tbody>
</table>

#### Additional Symbols

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>`'</td>
<td>&quot;</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>','</td>
<td>%</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>'.'</td>
<td></td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

### IPA and X-SAMPA Tables for Supported Languages

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʔ</td>
<td>?</td>
<td>glottal stop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>Bier</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>Dach</td>
<td>t</td>
</tr>
<tr>
<td>ç</td>
<td>C</td>
<td>voiceless palatal fricative</td>
<td>ich</td>
<td>k</td>
</tr>
<tr>
<td>̕d3</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>Dschungel</td>
<td>S</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>Voiceless labiodental fricative</td>
<td>Vogel</td>
<td>f</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>----------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>Voiced velar plosive</td>
<td>Gabel</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>Voiceless glottal fricative</td>
<td>Haus</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>Voiceless glottal fricative</td>
<td>jemand</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>Voiceless velar plosive</td>
<td>Kleid</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>Alveolar lateral approximant</td>
<td>Loch</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>Bilabial nasal</td>
<td>Milch</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>Alveolar nasal</td>
<td>Natur</td>
<td>t</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>Velar nasal</td>
<td>klingen</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>Voiceless bilabial plosive</td>
<td>Park</td>
<td>p</td>
</tr>
<tr>
<td>ŋpf</td>
<td>pf</td>
<td>Voiceless labiodental affricate</td>
<td>Apfel</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>R</td>
<td>Uvular trill</td>
<td>Regen</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>Voiceless alveolar fricative</td>
<td>Messer</td>
<td>s</td>
</tr>
<tr>
<td>f</td>
<td>s</td>
<td>Voiceless postalveolar fricative</td>
<td>Fischer</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>Voiceless alveolar plosive</td>
<td>Topf</td>
<td>T</td>
</tr>
<tr>
<td>ts</td>
<td>Ts</td>
<td>Voiceless alveolar affricate</td>
<td>Zahl</td>
<td></td>
</tr>
<tr>
<td>ŋf</td>
<td>tS</td>
<td>Voiceless postalveolar affricate</td>
<td>deutsch</td>
<td>S</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>Voiced labiodental fricative</td>
<td>Wasser</td>
<td>f</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>Voiceless velar fricative</td>
<td>kochen</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>Voiced alveolar fricative</td>
<td>See</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>Voiced postalveolar fricative</td>
<td>Orange</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| ø: | 2: | long close-mid front rounded vowel | böse | o |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>6</td>
<td>near-open central vowel</td>
<td>besser</td>
<td>a</td>
</tr>
<tr>
<td>øe</td>
<td>6_(^\wedge)</td>
<td>non-syllabic near-open central vowel</td>
<td>Klar</td>
<td>a</td>
</tr>
<tr>
<td>œe</td>
<td>9</td>
<td>open-mid front rounded vowel</td>
<td>können</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>@</td>
<td>mid central vowel</td>
<td>Rede</td>
<td>@</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>Salz</td>
<td>a</td>
</tr>
<tr>
<td>a:</td>
<td>a:</td>
<td>long open front unrounded vowel</td>
<td>Sahne</td>
<td>a</td>
</tr>
<tr>
<td>aɪ</td>
<td>aɪ</td>
<td>diphthong</td>
<td>nein</td>
<td>a</td>
</tr>
<tr>
<td>aʊ</td>
<td>aU</td>
<td>diphthong</td>
<td>Augen</td>
<td>a</td>
</tr>
<tr>
<td>ȏ</td>
<td>A~</td>
<td>nasal open back unrounded vowel</td>
<td>Restaurant</td>
<td>a</td>
</tr>
<tr>
<td>e:</td>
<td>e:</td>
<td>long close-mid front unrounded vowel</td>
<td>Rede</td>
<td>e</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>Keller</td>
<td>E</td>
</tr>
<tr>
<td>ɛ̃</td>
<td>E~</td>
<td>nasal open-mid front unrounded vowel</td>
<td>Terrain</td>
<td>E</td>
</tr>
<tr>
<td>i:</td>
<td>i:</td>
<td>long close front unrounded vowel</td>
<td>Lied</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>ɪ</td>
<td>near-close near-front unrounded vowel</td>
<td>bitte</td>
<td>i</td>
</tr>
<tr>
<td>o:</td>
<td>o:</td>
<td>long close-mid back rounded vowel</td>
<td>Kohl</td>
<td>o</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>Koffer</td>
<td>O</td>
</tr>
<tr>
<td>ɔ̃</td>
<td>O~</td>
<td>nasal open-mid back rounded vowel</td>
<td>Announce</td>
<td>O</td>
</tr>
<tr>
<td>ɔʏ</td>
<td>OY</td>
<td>diphthong</td>
<td>neu</td>
<td>O</td>
</tr>
<tr>
<td>u:</td>
<td>u:</td>
<td>long close back rounded vowel</td>
<td>Bruder</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>Wunder</td>
<td>u</td>
</tr>
<tr>
<td>y:</td>
<td>y:</td>
<td>long close front rounded vowel</td>
<td>kühl</td>
<td>u</td>
</tr>
</tbody>
</table>
Hindi (hi-IN)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the phoneme's sound type for the Hindi voices that are supported by Amazon Polly.

For additional phonemes used in conjunction with Hindi, see English, Indian (en-IN) (p. 37).

### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Consonants</strong></td>
<td></td>
</tr>
<tr>
<td>pʰ</td>
<td>p_h</td>
<td>voiceless aspirated bilabial plosive</td>
<td>फूल (phool)</td>
</tr>
<tr>
<td>bʰ</td>
<td>b_h</td>
<td>voiced aspirated bilabial plosive</td>
<td>भारी (bhaari)</td>
</tr>
<tr>
<td>t</td>
<td>t_d</td>
<td>voiceless dental plosive</td>
<td>तापमान (taapmaan)</td>
</tr>
<tr>
<td>tʰ</td>
<td>t_d_h</td>
<td>voiceless aspirated dental plosive</td>
<td>थोड़ा (thoda)</td>
</tr>
<tr>
<td>d</td>
<td>d_d</td>
<td>voiced dental plosive</td>
<td>दिल्ली (dilli)</td>
</tr>
<tr>
<td>dʰ</td>
<td>d_d_h</td>
<td>voiced aspirated dental plosive</td>
<td>धोबी (dhobi)</td>
</tr>
<tr>
<td>t`</td>
<td>t`_h</td>
<td>voiceless aspirated retroflex plosive</td>
<td>कटोरा (katora)</td>
</tr>
<tr>
<td>tʰ</td>
<td>t`_h</td>
<td>voiceless aspirated retroflex plosive</td>
<td>ठंड (thand)</td>
</tr>
<tr>
<td>d`</td>
<td>d`_h</td>
<td>voiced aspirated retroflex plosive</td>
<td>ढर (darr)</td>
</tr>
<tr>
<td>dʰ</td>
<td>d`_h</td>
<td>voiced aspirated retroflex plosive</td>
<td>ढाल (dhal)</td>
</tr>
<tr>
<td>tʃʰ</td>
<td>tS_h</td>
<td>voiceless aspirated palatal affricate</td>
<td>छाल (chaal)</td>
</tr>
<tr>
<td>dʒʰ</td>
<td>dZ_h</td>
<td>voiced aspirated palatal affricate</td>
<td>ज्याल (jhaal)</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>kʰ</td>
<td>k_h</td>
<td>voiceless aspirated velar plosive</td>
<td>खान (khan)</td>
</tr>
<tr>
<td>gʰ</td>
<td>g_h</td>
<td>voiced aspirated velar plosive</td>
<td>घान (ghaan)</td>
</tr>
<tr>
<td>n`</td>
<td>n`</td>
<td>retroflex nasal</td>
<td>क्षण (kshan)</td>
</tr>
<tr>
<td>r</td>
<td>4</td>
<td>alveolar flap</td>
<td>राम (ram)</td>
</tr>
<tr>
<td>r`</td>
<td>r`</td>
<td>plain retroflex flap</td>
<td>बड़ा (bada)</td>
</tr>
<tr>
<td>r`_h</td>
<td>r`_h</td>
<td>voiced aspirated retroflex flap</td>
<td>बढ़ी (barhi)</td>
</tr>
<tr>
<td>v\</td>
<td>v\</td>
<td>bilabial approximant</td>
<td>चमुल (wasool)</td>
</tr>
</tbody>
</table>

**Vowels**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>@_o</td>
<td>mid central vowel</td>
<td>अच्छा (achhaa)</td>
</tr>
<tr>
<td>ŋ</td>
<td>@~</td>
<td>nasalised mid central vowel</td>
<td>हँसना (hansnaa)</td>
</tr>
<tr>
<td>a</td>
<td>A_o</td>
<td>open front unrounded vowel</td>
<td>आग (aag)</td>
</tr>
<tr>
<td>a~</td>
<td>A~</td>
<td>nasalised open front unrounded vowel</td>
<td>खोंसिया (ghariyaan)</td>
</tr>
<tr>
<td>i</td>
<td>I_o</td>
<td>near-close near-front unrounded vowel</td>
<td>इक्कीस (ikkees)</td>
</tr>
<tr>
<td>ĭ</td>
<td>I~</td>
<td>nasalised near-close near front unrounded vowel</td>
<td>संचाई (sinchai)</td>
</tr>
<tr>
<td>i</td>
<td>i_o</td>
<td>close front unrounded vowel</td>
<td>बिल्ली (billee)</td>
</tr>
<tr>
<td>ĭ</td>
<td>i~</td>
<td>nasalised close front unrounded vowel</td>
<td>नहीं (nahin)</td>
</tr>
<tr>
<td>u</td>
<td>U_o</td>
<td>near-close near-back rounded vowel</td>
<td>उल्लूल (ullu)</td>
</tr>
<tr>
<td>ũ</td>
<td>U~</td>
<td>nasalised near-close near-back rounded vowel</td>
<td>मँह (munh)</td>
</tr>
<tr>
<td>u</td>
<td>u_o</td>
<td>close back rounded vowel</td>
<td>फूल (phool)</td>
</tr>
<tr>
<td>ũ</td>
<td>u~</td>
<td>nasalised close back rounded vowel</td>
<td>ऊंट (oont)</td>
</tr>
<tr>
<td>o</td>
<td>O_o</td>
<td>open-mid back rounded vowel</td>
<td>कौन (kaun)</td>
</tr>
<tr>
<td>ō</td>
<td>O~</td>
<td>nasalised open-mid back rounded vowel</td>
<td>भी (bhaun)</td>
</tr>
</tbody>
</table>
**IPA** | **X-SAMPA** | **Description** | **Example** |
--- | --- | --- | --- |
{o} | o | close-mid back rounded vowel | सोना (sona) |
{o} | o~ | nasalised close-mid back rounded vowel | क्यों (kyon) |
{ε} | E_o | open-mid front unrounded vowel | पैसा (paisa) |
{ε} | E~ | nasalised open-mid front unrounded vowel | मैं (main) |
{e} | e | close-mid front unrounded vowel | एक (ek) |
{e} | e~ | nasalised close-mid front unrounded vowel | किताबें (kitabein) |

### Icelandic (is-IS)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Icelandic voices that are supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>grasbakkunum</td>
<td>0</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>voiceless palatal plosive</td>
<td>pakkun</td>
<td>k</td>
</tr>
<tr>
<td>cʰ</td>
<td>C_h</td>
<td>aspirated voiceless palatal plosive</td>
<td>anarkistai</td>
<td>k</td>
</tr>
<tr>
<td>ç</td>
<td>C</td>
<td>voiceless palatal fricative</td>
<td>héðan</td>
<td>k</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>bóndi</td>
<td>t</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>borð</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>duft</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>holgóma</td>
<td>k</td>
</tr>
<tr>
<td>γ</td>
<td>G</td>
<td>voiced velar fricative</td>
<td>hugur</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>héili</td>
<td>k</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>---------------------------------</td>
<td>---------------</td>
<td>--------</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>jökull</td>
<td>i</td>
</tr>
<tr>
<td>kʰ</td>
<td>k_h</td>
<td>aspirated voiceless velar plosive</td>
<td>ósköpunum</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>gölf</td>
<td>t</td>
</tr>
<tr>
<td>ħ</td>
<td>l_0</td>
<td>voiceless alveolar lateral approximant</td>
<td>fólk</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>september</td>
<td>p</td>
</tr>
<tr>
<td>ñ</td>
<td>m_0</td>
<td>voiceless bilabial nasal</td>
<td>kompa</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>númer</td>
<td>t</td>
</tr>
<tr>
<td>ñ</td>
<td>n_0</td>
<td>voiceless alveolar nasal</td>
<td>pöntun</td>
<td>t</td>
</tr>
<tr>
<td>ñ</td>
<td>J</td>
<td>palatal nasal</td>
<td>pælingar</td>
<td>J</td>
</tr>
<tr>
<td>ñ</td>
<td>N</td>
<td>velar nasal</td>
<td>söngvarann</td>
<td>k</td>
</tr>
<tr>
<td>ñ</td>
<td>N_0</td>
<td>voiceless velar nasal</td>
<td>frænka</td>
<td>k</td>
</tr>
<tr>
<td>pʰ</td>
<td>p_h</td>
<td>aspirated voiceless bilabial plosive</td>
<td>afplánun</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>afskrifta</td>
<td>r</td>
</tr>
<tr>
<td>r</td>
<td>r_0</td>
<td>voiceless alveolar trill</td>
<td>andvörpum</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>baðhús</td>
<td>s</td>
</tr>
<tr>
<td>tʰ</td>
<td>t_h</td>
<td>aspirated voiceless alveolar plosive</td>
<td>tanki</td>
<td>t</td>
</tr>
<tr>
<td>θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>þeldökki</td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>silfur</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td></td>
<td>u</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>samfélags</td>
<td>k</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| ð   | ð     | open-mid front rounded vowel   | þrðskuldinum  | O      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>œː</td>
<td>9:</td>
<td>long open-mid front rounded vowel</td>
<td>tvö</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>nefna</td>
<td>a</td>
</tr>
<tr>
<td>a:</td>
<td>a:</td>
<td>long open front unrounded vowel</td>
<td>fara</td>
<td>a</td>
</tr>
<tr>
<td>au</td>
<td>au</td>
<td>diphthong</td>
<td>átta</td>
<td>a</td>
</tr>
<tr>
<td>au:</td>
<td>au:</td>
<td>diphthong</td>
<td>átján</td>
<td>a</td>
</tr>
<tr>
<td>ε</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>kennari</td>
<td>E</td>
</tr>
<tr>
<td>εː</td>
<td>E:</td>
<td>long open-mid front unrounded vowel</td>
<td>dreka</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>Gúlfver</td>
<td>i</td>
</tr>
<tr>
<td>iː</td>
<td>iː</td>
<td>long close front unrounded vowel</td>
<td>þrín</td>
<td>i</td>
</tr>
<tr>
<td>ð</td>
<td>l</td>
<td>near-close near-front unrounded vowel</td>
<td>samspíl</td>
<td>i</td>
</tr>
<tr>
<td>ðː</td>
<td>ðː</td>
<td>long near-close near-front unrounded vowel</td>
<td>stíg</td>
<td>i</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>regnðrópar</td>
<td>O</td>
</tr>
<tr>
<td>ɔː</td>
<td>O:</td>
<td>long open-mid back rounded vowel</td>
<td>ullarðólar</td>
<td>O</td>
</tr>
<tr>
<td>ðu</td>
<td>ðu</td>
<td>diphthong</td>
<td>tólf</td>
<td>O</td>
</tr>
<tr>
<td>ðu:</td>
<td>ðu:</td>
<td>diphthong</td>
<td>fjórir</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>stúlkan</td>
<td>u</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close back rounded vowel</td>
<td>frú</td>
<td>u</td>
</tr>
<tr>
<td>y</td>
<td>Y</td>
<td>near-close near-front rounded vowel</td>
<td>tíu</td>
<td>u</td>
</tr>
<tr>
<td>yː</td>
<td>Y</td>
<td>long near-close near-front rounded vowel</td>
<td>gruninn</td>
<td>u</td>
</tr>
</tbody>
</table>

**Additional Symbols**

| ˈ | " | primary stress | Alabama |
Italian (it-IT)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Italian voices that are supported by Amazon Polly.

### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>%</td>
<td>secondary stress</td>
<td>Alabama</td>
<td>p</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
<td></td>
</tr>
</tbody>
</table>

Consonants

<p>| b   | b       | voiced bilabial plosive   | bacca   | p      |
| d   | d       | voiced alveolar plosive   | dama    | t      |
| dz  | dz      | voiced alveolar affricate | zero    | s      |
| dʒ  | dZ      | voiced postalveolar affricate | giro   | S      |
| f   | f       | voiceless labiodental fricative | famiglia | f      |
| g   | g       | voiced velar plosive      | gatto   | k      |
| h   | h       | voiceless glottal fricative | horror  | k      |
| j   | j       | palatal approximant       | dieci   | i      |
| k   | k       | voiceless velar plosive   | campo   | k      |
| l   | l       | alveolar lateral approximant | lido   | t      |
| A   | L       | palatal lateral approximant | aglio  | J      |
| m   | m       | bilabial nasal            | mille   | p      |
| n   | n       | alveolar nasal            | nove    | t      |
| p   | p       | voiceless bilabial plosive | pizza   | p      |
| r   | r       | alveolar trill            | risata  | r      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>sei</td>
<td>s</td>
</tr>
<tr>
<td>f</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>scienza</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tavola</td>
<td>t</td>
</tr>
<tr>
<td>ts</td>
<td>ts</td>
<td>voiceless alveolar affricate</td>
<td>forza</td>
<td>s</td>
</tr>
<tr>
<td>tʃ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>cielo</td>
<td>S</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>venti</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>quattro</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>bisogno</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>bijou</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

| a   | a       | open front unrounded vowel | arco     | a      |
| e   | e       | close-mid front unrounded vowel | tre     | e      |
| ɛ   | E       | open-mid front unrounded vowel | ettaro  | E      |
| i   | i       | close front unrounded vowel | impero   | i      |
| o   | o       | close-mid back rounded vowel | cento   | o      |
| ɔ   | O       | open-mid back rounded vowel | otto    | O      |
| u   | u       | close back rounded vowel    | uno     | u      |

**Additional Symbols**

| | | primary stress | Alabama |
| ' | " |  | |
| , | % | secondary stress | Alabama |
| . | . | syllable boundary | A.la.ba.ma |
## Japanese (ja-JP)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Japanese voice supported by Amazon Polly.

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>4</td>
<td>alveolar flap</td>
<td>練習, renshuu</td>
<td>t</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>glottal stop</td>
<td>あつっ, atsu'</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>舞踊, buyou</td>
<td>p</td>
</tr>
<tr>
<td>β</td>
<td>B</td>
<td>voiced bilabial fricative</td>
<td>ヴィンテージ, vinteeji</td>
<td>B</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>voiceless palatal plosive</td>
<td>ききょう, kikyou</td>
<td>k</td>
</tr>
<tr>
<td>ç</td>
<td>C</td>
<td>voiceless palatal fricative</td>
<td>人, hito</td>
<td>k</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>濁点, dakuten</td>
<td>t</td>
</tr>
<tr>
<td>d͡ʑ</td>
<td>dz\</td>
<td>voiced alveolo-palatal affricate</td>
<td>純, jun</td>
<td>J</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>ご飯, gohan</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>本, hon</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>屋根, yane</td>
<td>i</td>
</tr>
<tr>
<td>j\</td>
<td>J</td>
<td>voiced palatal plosive</td>
<td>行儀, gyougi</td>
<td>J</td>
</tr>
<tr>
<td>k</td>
<td>k\</td>
<td>voiceless velar plosive</td>
<td>漢字, kanji</td>
<td>k</td>
</tr>
<tr>
<td>J\</td>
<td>Ｌ\</td>
<td>alveolar lateral flap</td>
<td>釣り, tsuri</td>
<td>r</td>
</tr>
<tr>
<td>Jj</td>
<td>\j</td>
<td>alveolar lateral flap, palatal approximant</td>
<td>流行, ryuukou</td>
<td>r</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>飯, meshi</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>猫, neko</td>
<td>t</td>
</tr>
<tr>
<td>n\</td>
<td>N\</td>
<td>palatal nasal</td>
<td>日本, nippon</td>
<td>J</td>
</tr>
<tr>
<td>n\</td>
<td>N\</td>
<td>uvular nasal</td>
<td>缶, kan</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p\</td>
<td>voiceless bilabial plosive</td>
<td>パン, pan</td>
<td>p</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------------</td>
<td>--------------</td>
<td>--------</td>
</tr>
<tr>
<td>ɸ</td>
<td>p\</td>
<td>voiceless bilabial fricative</td>
<td>福, huku</td>
<td>f</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>層, sou</td>
<td>s</td>
</tr>
<tr>
<td>ş</td>
<td>s\</td>
<td>voiceless alveolo-palatal fricative</td>
<td>書簡, shokan</td>
<td>J</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>手紙, tegami</td>
<td>t</td>
</tr>
<tr>
<td>ts</td>
<td>ts</td>
<td>voiceless alveolar affricate</td>
<td>釣り, tsuri</td>
<td>s</td>
</tr>
<tr>
<td>to</td>
<td>ts\</td>
<td>voiceless alveolo-palatal affricate</td>
<td>吉, kichi</td>
<td>J</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>電話, denwa</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>座敷, zashiki</td>
<td>s</td>
</tr>
</tbody>
</table>

**Vowels**

| äː | a:~   | long open central unrounded vowel | 羽蟻, haari | a |
| ä  | a~    | open central unrounded vowel      | 仮名, kana  | a |
| eː | e:o   | long mid front unrounded vowel    | 学生, gakusei | @ |
| e  | e_o   | mid front unrounded vowel         | 歴, reki    | @ |
| i  | i      | close front unrounded vowel       | 気, ki      | i |
| iː | i:    | long close front unrounded vowel  | 詩歌, shiika | i |
| u  | M      | close back unrounded vowel        | 運, un      | i |
| uː | M:    | long close back unrounded vowel   | 宗教, shuukyou | i |
| oː | o:o   | long mid back rounded vowel       | 購読, koodoku | o |
| o  | o_o   | mid back rounded vowel            | 読者, dokusha | o |
# Korean (ko-KR)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Korean voice supported by Amazon Polly.

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>강, [g]ang</td>
<td>k</td>
</tr>
<tr>
<td>k#</td>
<td>k_t</td>
<td>strong voiceless velar plosive</td>
<td>깨, [kk]e</td>
<td>k</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>남, [n]am</td>
<td>t</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>도, [d]o</td>
<td>t</td>
</tr>
<tr>
<td>t#</td>
<td>t_t</td>
<td>strong voiceless alveolar plosive</td>
<td>때, [tt]e</td>
<td>t</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar flap</td>
<td>사랑, sa[rr]ang</td>
<td>t</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>돌, do[l]</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>무, [m]u</td>
<td>p</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>봄, [b]om</td>
<td>p</td>
</tr>
<tr>
<td>p#</td>
<td>p_t</td>
<td>strong voiceless bilabial plosive</td>
<td>빼, [pp]eol</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>세, [s]e</td>
<td>s</td>
</tr>
<tr>
<td>s#</td>
<td>s_t</td>
<td>strong voiceless alveolar fricative</td>
<td>씨, [ss]i</td>
<td>s</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>방, ba[ng]</td>
<td>k</td>
</tr>
<tr>
<td>ɾ</td>
<td>4</td>
<td>alveolar flap</td>
<td>사랑, sa[r]ang</td>
<td>t</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>돌, do[l]</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>무, [m]u</td>
<td>p</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>봄, [b]om</td>
<td>p</td>
</tr>
<tr>
<td>p#</td>
<td>p_t</td>
<td>strong voiceless bilabial plosive</td>
<td>빼, [pp]eol</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>세, [s]e</td>
<td>s</td>
</tr>
<tr>
<td>s#</td>
<td>s_t</td>
<td>strong voiceless alveolar fricative</td>
<td>씨, [ss]i</td>
<td>s</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>방, ba[ng]</td>
<td>k</td>
</tr>
<tr>
<td>ɾ</td>
<td>4</td>
<td>alveolar flap</td>
<td>사랑, sa[r]ang</td>
<td>t</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>돌, do[l]</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>무, [m]u</td>
<td>p</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>봄, [b]om</td>
<td>p</td>
</tr>
<tr>
<td>p#</td>
<td>p_t</td>
<td>strong voiceless bilabial plosive</td>
<td>빼, [pp]eol</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>세, [s]e</td>
<td>s</td>
</tr>
<tr>
<td>s#</td>
<td>s_t</td>
<td>strong voiceless alveolar fricative</td>
<td>씨, [ss]i</td>
<td>s</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>방, ba[ng]</td>
<td>k</td>
</tr>
<tr>
<td>ɾ</td>
<td>4</td>
<td>alveolar flap</td>
<td>사랑, sa[r]ang</td>
<td>t</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>돌, do[l]</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>무, [m]u</td>
<td>p</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>봄, [b]om</td>
<td>p</td>
</tr>
<tr>
<td>p#</td>
<td>p_t</td>
<td>strong voiceless bilabial plosive</td>
<td>빼, [pp]eol</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>세, [s]e</td>
<td>s</td>
</tr>
<tr>
<td>s#</td>
<td>s_t</td>
<td>strong voiceless alveolar fricative</td>
<td>씨, [ss]i</td>
<td>s</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>방, ba[ng]</td>
<td>k</td>
</tr>
<tr>
<td>ɾ</td>
<td>4</td>
<td>alveolar flap</td>
<td>사랑, sa[r]ang</td>
<td>t</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>돌, do[l]</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>무, [m]u</td>
<td>p</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>봄, [b]om</td>
<td>p</td>
</tr>
</tbody>
</table>
IPA | X-SAMPA | Description | Example | Viseme
---|---|---|---|---
tʰ | t_h | aspirated voiceless alveolar plosive | 통, [t]ong | t
pʰ | p_h | aspirated voiceless bilabial plosive | 패, [p]e | p
h | h | voiceless glottal fricative | 헌, [h]im | k
j | j | palatal approximant | 양, [y]ang | i
w | w | labial-velar approximant | 왕, [w]ang | u
ɰ | M\ | velar approximant> | 의, [w]ji | i

Vowels

a | a | open front unrounded vowel | 밥, b[a]b | a
ʌ | V | open-mid back unrounded vowel | 정, j[eo]ng | E
ɛ | E | open-mid front unrounded vowel | 배, b[e] | E
ɔ | o | close-mid back rounded vowel | 노, n[o] | o
u | u | close back rounded vowel | 들, d[u]l | u
ʊ | M | close back unrounded vowel | 은, [eu]n | i
i | i | close front unrounded vowel | 김, k[i]m | i

Norwegian (nb-NO)

The following chart lists the full set of International Phonetic Alphabet (IPA) phonemes and the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols as well as the corresponding visemes as supported by Amazon Polly for Norwegian language voices.

| IPA | X-SAMPA | Description | Example | Viseme
---|---|---|---|---
ɾ | ɾ | alveolar flap | prøv | t
b | b | voiced bilabial plosive | labb | p
ζ | C | voiceless palatal fricative | kino | k
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>ladd</td>
<td>t</td>
</tr>
<tr>
<td>d'</td>
<td>d`</td>
<td>voiced retroflex plosive</td>
<td>verdi</td>
<td>t</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fot</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>tagg</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>ha</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>gi</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>takk</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>fall, ball</td>
<td>t</td>
</tr>
<tr>
<td>l`</td>
<td>l`</td>
<td>retroflex lateral approximant</td>
<td>ærlig</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>lam</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>vann</td>
<td>t</td>
</tr>
<tr>
<td>n`</td>
<td>n`</td>
<td>retroflex nasal</td>
<td>garn</td>
<td>t</td>
</tr>
<tr>
<td>ñ</td>
<td>ñ</td>
<td>velar nasal</td>
<td>sang</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>hopp</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>lass</td>
<td>s</td>
</tr>
<tr>
<td>ñ</td>
<td>ñ`</td>
<td>voiceless retroflex fricative</td>
<td>års</td>
<td>S</td>
</tr>
<tr>
<td>f</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>skyt</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>lat</td>
<td>t</td>
</tr>
<tr>
<td>t`</td>
<td>t`</td>
<td>voiceless retroflex plosive</td>
<td>hardt</td>
<td>t</td>
</tr>
<tr>
<td>u</td>
<td>v\</td>
<td>labiodental approximant</td>
<td>vin</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>will</td>
<td>x</td>
</tr>
</tbody>
</table>

**Vowels**

<table>
<thead>
<tr>
<th>IPA</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Amazon Polly Developer Guide**

**Phoneme and Viseme Tables for Supported Languages**

---

67
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>øː</td>
<td>2ː</td>
<td>long close-mid front rounded vowel</td>
<td>søt</td>
<td>o</td>
</tr>
<tr>
<td>œ</td>
<td>9</td>
<td>open-mid front rounded vowel</td>
<td>søtt</td>
<td>O</td>
</tr>
<tr>
<td>æ</td>
<td>@</td>
<td>mid central vowel</td>
<td>ape</td>
<td>@</td>
</tr>
<tr>
<td>æː</td>
<td>{:</td>
<td>long near-open front unrounded vowel</td>
<td>vær</td>
<td>a</td>
</tr>
<tr>
<td>u</td>
<td>}</td>
<td>close central rounded vowel</td>
<td>lund</td>
<td>u</td>
</tr>
<tr>
<td>uː</td>
<td>}ː</td>
<td>long close central rounded vowel</td>
<td>lun</td>
<td>u</td>
</tr>
<tr>
<td>æː</td>
<td>{</td>
<td>near-open front unrounded vowel</td>
<td>vært</td>
<td>a</td>
</tr>
<tr>
<td>øː</td>
<td>A</td>
<td>open back unrounded vowel</td>
<td>hatt</td>
<td>a</td>
</tr>
<tr>
<td>øː</td>
<td>Aː</td>
<td>long open back unrounded vowel</td>
<td>hat</td>
<td>a</td>
</tr>
<tr>
<td>eː</td>
<td>eː</td>
<td>long close-mid front unrounded vowel</td>
<td>sen</td>
<td>e</td>
</tr>
<tr>
<td>e</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>send</td>
<td>E</td>
</tr>
<tr>
<td>iː</td>
<td>iː</td>
<td>long close front unrounded vowel</td>
<td>vin</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>ɪ</td>
<td>near-close near-front unrounded vowel</td>
<td>vind</td>
<td>i</td>
</tr>
<tr>
<td>oː</td>
<td>oː</td>
<td>long close-mid back rounded vowel</td>
<td>våt</td>
<td>o</td>
</tr>
<tr>
<td>ó</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>vått</td>
<td>O</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close back rounded vowel</td>
<td>bok</td>
<td>u</td>
</tr>
<tr>
<td>uː</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>bukk</td>
<td>u</td>
</tr>
<tr>
<td>yː</td>
<td>yː</td>
<td>long close front rounded vowel</td>
<td>lyn</td>
<td>u</td>
</tr>
<tr>
<td>y</td>
<td>Y</td>
<td>near-close near-front rounded vowel</td>
<td>lynne</td>
<td>u</td>
</tr>
</tbody>
</table>

**Additional Symbols**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>′</td>
<td>primary stress</td>
</tr>
</tbody>
</table>
The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Polish voices that are supported by Amazon Polly.

### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>%</td>
<td>secondary stress</td>
<td>Alabama</td>
<td>p</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
<td>p</td>
</tr>
</tbody>
</table>

**Polish (pl-PL)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Polish voices that are supported by Amazon Polly.
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>sum, pas</td>
<td>s</td>
</tr>
<tr>
<td>ç</td>
<td>s\</td>
<td>voiceless alveolo-palatal fricative</td>
<td>śruba, śnieg</td>
<td>J</td>
</tr>
<tr>
<td>ş</td>
<td>s `</td>
<td>voiceless retroflex fricative</td>
<td>szum, masz</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tok, stół</td>
<td>t</td>
</tr>
<tr>
<td>ĭs</td>
<td>ts</td>
<td>voiceless alveolar affricate</td>
<td>car, co</td>
<td>s</td>
</tr>
<tr>
<td>ğq</td>
<td>ts\</td>
<td>voiceless alveolo-palatal affricate</td>
<td>ćma, mieć</td>
<td>J</td>
</tr>
<tr>
<td>įs</td>
<td>ts `</td>
<td>voiceless retroflex affricate</td>
<td>czas, raczej</td>
<td>S</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>worek, mewa</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>łąska, mało</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zero</td>
<td>s</td>
</tr>
<tr>
<td>ʑ</td>
<td>z\</td>
<td>voiced alveolo-palatal fricative</td>
<td>źrebię, bieliźnie</td>
<td>J</td>
</tr>
<tr>
<td>ʐ</td>
<td>z `</td>
<td>voiced retroflex fricative</td>
<td>żar, żona</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<table>
<thead>
<tr>
<th>IPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>open front unrounded vowel</td>
<td>ja</td>
<td>a</td>
</tr>
<tr>
<td>ɛ</td>
<td>open-mid front unrounded vowel</td>
<td>echo</td>
<td>E</td>
</tr>
<tr>
<td>ɛ̃</td>
<td>nasal open-mid front unrounded vowel</td>
<td>węże</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>close front unrounded vowel</td>
<td>ile</td>
<td>i</td>
</tr>
<tr>
<td>ɔ</td>
<td>open-mid back rounded vowel</td>
<td>oczy</td>
<td>O</td>
</tr>
<tr>
<td>ɔ̃</td>
<td>nasal open-mid back rounded vowel</td>
<td>wąź</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>close back rounded vowel</td>
<td>uczta</td>
<td>u</td>
</tr>
</tbody>
</table>
### Additional Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>primary stress</td>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>secondary stress</td>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
<td></td>
</tr>
</tbody>
</table>

### Portuguese (pt-PT)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Portuguese voices that are supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example 1</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>ɨ</td>
<td>close central unrounded vowel</td>
<td>byk</td>
<td>i</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consonants</th>
</tr>
</thead>
<tbody>
<tr>
<td>r 4</td>
</tr>
<tr>
<td>b b</td>
</tr>
<tr>
<td>d d</td>
</tr>
<tr>
<td>f f</td>
</tr>
<tr>
<td>g g</td>
</tr>
<tr>
<td>j j</td>
</tr>
<tr>
<td>k k</td>
</tr>
<tr>
<td>l l</td>
</tr>
<tr>
<td>ʎ L</td>
</tr>
<tr>
<td>m m</td>
</tr>
<tr>
<td>n n</td>
</tr>
<tr>
<td>p J</td>
</tr>
<tr>
<td>p p</td>
</tr>
<tr>
<td>R R\</td>
</tr>
<tr>
<td>IPA</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
</tr>
<tr>
<td>t</td>
</tr>
<tr>
<td>v</td>
</tr>
<tr>
<td>w</td>
</tr>
<tr>
<td>z</td>
</tr>
<tr>
<td>ʒ</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| a   | a       | open front unrounded vowel          | parto   | a      |
| a~  | a~      | nasal open front unrounded vowel    | pega    | a      |
| e   | e       | close-mid front unrounded vowel     | pega    | e      |
| e~  | e~      | nasal close-mid front unrounded vowel| movem  | e      |
| E   | E       | open-mid front unrounded vowel      | café    | E      |
| i   | i       | close front unrounded vowel         | lingueta| i      |
| i~  | i~      | nasal close front unrounded vowel   | cinto   | i      |
| o   | o       | close-mid back rounded vowel        | poder   | o      |
| o~  | o~      | nasal close-mid back rounded vowel  | compra  | o      |
| O   | O       | open-mid back rounded vowel         | cotó    | O      |
| u   | u       | close back rounded vowel            | fui     | u      |
| u~  | u~      | nasal close back rounded vowel      | sunto   | u      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional Symbols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'</td>
<td>&quot;</td>
<td>primary stress</td>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>%</td>
<td>secondary stress</td>
<td>Alabama</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
<td></td>
</tr>
</tbody>
</table>

**Portuguese, Brazilian (pt-BR)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Brazilian Portuguese voices that are supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>4</td>
<td>alveolar flap</td>
<td>pira</td>
<td>t</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bato</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dato</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>idade</td>
<td>S</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>facto</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>gato</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>paraguay</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cacto</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>galo</td>
<td>t</td>
</tr>
<tr>
<td>ʎ</td>
<td>L</td>
<td>palatal lateral approximant</td>
<td>galho</td>
<td>J</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mato</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nato</td>
<td>t</td>
</tr>
<tr>
<td>ŋ</td>
<td>J</td>
<td>palatal nasal</td>
<td>pinha</td>
<td>J</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pato</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>saca</td>
<td>s</td>
</tr>
</tbody>
</table>
### IPA and X-SAMPA Phoneme and Viseme Tables

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>chato</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tacto</td>
<td>t</td>
</tr>
<tr>
<td>ɬ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>noite</td>
<td>S</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vaca</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>mau</td>
<td>u</td>
</tr>
<tr>
<td>χ</td>
<td>X</td>
<td>voiceless uvular fricative</td>
<td>carro</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>zaca</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>jacto</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>parto</td>
<td>a</td>
</tr>
<tr>
<td>a~</td>
<td>a~</td>
<td>nasal open front unrounded vowel</td>
<td>pensamos</td>
<td>a</td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>close-mid front unrounded vowel</td>
<td>pega</td>
<td>e</td>
</tr>
<tr>
<td>e~</td>
<td>e~</td>
<td>nasal close-mid front unrounded vowel</td>
<td>movem</td>
<td>e</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>café</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>lingueta</td>
<td>i</td>
</tr>
<tr>
<td>i~</td>
<td>i~</td>
<td>nasal close front unrounded vowel</td>
<td>cinto</td>
<td>i</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>poder</td>
<td>o</td>
</tr>
<tr>
<td>o~</td>
<td>o~</td>
<td>nasal close-mid back rounded vowel</td>
<td>compra</td>
<td>o</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>cotó</td>
<td>O</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>fui</td>
<td>u</td>
</tr>
</tbody>
</table>
Phoneme and Viseme Tables for Supported Languages

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>u~</td>
<td>u~</td>
<td>nasal close back rounded vowel</td>
<td>sunto</td>
<td>u</td>
</tr>
</tbody>
</table>

Additional Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>`'</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>`,</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td><code>.</code></td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

Romanian (ro-RO)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Romanian voice supported by Amazon Polly.

Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bubă</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>după</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>george</td>
<td>S</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>afacere</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>a griș</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>harpă</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>baie</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>coș</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lampa</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>mama</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nor</td>
<td>t</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pilă</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>rampă</td>
<td>r</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>soare</td>
<td>s</td>
</tr>
<tr>
<td>ʃ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>maşină</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tata</td>
<td>t</td>
</tr>
<tr>
<td>ʦ</td>
<td>ts</td>
<td>voiceless alveolar affricate</td>
<td>țără</td>
<td>s</td>
</tr>
<tr>
<td>ʧ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>ceai</td>
<td>S</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>viață</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>beau</td>
<td>u</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>mozol</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>joacă</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

| a   | @       | mid central vowel | babă | @       |
| a   | a       | open front unrounded vowel | casa | a       |
| e   | e       | close-mid front unrounded vowel | elan | e       |
| e̯  | e^      | non-syllabic close-mid front unrounded vowel | beau | e       |
| i   | i       | close front unrounded vowel | mie   | i       |
| o   | o       | close-mid back rounded vowel | oră   | o       |
| oa  | o^a     | diphthong | oare | o       |
| u   | u       | close back rounded vowel | unde  | u       |
| ɨ   | 1       | close central unrounded vowel | România | i |

**Additional Symbols**

| '   | "      | primary stress | Alabama |

---

76
IPA | X-SAMPA | Description | Example | Viseme
--- | --- | --- | --- | ---
ˌ | % | secondary stress | Alabama | 
ˌ |  | syllable boundary | A.la.ba.ma | 

**Russian (ru-RU)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Russian voices that are supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consonants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>борт</td>
<td>p</td>
</tr>
<tr>
<td>bʲ</td>
<td>b'</td>
<td>palatalized voiced bilabial plosive</td>
<td>бюро</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>дом</td>
<td>t</td>
</tr>
<tr>
<td>dʲ</td>
<td>d'</td>
<td>palatalized voiced alveolar plosive</td>
<td>дядя</td>
<td>t</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>флаг</td>
<td>f</td>
</tr>
<tr>
<td>fʲ</td>
<td>f'</td>
<td>palatalized voiceless labiodental fricative</td>
<td>февраль</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>нога</td>
<td>k</td>
</tr>
<tr>
<td>gʲ</td>
<td>g'</td>
<td>palatalized voiced velar plosive</td>
<td>герой</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>дизайн, ящик</td>
<td>i</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>кот</td>
<td>k</td>
</tr>
<tr>
<td>kʲ</td>
<td>k'</td>
<td>palatalized voiceless velar plosive</td>
<td>кино</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>лампа</td>
<td>t</td>
</tr>
<tr>
<td>lʲ</td>
<td>l'</td>
<td>palatalized alveolar lateral approximant</td>
<td>лес</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>мама</td>
<td>p</td>
</tr>
<tr>
<td>mʲ</td>
<td>m'</td>
<td>palatalized bilabial nasal</td>
<td>мяч</td>
<td>p</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>--------</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>нос</td>
<td>t</td>
</tr>
<tr>
<td>n'</td>
<td>n'</td>
<td>palatalized alveolar nasal</td>
<td>няня</td>
<td>t</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>пapa</td>
<td>p</td>
</tr>
<tr>
<td>p'</td>
<td>p'</td>
<td>palatalized voiceless bilabial plosive</td>
<td>перо</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>роза</td>
<td>r</td>
</tr>
<tr>
<td>r'</td>
<td>r'</td>
<td>palatalized alveolar trill</td>
<td>рюмка</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>сыр</td>
<td>s</td>
</tr>
<tr>
<td>s'</td>
<td>s'</td>
<td>palatalized voiceless alveolar fricative</td>
<td>сердце, русь</td>
<td>s</td>
</tr>
<tr>
<td>c:</td>
<td>s:</td>
<td>long voiceless alveolo-palatal fricative</td>
<td>щека</td>
<td>J</td>
</tr>
<tr>
<td>s</td>
<td>s`</td>
<td>voiceless retroflex fricative</td>
<td>шум</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>точка</td>
<td>t</td>
</tr>
<tr>
<td>t'</td>
<td>t'</td>
<td>palatalized voiceless alveolar plosive</td>
<td>тётя</td>
<td>t</td>
</tr>
<tr>
<td>ts</td>
<td>ts</td>
<td>voiceless alveolar affricate</td>
<td>царь</td>
<td>s</td>
</tr>
<tr>
<td>ts\</td>
<td>ts\</td>
<td>voiceless alveolo-palatal affricate</td>
<td>час</td>
<td>J</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>вор</td>
<td>f</td>
</tr>
<tr>
<td>v'</td>
<td>v'</td>
<td>palatalized voiced labiodental fricative</td>
<td>верфь</td>
<td>f</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>хор</td>
<td>k</td>
</tr>
<tr>
<td>x'</td>
<td>x'</td>
<td>palatalized voiceless velar fricative</td>
<td>химия</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>зуб</td>
<td>s</td>
</tr>
<tr>
<td>z'</td>
<td>z'</td>
<td>palatalized voiced alveolar fricative</td>
<td>зима</td>
<td>s</td>
</tr>
</tbody>
</table>
### Phoneme and Viseme Tables for Supported Languages

#### IPA and X-SAMPA for Supported Languages

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʑ⁵</td>
<td>ʑ \ː</td>
<td>long voiced alveolo-palatal fricative</td>
<td>уезжать</td>
<td>J</td>
</tr>
<tr>
<td>ʐ`</td>
<td>ʐ`</td>
<td>voiced retroflex fricative</td>
<td>жена</td>
<td>S</td>
</tr>
</tbody>
</table>

#### Vowels

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ə</td>
<td>ə</td>
<td>mid central vowel</td>
<td>канарейка</td>
<td>@</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>два, яблоко</td>
<td>a</td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>close-mid front unrounded vowel</td>
<td>печь</td>
<td>e</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>это</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>один, четыре</td>
<td>i</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>кот</td>
<td>o</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>муж, вьюга</td>
<td>u</td>
</tr>
<tr>
<td>ɨ</td>
<td>ɨ</td>
<td>close central unrounded vowel</td>
<td>мышь</td>
<td>i</td>
</tr>
</tbody>
</table>

### Spanish (es-ES)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Spanish voices that are supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ɾ⁴</td>
<td>ɾ⁴</td>
<td>alveolar flap</td>
<td>pero, bravo, amor, eterno</td>
<td>t</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bestia</td>
<td>p</td>
</tr>
<tr>
<td>β</td>
<td>B</td>
<td>voiced bilabial fricative</td>
<td>bebé</td>
<td>B</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>cuando</td>
<td>t</td>
</tr>
<tr>
<td>δ</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>arder</td>
<td>T</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------------------------------</td>
<td>------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fase, café</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>gato, lengua, guerra</td>
<td>k</td>
</tr>
<tr>
<td>γ</td>
<td>G</td>
<td>voiced velar fricative</td>
<td>trigo, Argos</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>hacia, tierra, radio, viuda</td>
<td>i</td>
</tr>
<tr>
<td>ʝ</td>
<td>j\</td>
<td>voiced palatal fricative</td>
<td>enhielar, sayo, inyectado, desyerba</td>
<td>J</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>caña, laca, quisimos</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lino, calor, principal</td>
<td>t</td>
</tr>
<tr>
<td>ʎ</td>
<td>L</td>
<td>palatal lateral approximant</td>
<td>llave, pollo</td>
<td>J</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>madre, comer, anfibio</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nido, anillo, sin</td>
<td>t</td>
</tr>
<tr>
<td>ñ</td>
<td>J</td>
<td>palatal nasal</td>
<td>cabaña, ñoquis</td>
<td>J</td>
</tr>
<tr>
<td>η</td>
<td>N</td>
<td>velar nasal</td>
<td>cinco, venga</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pozo, topo</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>perro, enrachado</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>saco, casa, puertas</td>
<td>s</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tamiz, átomo</td>
<td>t</td>
</tr>
<tr>
<td>ʝ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chubasco</td>
<td>S</td>
</tr>
<tr>
<td>θ</td>
<td>T</td>
<td>voiceless dental fricative</td>
<td>cereza, zorro, lacero, paz</td>
<td>T</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>fuego, fuimos, cuota, cuadro</td>
<td>u</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>jamón, general, suje, reloj</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>rasgo, mismo</td>
<td>s</td>
</tr>
</tbody>
</table>

**Vowels**
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>open front</td>
<td>tanque</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unrounded</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>vowel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>close-mid</td>
<td>peso</td>
<td>e</td>
</tr>
<tr>
<td></td>
<td></td>
<td>front</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>unrounded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front</td>
<td>cinco</td>
<td>i</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vowel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid</td>
<td>bosque</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td></td>
<td>back</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rounded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close-mid</td>
<td>publicar</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td></td>
<td>front</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>unrounded</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>vowel</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Additional Symbols**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ˈ</td>
<td>&quot;</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>′</td>
<td>%</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>‾</td>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

**Spanish, Mexican (es-MX)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Mexican Spanish voice that is supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>alveolar flap</td>
<td>pero, bravo, amor, eterno</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bestia</td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>voiced bilabial fricative</td>
<td>bebé</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>cuando</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>ð</td>
<td>voiced dental fricative</td>
<td>arder</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fase, café</td>
<td>f</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>voiced velar plosive</td>
<td>gato, lengua, guerra</td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>voiced velar fricative</td>
<td>trigo, Argos</td>
<td>k</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>---------------------------------</td>
<td>------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>hacia, tierra, radio, lluvia</td>
<td>i</td>
</tr>
<tr>
<td>j</td>
<td>j\</td>
<td>voiced palatal fricative</td>
<td>enhielar, sayo, inyectado, desyerba</td>
<td>J</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>cana, laca, quisimos</td>
<td>K</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>lateral alveolar approximant</td>
<td>lino, calor, principal</td>
<td>T</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>madre, comer, anfibio</td>
<td>P</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nido, anillo, sin</td>
<td>T</td>
</tr>
<tr>
<td>ñ</td>
<td>ñ</td>
<td>palatal nasal</td>
<td>cabaña, ñoquis</td>
<td>J</td>
</tr>
<tr>
<td>ñ</td>
<td>ñ</td>
<td>velar nasal</td>
<td>angosto, increíble</td>
<td>K</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pozo, topo</td>
<td>P</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>perro, enraizado</td>
<td>R</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>saco, casa, puertas</td>
<td>S</td>
</tr>
<tr>
<td>S</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>show, flash</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tamiz, átomo</td>
<td>T</td>
</tr>
<tr>
<td>tʃ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chubasco</td>
<td>S</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>fuego, fuimos, cuota, cuadro</td>
<td>U</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>jamón, general, peaje, reloj</td>
<td>K</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>rasgo, mismo</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

| a   | a       | central open unrounded vowel   | tanque                       | a      |
| e   | e       | close-mid front unrounded vowel| peso                         | e      |
| i   | i       | close front unrounded vowel    | cinco                        | i      |
### IPA and X-SAMPA Codes for Spanish, US (es-US)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the US Spanish voices that are supported by Amazon Polly.

#### Phoneme/Viseme Table

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>bosque</td>
<td>o</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>publicar</td>
<td>u</td>
</tr>
</tbody>
</table>

### Additional Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>,</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>4</td>
<td>alveolar flap</td>
<td>pero, bravo, amor, eterno</td>
<td>t</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>bestia</td>
<td>p</td>
</tr>
<tr>
<td>β</td>
<td>B</td>
<td>voiced bilabial fricative</td>
<td>bebé</td>
<td>B</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>cuando</td>
<td>t</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>voiced dental fricative</td>
<td>arder</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fase, café</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>gato, lengua, guerra</td>
<td>k</td>
</tr>
<tr>
<td>γ</td>
<td>G</td>
<td>voiced velar fricative</td>
<td>trigo, Argos</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>hacia, tierra, radio, viuda</td>
<td>i</td>
</tr>
<tr>
<td>j\</td>
<td>j\</td>
<td>voiced palatal fricative</td>
<td>enhielar, sayo, inyectado, desyerba</td>
<td>J</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>caña, laca, quisimos</td>
<td>k</td>
</tr>
</tbody>
</table>
## Phoneme and Viseme Tables for Supported Languages

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>l</td>
<td>l</td>
<td>lateral alveolar approximant</td>
<td>lino, calor, principal</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>madre, comer, anfíbio</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>nido, anillo, sin</td>
<td>t</td>
</tr>
<tr>
<td>ñ</td>
<td>J</td>
<td>palatal nasal</td>
<td>cabaña, ñoquis</td>
<td>J</td>
</tr>
<tr>
<td>nj</td>
<td>N</td>
<td>velar nasal</td>
<td>angosto, increíble</td>
<td>k</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pozo, topo</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>perro, enrachado</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>saco, casa, puertas</td>
<td>s</td>
</tr>
<tr>
<td>ſ</td>
<td>S</td>
<td>voiceless postalveolar fricative</td>
<td>show, flash</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tamiz, átomo</td>
<td>t</td>
</tr>
<tr>
<td>ſ</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>chubasco</td>
<td>S</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>fuego, fuimos, cuota, cuadro</td>
<td>u</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>voiceless velar fricative</td>
<td>jamón, general, peaje, reloj</td>
<td>k</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>rasgo, mismo</td>
<td>s</td>
</tr>
</tbody>
</table>

### Vowels

<table>
<thead>
<tr>
<th>IPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>central open unrounded vowel</td>
<td>tanque</td>
<td>a</td>
</tr>
<tr>
<td>e</td>
<td>close-mid front unrounded vowel</td>
<td>peso</td>
<td>e</td>
</tr>
<tr>
<td>i</td>
<td>close front unrounded vowel</td>
<td>cinco</td>
<td>i</td>
</tr>
<tr>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>bosque</td>
<td>o</td>
</tr>
<tr>
<td>u</td>
<td>close back rounded vowel</td>
<td>publicar</td>
<td>u</td>
</tr>
</tbody>
</table>

### Additional Symbols

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>.</td>
<td>%</td>
<td>secondary stress</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
</tr>
</tbody>
</table>

**Swedish (sv-SE)**

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Swedish voice supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
</table>
| Consonants
<p>| b   | b       | voiced bilabial plosive  | bil     | p      |
| d   | d       | voiced alveolar plosive  | dal     | t      |
| d<code>  | d</code>      | voiced retroflex plosive | bord    | t      |
| f   | f       | voiceless labiodental fricative | fil | f |
| g   | g       | voiced velar plosive     | gås     | k      |
| h   | h       | voiceless glottal fricative | hal | k |
| j   | j       | palatal approximant      | jag     | i      |
| k   | k       | voiceless velar plosive  | kal     | k      |
| l   | l       | alveolar lateral approximant | lös | t |
| l<code>  | l</code>      | retroflex lateral approximant | härlig | t |
| m   | m       | bilabial nasal           | mil     | p      |
| n   | n       | alveolar nasal           | nålar   | t      |
| n<code>  | n</code>      | retroflex nasal          | barn    | t      |
| N   | N       | velar nasal              | ring    | k      |
| p   | p       | voiceless bilabial plosive | pil | p |
| r   | r       | alveolar trill           | ris     | r      |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>sil</td>
<td>s</td>
</tr>
<tr>
<td>ş</td>
<td>ş\</td>
<td>voiceless alveolo-palatal fricative</td>
<td>tjock</td>
<td>J</td>
</tr>
<tr>
<td>ʂ</td>
<td>ʂ`</td>
<td>voiceless retroflex fricative</td>
<td>fors, schlager</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tal</td>
<td>t</td>
</tr>
<tr>
<td>ť</td>
<td>t`</td>
<td>voiceless retroflex plosive</td>
<td>hjort</td>
<td>t</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>vår</td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>labial-velar approximant</td>
<td>aula, airways</td>
<td>u</td>
</tr>
<tr>
<td>ɻ</td>
<td>x\</td>
<td>voiceless palatal-velar fricative</td>
<td>sjuk</td>
<td>k</td>
</tr>
</tbody>
</table>

**Vowels**

| ø   | 2       | close-mid front rounded vowel | föll, förr | o      |
| øː  | 2ː      | long close-mid front rounded vowel | föl, nöt, för | o      |
| øː  | 8       | close-mid central rounded vowel | buss, full | o      |
| a   | @       | mid central vowel | pojken | @      |
| uː  | }ː      | long close central rounded vowel | hus, ful | u      |
| a   | a       | open front unrounded vowel | hall, matt | a      |
| æː  | {       | near-open front unrounded vowel | herr    | a      |
| øː  | Aː      | long open back unrounded vowel | hal, mat | a      |
| eː  | eː      | long close-mid front unrounded vowel | vet, hel | e      |
| ɛ   | E       | open-mid front unrounded vowel | vett, rätt, hetta, håll | E      |
| ɛː  | Eː      | long open-mid front unrounded vowel | säi, häl, här | E:     |
Table of Phonemes and Visemes for Supported Languages

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>i:</td>
<td>i:</td>
<td>long close front unrounded vowel</td>
<td>vit, sil</td>
<td>i:</td>
</tr>
<tr>
<td>ɾ</td>
<td>ɾ̝̊</td>
<td>voiceless fricated alveolar flap</td>
<td>bir</td>
<td>t</td>
</tr>
<tr>
<td>ɾ̝</td>
<td>ɾ̝̊</td>
<td>fricated alveolar flap</td>
<td>raf</td>
<td>t</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>raf</td>
<td>p</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>voiceless palatal plosive</td>
<td>kedi</td>
<td>k</td>
</tr>
</tbody>
</table>

Additional Symbols

- `ˈ` primary stress
- `ˌ` secondary stress
- `.` syllable boundary

Turkish (tr-TR)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Turkish voice supported by Amazon Polly.

**Phoneme/Viseme Table**

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>ɾ</td>
<td>alveolar flap</td>
<td>durum</td>
<td>t</td>
</tr>
<tr>
<td>ɾ̈</td>
<td>ɾ̈</td>
<td>voiceless fricated alveolar flap</td>
<td>bir</td>
<td>t</td>
</tr>
<tr>
<td>ɾ̝̊</td>
<td>ɾ̝̊</td>
<td>fricated alveolar flap</td>
<td>raf</td>
<td>t</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>voiced bilabial plosive</td>
<td>raf</td>
<td>p</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>voiceless palatal plosive</td>
<td>kedi</td>
<td>k</td>
</tr>
<tr>
<td>IPA</td>
<td>X-SAMPA</td>
<td>Description</td>
<td>Example</td>
<td>Viseme</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>----------------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>voiced alveolar plosive</td>
<td>dede</td>
<td>t</td>
</tr>
<tr>
<td>dʒ</td>
<td>dZ</td>
<td>voiced postalveolar affricate</td>
<td>cam</td>
<td>S</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>voiceless labiodental fricative</td>
<td>fare</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>voiced velar plosive</td>
<td>galibi</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>voiceless glottal fricative</td>
<td>hasta</td>
<td>k</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>palatal approximant</td>
<td>yat</td>
<td>i</td>
</tr>
<tr>
<td>j̥</td>
<td>J̥</td>
<td>voiced palatal plosive</td>
<td>genç</td>
<td>J</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>voiceless velar plosive</td>
<td>akıl</td>
<td>k</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>alveolar lateral approximant</td>
<td>lale</td>
<td>t</td>
</tr>
<tr>
<td>ł</td>
<td>Ł</td>
<td>velarized alveolar lateral approximant</td>
<td>labirent</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>bilabial nasal</td>
<td>maaş</td>
<td>p</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>alveolar nasal</td>
<td>anı</td>
<td>t</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>ip</td>
<td>p</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>ses</td>
<td>s</td>
</tr>
<tr>
<td>ş</td>
<td>Ş</td>
<td>voiceless postalveolar fricative</td>
<td>aşş</td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>üttü</td>
<td>t</td>
</tr>
<tr>
<td>y̥</td>
<td>tS</td>
<td>voiceless postalveolar affricate</td>
<td>çaba</td>
<td>S</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>voiced labiodental fricative</td>
<td>ekvator, kahveci, akvaryum, isveçli, teşviki, cetvel</td>
<td>f</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>voiced alveolar fricative</td>
<td>ver</td>
<td>s</td>
</tr>
<tr>
<td>ʒ</td>
<td>Z</td>
<td>voiced postalveolar fricative</td>
<td>azik</td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ø</td>
<td>2</td>
<td>close-mid front rounded vowel</td>
<td>göl</td>
<td>0</td>
</tr>
<tr>
<td>öe</td>
<td>9</td>
<td>open-mid front rounded vowel</td>
<td>banliyö</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>open front unrounded vowel</td>
<td>kal</td>
<td>a</td>
</tr>
<tr>
<td>aː</td>
<td>aː</td>
<td>long open front unrounded vowel</td>
<td>davacı</td>
<td>a</td>
</tr>
<tr>
<td>æ</td>
<td>{</td>
<td>near-open front unrounded vowel</td>
<td>özlem, güvenlik, gürel, somersault</td>
<td>a</td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>close-mid front unrounded vowel</td>
<td>keçi</td>
<td>e</td>
</tr>
<tr>
<td>ɛ</td>
<td>E</td>
<td>open-mid front unrounded vowel</td>
<td>dede</td>
<td>E</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>close front unrounded vowel</td>
<td>bir</td>
<td>i</td>
</tr>
<tr>
<td>iː</td>
<td>iː</td>
<td>long close front unrounded vowel</td>
<td>izah</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>l</td>
<td>near-close near-front unrounded vowel</td>
<td>keçi</td>
<td>i</td>
</tr>
<tr>
<td>ʌ</td>
<td>M</td>
<td>close back unrounded vowel</td>
<td>kıl</td>
<td>i</td>
</tr>
<tr>
<td>ɔ</td>
<td>o</td>
<td>close-mid back rounded vowel</td>
<td>kol</td>
<td>o</td>
</tr>
<tr>
<td>ɔː</td>
<td>ɔː</td>
<td>long close-mid back rounded vowel</td>
<td>dolar</td>
<td>o</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>close back rounded vowel</td>
<td>durum</td>
<td>u</td>
</tr>
<tr>
<td>uː</td>
<td>uː</td>
<td>long close back rounded vowel</td>
<td>ruhum</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>dolu</td>
<td>u</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>close front rounded vowel</td>
<td>güvenlik</td>
<td>u</td>
</tr>
<tr>
<td>yː</td>
<td>Y</td>
<td>near-close near-front rounded vowel</td>
<td>aşı</td>
<td>u</td>
</tr>
</tbody>
</table>

**Additional Symbols**

| ˈ  | ˈ       | primary stress | Alabama |
| ˌ  | %       | secondary stress | Alabama |
Welsh (cy-GB)

The following table lists the International Phonetic Alphabet (IPA) phonemes, the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols, and the corresponding visemes for the Welsh voice supported by Amazon Polly.

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
<td></td>
</tr>
</tbody>
</table>

### Consonants

- **b**: <b>voiced bilabial plosive</b>  
  **Example**: baban  
  **Viseme**: p
- **d**: <d>voiced alveolar plosive</d>  
  **Example**: deg  
  **Viseme**: t
- **dʒ**: <dʒ>voiced postalveolar affricate</dʒ>  
  **Example**: garej  
  **Viseme**: S
- **ð**: <ð>voiced dental fricative</ð>  
  **Example**: deuddeg  
  **Viseme**: T
- **f**: <f>voiceless labiodental fricative</f>  
  **Example**: ffacs  
  **Viseme**: f
- **g**: <g>voiced velar plosive</g>  
  **Example**: gadael  
  **Viseme**: k
- **h**: <h>voiceless glottal fricative</h>  
  **Example**: haearn  
  **Viseme**: k
- **j**: <j>palatal approximant</j>  
  **Example**: astudio  
  **Viseme**: i
- **k**: <k>voiceless velar plosive</k>  
  **Example**: cant  
  **Viseme**: k
- **l**: <l>alveolar lateral approximant</l>  
  **Example**: lan  
  **Viseme**: t
- **ɬ**: <ɬ>voiceless alveolar lateral fricative</ɬ>  
  **Example**: llan  
  **Viseme**: t
- **m**: <m>bilabial nasal</m>  
  **Example**: mae  
  **Viseme**: p
- **m̥**: <m̥>voiceless bilabial nasal</m̥>  
  **Example**: ymhen  
  **Viseme**: p
- **n**: <n>alveolar nasal</n>  
  **Example**: naw  
  **Viseme**: t
- **n̥**: <n̥>voiceless alveolar nasal</n̥>  
  **Example**: anhawster  
  **Viseme**: t
- **ŋ**: <ŋ>velar nasal</ŋ>  
  **Example**: argyfwng  
  **Viseme**: k
- **ŋ̊**: <ŋ̊>voiceless velar nasal</ŋ̊>  
  **Example**: anghenion  
  **Viseme**: k
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p</td>
<td>voiceless bilabial plosive</td>
<td>pump</td>
<td>p</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>alveolar trill</td>
<td>rhoi</td>
<td>r</td>
</tr>
<tr>
<td>r_0</td>
<td>r_0</td>
<td>voiceless alveolar trill</td>
<td>garw</td>
<td>r</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>voiceless alveolar fricative</td>
<td>saith</td>
<td>s</td>
</tr>
<tr>
<td>S</td>
<td>siawns</td>
<td>voiceless postalveolar fricative</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>voiceless alveolar plosive</td>
<td>tegan</td>
<td>t</td>
</tr>
<tr>
<td>tS</td>
<td>cytsain</td>
<td>voiceless postalveolar affricate</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>T</td>
<td>aberth</td>
<td>voiceless dental fricative</td>
<td></td>
<td>T</td>
</tr>
<tr>
<td>v</td>
<td>prawf</td>
<td>voiced labiodental fricative</td>
<td></td>
<td>f</td>
</tr>
<tr>
<td>w</td>
<td>rhagwel</td>
<td>labial-velar approximant</td>
<td></td>
<td>u</td>
</tr>
<tr>
<td>X</td>
<td>chwech</td>
<td>voiceless uvular fricative</td>
<td></td>
<td>k</td>
</tr>
<tr>
<td>Z</td>
<td>aids</td>
<td>voiced alveolar fricative</td>
<td></td>
<td>s</td>
</tr>
<tr>
<td>S</td>
<td>rouge</td>
<td>voiced postalveolar fricative</td>
<td></td>
<td>S</td>
</tr>
</tbody>
</table>

**Vowels**

<p>| a   | @       | mid central vowel                  | ychwanega | @     |
| a   | a       | open front unrounded vowel          | acen      | a     |
| ai  | ai      | diphthong                           | dau       | a     |
| au  | au      | diphthong                           | awdur     | a     |
| o:  | A:      | long open back unrounded vowel      | mab       | a     |
| a:i | A:1     | diphthong                           | aelod     | a     |
| e:  | e:      | long close-mid front unrounded vowel| peth      | e     |
| E   | pedwar  | open-mid front unrounded vowel      |           | E     |</p>
<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ɛi</td>
<td>Ei</td>
<td>diphthong</td>
<td>beic</td>
<td>E</td>
</tr>
<tr>
<td>i:</td>
<td>i:</td>
<td>long close front unrounded vowel</td>
<td>tri</td>
<td>i</td>
</tr>
<tr>
<td>ɪ</td>
<td>I</td>
<td>near-close near-front unrounded vowel</td>
<td>miliwn</td>
<td>i</td>
</tr>
<tr>
<td>iu</td>
<td>1u</td>
<td>diphthong</td>
<td>unigrwy</td>
<td>i</td>
</tr>
<tr>
<td>o:</td>
<td>o:</td>
<td>long close-mid back rounded vowel</td>
<td>oddi</td>
<td>o</td>
</tr>
<tr>
<td>ɔ</td>
<td>O</td>
<td>open-mid back rounded vowel</td>
<td>oddieithr</td>
<td>O</td>
</tr>
<tr>
<td>ɔi</td>
<td>Oi</td>
<td>diphthong</td>
<td>troi</td>
<td>O</td>
</tr>
<tr>
<td>ɔu</td>
<td>Ou</td>
<td>diphthong</td>
<td>rownd</td>
<td>O</td>
</tr>
<tr>
<td>u:</td>
<td>u:</td>
<td>long close back rounded vowel</td>
<td>cwch</td>
<td>u</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>near-close near-back rounded vowel</td>
<td>acwstig</td>
<td>u</td>
</tr>
<tr>
<td>ui</td>
<td>Ui</td>
<td>diphthong</td>
<td>wyth</td>
<td>u</td>
</tr>
</tbody>
</table>

**Additional Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>primary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>,</td>
<td>secondary stress</td>
<td>Alabama</td>
</tr>
<tr>
<td>.</td>
<td>syllable boundary</td>
<td>A.la.ba.ma</td>
</tr>
</tbody>
</table>
Amazon Polly has a *Neural TTS (NTTS)* system that can produce even higher quality voices than its standard voices. The NTTS system produces the most natural and human-like text-to-speech voices possible.

Standard TTS voices use concatenative synthesis. This method strings together (concatenates) the phonemes of recorded speech, producing very natural-sounding synthesized speech. However, the inevitable variations in speech and the techniques used to segment the waveforms limits the quality of speech.

The Amazon Polly Neural TTS system doesn't use standard concatenative synthesis to produce speech. It has two parts:

- A neural network that converts a sequence of phonemes—the most basic units of language—into a sequence of *spectrograms*, which are snapshots of the energy levels in different frequency bands
- A vocoder, which converts the spectrograms into a continuous audio signal.

The first component of the neural TTS system is a sequence-to-sequence model. This model doesn't create its results solely from the corresponding input but also considers how the sequence of the elements of the input work together. The model chooses the spectrograms that it outputs so that their frequency bands emphasize acoustic features that the human brain uses when processing speech.

The output of this model then passes to a neural vocoder. This converts the spectrograms into speech waveforms. When trained on the large data sets used to build general-purpose concatenative-synthesis systems, this sequence-to-sequence approach will yield higher-quality, more natural-sounding voices.

The Aria (New Zealand English), Ayanda (South African English), Gabrielle (Canadian French), Kevin (US English), and Olivia (Australian English) voices are only supported by Amazon Polly when using NTTS. All other voices have a counterpart created using the standard TTS method. When using an NTTS-only voice, the TTS engine parameter must be set to `neural`, whether using the console or API.

**Topics**

- Feature and Region Compatibility (p. 93)
- The Voice Engine (p. 94)
- Neural Voices (p. 96)
- NTTS Newscaster Speaking Style (p. 97)

**Feature and Region Compatibility**

Neural voices aren't available in all AWS Regions, nor do they support all Amazon Polly features.

Neural voices are supported in the following Regions:

- US East (N. Virginia): us-east-1
- US West (Oregon): us-west-2
Endpoints and protocols for these Regions are identical to those used for standard voices. For more information, see Amazon Polly endpoints and quotas.

The following features are supported for neural voices:

- Real-time and asynchronous speech synthesis operations.
- Newscaster speaking style. For more information about the speaking styles, see NTTS Newscaster Speaking Style (p. 97).
- All speechmarks.
- Many (but not all) of the SSML tags that are supported by Amazon Polly. For more information about NTTS-supported SSML tags, see Supported SSML Tags (p. 112).

As with standard voices, you can choose from various sampling rates to optimize the bandwidth and audio quality for your application. Valid sampling rates for standard and neural voices are 8 kHz, 16 kHz, 22 kHz, or 24 kHz. The default for standard voices is 22 kHz. The default for neural voices is 24 kHz. Amazon Polly supports MP3, OGG (Vorbis), and raw PCM audio stream formats.

The Voice Engine

Amazon Polly enables you to use either neural or standard voice with the engine property. It has two possible values: **Standard** or **Neural**, indicating whether to use a standard or neural voice. **Standard** is the default value.

**Important**

If you are not in one of the regions where NTTS is supported, only the standard voice engine will be displayed in the console. If the neural engine is not displayed, check your region. For more information on the regions where NTTS can be used, see Feature and Region Compatibility (p. 93).

When using an NTTS-only voice, the TTS engine parameter must be set to **neural**, whether using the console or API.

Choosing the Voice Engine (Console)

**To choose a voice engine (console)**

1. Open the Amazon Polly console at [https://console.aws.amazon.com/polly/](https://console.aws.amazon.com/polly/).
2. On the Text-to-Speech page, for **Engine**, choose **Standard** or **Neural**.
If you choose **Neural**, only neural voices are available and standard-only voices are disabled.

### Choosing the Voice Engine (CLI)

**To choose a voice engine (CLI)**

The `engine` parameter is optional, with two possible values: `standard` or `Neural`. Use this property when creating a `SynthesisTask` operation.
For example, you can use the following code to run the `start-speech-synthesis-task` AWS CLI command in the US West-2 (Oregon) region.

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (') for interior tags.

```sh
aws polly start-speech-synthesis-task \
  --engine neural \
  --region us-west-2 \
  --endpoint-url "https://polly.us-west-1.amazonaws.com/" \
  --output-format mp3 \
  --output-s3-bucket-name your-bucket-name \
  --output-s3-key-prefix optional/prefix/path/file \
  --voice-id Joanna \
  --text file://text_file.txt
```

This will result in a response that looks similar to this:

```json
"SynthesisTask": {
  "CreationTime": [..],
  "Engine": "neural",
  "OutputFormat": "mp3",
  "OutputUri": "https://s3.us-west-1.amazonaws.com/your-bucket-name/optional/prefix/path/file.<task_id>.mp3",
  "TextType": "text",
  "RequestCharacters": [..],
  "TaskStatus": "scheduled",
  "TaskId": [task_id],
  "VoiceId": "Joanna"
}
```

## Neural Voices

Neural voices are available in Australian English (en-AU), British English (en-GB), New Zealand English (en-NZ), South African English (en-ZA), US English (en-US), Canadian French (fr-CA), German (de-DE), Korean (ko-KR), Brazilian Portuguese (pt-BR), and US Spanish (es-US). The following table lists the voices.

<table>
<thead>
<tr>
<th>Language</th>
<th>Name/ID</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>English (Australian) (en-AU)</td>
<td>Olivia**</td>
<td>Female</td>
</tr>
<tr>
<td>English (British) (en-GB)</td>
<td>Amy*</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>Emma</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>Brian</td>
<td>Male</td>
</tr>
<tr>
<td>English (New Zealand) (en-NZ)</td>
<td>Aria**</td>
<td>Female</td>
</tr>
<tr>
<td>English (South African) (en-ZA)</td>
<td>Ayanda**</td>
<td>Female</td>
</tr>
<tr>
<td>English (US) (en-US)</td>
<td>Ivy</td>
<td>Female (child)</td>
</tr>
<tr>
<td></td>
<td>Joanna*</td>
<td>Female</td>
</tr>
</tbody>
</table>
NTTS Newscaster Speaking Style

People use different speaking styles, depending on context. Casual conversation, for example, sounds very different from a TV or radio newscast. Because of the way standard voices are made, they can't produce different speaking styles. However, neural voices can. They can be trained for a specific speaking style, with the variations and emphasis on certain parts of speech inherent in that style.

In addition to the default neural voices, Amazon Polly provides a newscaster speaking style that uses the neural system to generate speech in the style of a TV or radio newscaster. The Newscaster style is available with the Matthew and Joanna voices in US English (en-US), the Lupe voice in US Spanish (es-US), and the Amy voice in British English (en-GB).

To use the Newscaster style, first choose the neural engine and then use the syntax described in the following steps in your input text.

**Note**
To use any neural speaking style, you must use one of the AWS Regions that support neural voices. This option is not available in all Regions. For more information, see Feature and Region Compatibility (p. 93).

**To apply the Newscaster style (console)**

1. Open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Make sure that you are using an AWS Region where neural voices are supported.
3. On the Text-to-Speech page, for Engine, choose Neural.
4. Choose the language and voice you want to use.

Only Matthew and Joanna for US English (en-US), Lupe for US Spanish (es-US), and Amy for British English (en-GB) are available in the newscaster voice.

5. Turn on **SSML**.

6. Add input text to your text-to-speech request using the Newscaster style SSML syntax.

```xml
<amazon:domain name="news">text</amazon:domain>
```

For example, you might use the newscaster tag as follows:

```xml
<speak>
<amazon:domain name="news">
From the Tuesday, April 16th, 1912 edition of The Guardian newspaper:

The maiden voyage of the White Star liner Titanic, the largest ship ever launched ended in disaster.

The Titanic started her trip from Southampton for New York on Wednesday. Late on Sunday night she struck an iceberg off the Grand Banks of Newfoundland. By wireless telegraphy she sent out signals of distress, and several liners were near enough to catch and respond to the call.
</amazon:domain>
</speak>
```

7. Choose **Listen**.

**To apply the Newscaster style (CLI)**

1. In your API request, include the engine parameter with the `neural` value:

   ```
   --engine neural
   ```

2. Add input text to your API request using the Newscaster style SSML syntax.

   ```xml
   <amazon:domain name="news">text</amazon:domain>
   ```

For example, you might use the newscaster tag as follows:

```xml
<speak>
<amazon:domain name="news">
From the Tuesday, April 16th, 1912 edition of The Guardian newspaper:

The maiden voyage of the White Star liner Titanic, the largest ship ever launched ended in disaster.

The Titanic started her trip from Southampton for New York on Wednesday. Late on Sunday night she struck an iceberg off the Grand Banks of Newfoundland. By wireless telegraphy she sent out signals of distress, and several liners were near enough to catch and respond to the call.
</amazon:domain>
</speak>
```

For more information about SSML, see **Supported SSML Tags (p. 112)**.
Speech Marks

Speech marks are metadata that describe the speech that you synthesize, such as where a sentence or word starts and ends in the audio stream. When you request speech marks for your text, Amazon Polly returns this metadata instead of synthesized speech. By using speech marks in conjunction with the synthesized speech audio stream, you can provide your applications with an enhanced visual experience.

For example, combining the metadata with the audio stream from your text can enable you to synchronize speech with facial animation (lip-syncing) or to highlight written words as they're spoken.

Speech marks are available when using either neural or standard text-to-speech formats.

Topics
- Speech Mark Types (p. 99)
- Using Speech Marks (p. 100)
- Requesting Speech Marks (Console) (p. 103)

Speech Mark Types

You request speech marks using the SpeechMarkTypes option for either the SynthesizeSpeech or StartSpeechSynthesisTask commands. You specify the metadata elements that you want to return from your input text. You can request as many as four types of metadata but you must specify at least one per request. No audio output is generated with the request.

In the AWS CLI, for example:

```
--speech-mark-types='["sentence", "word", "viseme", "ssml"]'
```

Amazon Polly generates speech marks using the following elements:

- **sentence** – Indicates a sentence element in the input text.
- **word** – Indicates a word element in the text.
- **viseme** – Describes the face and mouth movements corresponding to each phoneme being spoken. For more information, see Visemes and Amazon Polly (p. 99).
- **ssml** – Describes a <mark> element from the SSML input text. For more information, see Generating Speech from SSML Documents (p. 105).

Visemes and Amazon Polly

A viseme represents the position of the face and mouth when saying a word. It is the visual equivalent of a phoneme, which is the basic acoustic unit from which a word is formed. Visemes are the basic visual building blocks of speech.

Each language has a set of viseme that correspond to their specific phonemes. In a language, each phoneme has a corresponding viseme that represents the shape that the mouth makes when forming the sound. However, not all visemes can be mapped to a particular phoneme because numerous
phonemes appear the same when spoken, even though they sound different. For example, in English, the words "pet" and "bet" are acoustically different. However, when observed visually (without sound), they look exactly the same.

The following chart shows a partial list of International Phonetic Alphabet (IPA) phonemes and Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) symbols as well as their corresponding visemes for US English voices.

For the complete table and tables for all available languages, see Phoneme and Viseme Tables for Supported Languages (p. 19).

<table>
<thead>
<tr>
<th>IPA</th>
<th>X-SAMPA</th>
<th>Description</th>
<th>Example</th>
<th>Viseme</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>Voiced bilabial plosive</td>
<td>bed</td>
<td>p</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>Voiced alveolar plosive</td>
<td>dig</td>
<td>t</td>
</tr>
<tr>
<td>ðʒ</td>
<td>dZ</td>
<td>Voiced postalveolar affricate</td>
<td>jump</td>
<td>S</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>Voiced dental fricative</td>
<td>then</td>
<td>T</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>Voiceless labiodental fricative</td>
<td>five</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>Voiced velar plosive</td>
<td>game</td>
<td>k</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>Voiceless glottal fricative</td>
<td>house</td>
<td>k</td>
</tr>
</tbody>
</table>

Using Speech Marks

Requesting Speech Marks

To request speech marks for input text, use the `synthesize-speech` command. Besides the input text, the following elements are required to return this metadata:

- `output-format`
  
  Amazon Polly supports only the JSON format when returning speech marks.

```bash
--output-format json
```

If you use an unsupported output format, Amazon Polly throws an exception.

- `voice-id`
  
  To ensure that the metadata matches the associated audio stream, specify the same voice that is used to generate the synthesized speech audio stream. The available voices don't have identical speech...
rates. If you use a voice other than the one used to generate the speech, the metadata will not match the audio stream.

```bash
--voice-id Joanna
```

- **speech-mark-types**

  Specify the type or types of speech marks you want. You can request any or all of the speech mark types, but must specify at least one type.

```bash
--speech-mark-types='["sentence", "word", "viseme", "ssml"]'
```

- **text-type**

  Plain text is the default input text for Amazon Polly, so you must use `text-type ssml` if you want to return SSML speech marks.

- **outfile**

  Specify the output file to which the metadata is written.

```bash
MaryLamb.txt
```

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (`\`) Unix continuation character at the end of each line with a caret (`^`) and use full quotation marks (`"`) around the input text with single quotes (`'`) for interior tags.

```bash
aws polly synthesize-speech \
  --output-format json \
  --voice-id <Voice ID> \
  --text 'Input text' \
  --speech-mark-types='["sentence", "word", "viseme"]' \
  outfile
```

## Speech Mark Output

Amazon Polly returns speech mark objects in a line-delimited JSON stream. A speech mark object contains the following fields:

- **time** – the timestamp in milliseconds from the beginning of the corresponding audio stream
- **type** – the type of speech mark (sentence, word, viseme, or ssml)
- **start** – the offset in bytes (not characters) of the start of the object in the input text (not including viseme marks)
- **end** – the offset in bytes (not characters) of the object's end in the input text (not including viseme marks)
- **value** – this varies depending on the type of speech mark
  - **SSML**: `<mark>` SSML tag
  - **viseme**: the viseme name
  - **word** or **sentence**: a substring of the input text, as delimited by the start and end fields

For example, Amazon Polly generates the following word speech mark object from the text "Mary had a little lamb":

```json
101
```
Speech Mark Examples

The following examples of speech mark requests show how to make common requests and the output that they generate.

Example 1: Speech Marks Without SSML

The following example shows you what requested metadata looks like on your screen for the simple sentence: “Mary had a little lamb.” For simplicity, we don't include SSML speech marks in this example.

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (') for interior tags.

```bash
aws polly synthesize-speech \
--output-format json \
--voice-id Joanna \
--text 'Mary had a little lamb.' \
--speech-mark-types='["viseme", "word", "sentence"]'
```

When you make this request, Amazon Polly returns the following in the .txt file:

```
{"time":0,"type":"sentence","start":0,"end":23,"value":"Mary had a little lamb."}
{"time":6,"type":"word","start":0,"end":4,"value":"Mary"}
{"time":180,"type":"viseme", "value":"t"}
{"time":373,"type":"word","start":5,"end":8,"value":"had"}
```

In this output, each part of the text is broken out in terms of speech marks:

- The sentence "Mary had a little lamb."
Amazon Polly Developer Guide
Requesting Speech Marks (Console)

• Each word in the text: "Mary", "had", "a", "little", and "lamb."
• The viseme for each sound in the corresponding audio stream: "p", "E", "r", "i", and so on. For more
information on visemes see Visemes and Amazon Polly (p. 99).

Example 2: Speech Marks with SSML
The process of generating speech marks from SSML-enhanced text is similar to the process when SSML is
not present. Use the synthesize-speech command, and specify the SSML-enhanced text and the type
of speech marks that you want, as shown in the following example. To make the example easier to read,
we do not include viseme speech marks, but these could be included as well.
The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the
backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation
marks (") around the input text with single quotes (') for interior tags.
aws polly synthesize-speech \
--output-format json \
--voice-id Joanna \
--text-type ssml \
--text '<speak><prosody volume="+20dB">Mary had <break time="300ms"/>a little <mark
name="animal"/>lamb</prosody></speak>' \
--speech-mark-types='["sentence", "word", "ssml"]' \
output.txt

When you make this request, Amazon Polly returns the following in the .txt ﬁle:
{"time":0,"type":"sentence","start":31,"end":95,"value":"Mary had <break time=\"300ms\"\/>a
little <mark name=\"animal\"\/>lamb"}
{"time":6,"type":"word","start":31,"end":35,"value":"Mary"}
{"time":325,"type":"word","start":36,"end":39,"value":"had"}
{"time":897,"type":"word","start":40,"end":61,"value":"<break time=\"300ms\"\/>"}
{"time":1291,"type":"word","start":61,"end":62,"value":"a"}
{"time":1373,"type":"word","start":63,"end":69,"value":"little"}
{"time":1635,"type":"ssml","start":70,"end":91,"value":"animal"}
{"time":1635,"type":"word","start":91,"end":95,"value":"lamb"}

Requesting Speech Marks (Console)
You can use the console to request speech marks from Amazon Polly. You can then view the metadata or
save it to a ﬁle.

To generate speech marks (console)
1.

Sign in to the AWS Management Console and open the Amazon Polly console at https://
console.aws.amazon.com/polly/.

2.

Choose the Text-to-Speech tab.

3.

Turn on SSML to use SSML.

4.

Type or paste your text into the input box.

5.

For Language, choose the language for your text.

6.

For Voice, choose the voice you want to use for the text.

7.

To change text pronunciation, expand Additional settings, turn on Customize pronunciation, and
for Apply lexicon, choose the desired lexicon.

8.

To verify that the speech is in its ﬁnal form, choose Listen.

103


9. Turn on **Speech file format settings**.

   **Note**
   Downloading MP3, OGG, or PCM formats will not generate speech marks.

10. For **File Format**, choose **Speech marks**.

11. For **Speech mark types**, choose the types of speech marks to generate. The option to choose **SSML** metadata is only available when **SSML** is on. For more information on using SSML with Amazon Polly see Generating Speech from SSML Documents (p. 105).

12. Choose **Download**.
Generating Speech from SSML Documents

You can use Amazon Polly to generate speech from either plain text or from documents marked up with Speech Synthesis Markup Language (SSML). Using SSML-enhanced text gives you additional control over how Amazon Polly generates speech from the text you provide.

For example, you can include a long pause within your text, or change the speech rate or pitch. Other options include:

- emphasizing specific words or phrases
- using phonetic pronunciation
- including breathing sounds
- whispering
- using the Newscaster speaking style.

For complete details on the SSML tags supported by Amazon Polly and how to use them, see Supported SSML Tags (p. 112)

When using SSML, there are several reserved characters that require special treatment. This is because SSML uses these characters as part of its code. In order to use them, you use a specific entity to escape them. For more information, see Reserved Characters in SSML (p. 105)

Amazon Polly provides these types of control with a subset of the SSML markup tags that are defined by Speech Synthesis Markup Language (SSML) Version 1.1, W3C Recommendation.

You can use SSML within the Amazon Polly console or by using the AWS CLI. The following topics show you how you can use SSML to generate speech and control the output so that it precisely fits your needs.

Topics
- Reserved Characters in SSML (p. 105)
- Using SSML (Console) (p. 107)
- Using SSML (AWS CLI) (p. 108)
- Supported SSML Tags (p. 112)

Reserved Characters in SSML

There are five predefined characters that can't normally be used within an SSML statement. These entities are reserved by the language specification. These characters are

<table>
<thead>
<tr>
<th>Name</th>
<th>Character</th>
<th>Escape code</th>
</tr>
</thead>
<tbody>
<tr>
<td>quotation mark</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
Reserved Characters

Name | Character | Escape code
--- | --- | ---
& | &amp; | &ampersand
’ | ’ | &apos; or single quotation mark
< | < | &lt; than sign
> | > | &gt; than sign

Because SSML uses these characters as part of its code, to use these symbols in SSML, you must escape the character when you use it. You use the escape code instead of the actual character so it displays properly while still creating a valid SSML document. For example, the following sentence

> We're using the lawyer at Peabody & Chambers, attorneys-at-law. 

would be rendered in SSML as

```
<speak>We’re using the lawyer at Peabody & Chambers, attorneys-at-law.</speak>
```

In this case, the special characters for the apostrophe and ampersand are escaped so the SSML document remains valid.

For the & , <, and > symbols, escape codes are always necessary when you use SSML. Additionally, when you use the apostrophe/single quotation mark (’) as an apostrophe, you must also use the escape code.

However, when you use the double quotation mark ("), or the apostrophe/single quotation mark (’) as a quotation mark, then whether or not you use the escape code is dependent on context.

Double quotation marks

- Must be escaped when in a attribute value delimited by double quotes. For example, in the following AWS CLI code

  ```
  --text "Pete 'Maverick' Mitchell"
  ```

- Do not need to be escaped when in textual context. For example, in the following

  ```
  He said, "Turn right at the corner."
  ```

- Do not need to be escaped when in a attribute value delimited by single quotes. For example, in the following AWS CLI code

  ```
  --text 'Pete "Maverick" Mitchell'
  ```
Using SSML (Console)

With SSML tags, you can customize and control aspects of speech such as pronunciation, volume, and speech rate. In the AWS Management Console, the SSML-enhanced text that you want to convert to audio is entered on the SSML tab of the Text-to-Speech page. Although text entered in plain text relies on default settings for the language and voice you've chosen, text enhanced with SSML tells Amazon Polly not only what you want to say, but how you want to say it. Except for the added SSML tags, Amazon Polly synthesizes SSML-enhanced text in the same way as it synthesizes plain text. See Exercise 2: Synthesizing Speech with Plain Text Input (Console) (p. 6) for more information.

When using SSML, you enclose the entire text in a `<speak>` tag to let Amazon Polly know that you're using SSML. For example:

```
<speak>Hi! My name is Joanna. I will read any text you type here.</speak>
```

You then use specific SSML tags on the text inside the `<speak>` tags to customize the way you want the text to sound. You can add a pause, change the pace of the speech, lower or raise the volume of the voice, or add many other customizations so that the text sounds right for you. For a full list of the SSML tags that you can use, see Supported SSML Tags (p. 112).

In the following example, you use an SSML tag to tell Amazon Polly to substitute "World Wide Web Consortium" for "W3C" when it speaks a short paragraph. You also use tags to introduce a pause and whisper a word. Compare the results of this exercise with that of Applying Lexicons Using the Console (Synthesize Speech) (p. 135).

For more information on SSML, with examples, see Supported SSML Tags (p. 112).

To synthesize speech from SSML-enhanced text (console)

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. If it isn't already displayed, choose the Text-to-Speech tab.
3. Turn on SSML.
4. Type or paste the following text in the text box:

```
<speak>
  He was caught up in the game.<break time="1s"/>
  In the middle of the
```

---

Single quotation marks

- Must be escaped when used as an apostrophe. For example, in the following

```
We've got to leave quickly.
```

- Do not need to be escaped when in textual context. For example, in the following

```
"And then I said, 'Don't quote me.'"
```

- Do not need to be escaped when in a code attribute delimited by double quotes. For example, in the following AWS CLI code

```
--text "Pete 'Maverick' Mitchell"
```
</speak>

The SSML tags tell Amazon Polly how to render the text:

- `<break time="1s"/>` tells Amazon Polly to pause 1 second between the first two sentences.
- `<sub alias="World Wide Web Consortium">W3C</sub>` tells Amazon Polly to substitute World Wide Web Consortium for the acronym W3C.
- `<amazon:effect name="whispered">Nice job</amazon:effect>` tells Amazon Polly to whisper the second instance of "Nice job.".

**Note**
When you use the AWS CLI, you enclose the input text in quotation marks to differentiate it from the surrounding code. The Amazon Polly console doesn't show you code, so you don't enclose input text in quotation marks when you use it.

5. For **Language**, choose **English, US**, then choose a voice.
6. To listen to the speech, choose **Listen**.
7. To save the speech file, choose **Download**. If you want to save it in a different format, expand **Additional settings**, turn on **Speech file format settings** and choose the format that you want, then choose **Download**.

### Using SSML (AWS CLI)

You can use the AWS CLI to synthesize SSML input text. The following examples show how to perform common tasks using the AWS CLI.

**Topics**
- Using SSML With the Synthesize-Speech Command (p. 108)
- Synthesizing an SSML-enhanced Document (p. 109)
- Using SSML for Common Amazon Polly Tasks (p. 110)

### Using SSML With the Synthesize-Speech Command

This example shows how to use the `synthesize-speech` command with an SSML string. When you use the `synthesize-speech` command, you typically provide the following:

- The input text (required)
- Opening and closing tags (required)
- The output format
- A voice

In this example, you specify a simple text string in quotation marks along with the required opening and closing `<speak>` tags.

**Important**
Although you don't use quotation marks around input text in the Amazon Polly console, you must use them in use the AWS CLI It's also important that you differentiate between the quotation marks around input text and quotations required for individual tags.
For example, you can use standard quotation marks ("), to enclose the input text, and single quotation marks (’) for interior tags, or vice versa. Either option works for Unix, Linux, and macOS. However, with Windows you must enclose the input text in standard quotations marks and use single quotation marks for the tags.

For all operating systems, you can use standard quotation marks ("), to enclose the input text, and single quotation marks (’) for interior tags. For example:

```bash
--text "<speak>Hello <break time='300ms'/> World</speak>"
```

For Unix, Linux, and macOS, you can also use the reverse, with single quotation marks (’) enclosing the input text and standard quotation marks ("), for interior tags:

```bash
--text '<speak>Hello <break time="300ms"/> World</speak>'
```

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (’) for interior tags.

```
aws polly synthesize-speech \
--text-type ssml \ 
--text '<speak>Hello world</speak>' \ 
--output-format mp3 \ 
--voice-id Joanna \
speech.mp3
```

To hear the synthesized speech, play the resulting speech.mp3 file using any audio player.

### Synthesizing an SSML-enhanced Document

For longer input text, you may find it easier to save your SSML content to a file and simply specify the file name in the `synthesize-speech` command. For example you could save the following to a file called example.xml:

```xml
<?xml version="1.0"?>
<speak version="1.1"
xmlns="http://www.w3.org/2001/10/synthesis"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2001/10/synthesis http://www.w3.org/TR/speech-synthesis11/synthesis.xsd"
xml:lang="en-US">Hello World</speak>
```

The `xml:lang` attribute specifies `en-US` (US English) as the language of the input text. For information about how the language of the input text and the language of the chosen voice affect the SynthesizeSpeech operation, see [Improving the Pronunciation of Foreign Words](#) (p. 112).

### To run an SSML-enhanced file

1. Save the SSML to a file (for example, example.xml).
2. Run the following `synthesize-speech` command from the path where the XML file is stored and specify the SSML file as input by substituting `file:\\example.xml` for the input text. Because this command points to a file instead of containing the actual input text, you don't use quotation marks.

   ```bash
   aws polly synthesize-speech --text-type ssml --text 'file:\example.xml' --output-format mp3 --voice-id Joanna output.mp3
   ```

   **Note**
   The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^).
aws polly synthesize-speech \
--text-type ssml \
--text file://example.xml \
--output-format mp3 \
--voice-id Joanna \
speech.mp3

3. To hear the synthesized speech, play the resulting speech.mp3 file using any audio player.

Using SSML for Common Amazon Polly Tasks

The following examples show how to use SSML tags to complete common Amazon Polly tasks. For more SSML tags, see Supported SSML Tags (p. 112).

To test the following examples, use the following synthesize-speech command with the appropriate SSML-enhanced text:

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (') for interior tags.

aws polly synthesize-speech \
--text-type ssml \
--text '<speak>Hello <break time="300ms"/> World</speak>' \
--output-format mp3 \
--voice-id Joanna \
speech.mp3

Adding a Pause

To add a pause between words, use the <break> element. The following SSML synthesize-speech command uses the <break> element to add a 300-millisecond delay between the words "Hello" and "World."

<speak>  
Hello <break time="300ms"/> World.  
</speak>

Controlling Volume, Pitch, and Speed

To control pitch, speaking rate, and speech volume, use the <prosody> element.

• The following synthesize-speech command uses the <prosody> element to control volume:

<speak>  
<prosody volume="+20dB">Hello world</prosody>  
</speak>

• The following synthesize-speech command uses the <prosody> element to control pitch:

<speak>  
<prosody pitch="x-high">Hello world</prosody>  
</speak>
• The following `synthesize-speech` command uses the `<prosody>` element to specify the speech rate (speaking speed):

```xml
<speak>
  <prosody rate="x-fast">Hello world.</prosody>
</speak>
```

• You can specify multiple attributes in a `<prosody>` element, as shown in the following examples:

```xml
<speak>
  <prosody volume="x-loud" pitch="x-high" rate="x-fast">Hello world.</prosody>
</speak>
```

### Whispering

To whisper words, use the `<amazon:effect name="whispered">` element. In the following example, the `<amazon:effect name="whispered">` element tells Amazon Polly to whisper "little lamb":

```xml
<speak>
  Mary has a <amazon:effect name="whispered">little lamb</amazon:effect>.
</speak>
```

To enhance this effect, use the `<prosody>` element to slightly slow down the whispered speech.

### Emphasizing Words

To stress a word or phrase, use the `<emphasis>` element.

```xml
<speak>
  <emphasis level="strong">Hello</emphasis> world how are you?
</speak>
```

### Specifying How to Say Certain Words

To provide information about the type of text to be spoken, use the `<say-as>` element.

For instance, in the following SSML, `<say-as>` indicates that the text 4/6 should be interpreted as a date. The attribute `interpret-as="date" format="dm"` indicates that it should be spoken as a date with the format month/day.

You can also use the `<say-as>` element to tell Amazon Polly to say numbers as fractions, telephone numbers, measurement units, and more.

```xml
<speak>
  Today is <say-as interpret-as="date" format="md">4/6</say-as>
</speak>
```

The resulting speech is "Today is June 4th." The `<say-as>` tag describes how the text should be interpreted by providing additional context with the `interpret-as` attribute.

To verify the accuracy of the synthesized speech, play the resulting `speech.mp3` file.
For more information on this element, see Controlling How Special Types of Words Are Spoken (p. 122).

**Improving the Pronunciation of Foreign Words**

Amazon Polly assumes that the input text is in the same language as the language spoken by the voice you choose. To improve the pronunciation of foreign words within input text, in the `synthesize-speech` call. Specify the target language with the `xml:lang` attribute. This tells Amazon Polly to apply different pronunciation rules for the foreign words that you tag.

The following examples show how to use different combinations of languages in the input text, and how to specify voices and the pronunciation of foreign words. For a complete list of available languages, see Languages Supported by Amazon Polly (p. 18).

In the following example, the voice (Joanna) is a US English voice. By default, Amazon Polly assumes that the input text is in the same language as the voice (in this case, US English). When you use the `xml:lang` tag, Amazon Polly interprets the text as Spanish and the text is spoken as the selected voice would pronounce Spanish words, according to the pronunciation rules of the foreign language. Without this tag, the text is spoken using the pronunciation rules of the selected voice.

```xml
<speak>
    That restaurant is terrific. <lang xml:lang="es-ES">Mucho gusto.</lang>
</speak>
```

Because the language of the input text is English, Amazon Polly maps the Spanish phonemes to the closest English phonemes. As a result, Joanna speaks the text as a native US speaker who pronounces the works correctly in Spanish, but with a US English accent.

**Note**

Some languages are more similar than others, and so some language combinations work better than others.

## Supported SSML Tags

Amazon Polly supports the following SSML tags:

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding a Pause (p. 113)</td>
<td><code>&lt;break&gt;</code></td>
<td>Full availability</td>
</tr>
<tr>
<td>Emphasizing Words (p. 114)</td>
<td><code>&lt;emphasis&gt;</code></td>
<td>Not available</td>
</tr>
<tr>
<td>Specifying Another Language for Specific Words (p. 115)</td>
<td><code>&lt;lang&gt;</code></td>
<td>Full availability</td>
</tr>
<tr>
<td>Placing a Custom Tag in Your Text (p. 116)</td>
<td><code>&lt;mark&gt;</code></td>
<td>Full availability</td>
</tr>
<tr>
<td>Adding a Pause Between Paragraphs (p. 116)</td>
<td><code>&lt;p&gt;</code></td>
<td>Full availability</td>
</tr>
<tr>
<td>Using Phonetic Pronunciation (p. 116)</td>
<td><code>&lt;phoneme&gt;</code></td>
<td>Full availability</td>
</tr>
<tr>
<td>Action</td>
<td>SSML Tag</td>
<td>Availability with Neural Voices</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Controlling Volume, Speaking Rate, and Pitch (p. 117)</td>
<td>&lt;prosody&gt;</td>
<td>Partial availability</td>
</tr>
<tr>
<td>Setting a Maximum Duration for Synthesized Speech (p. 119)</td>
<td>&lt;prosody amazon:max-duration&gt;</td>
<td>Not available</td>
</tr>
<tr>
<td>Adding a Pause Between Sentences (p. 121)</td>
<td>&lt;s&gt;</td>
<td>Full availability</td>
</tr>
<tr>
<td>Controlling How Special Types of Words Are Spoken (p. 122)</td>
<td>&lt;say-as&gt;</td>
<td>Partial availability</td>
</tr>
<tr>
<td>Identifying SSML-Enhanced Text (p. 113)</td>
<td>&lt;speak&gt;</td>
<td>Full availability</td>
</tr>
<tr>
<td>Pronouncing Acronyms and Abbreviations (p. 124)</td>
<td>&lt;sub&gt;</td>
<td>Full availability</td>
</tr>
<tr>
<td>Improving Pronunciation by Specifying Parts of Speech (p. 125)</td>
<td>&lt;w&gt;</td>
<td>Full availability</td>
</tr>
<tr>
<td>Adding the Sound of Breathing (p. 126)</td>
<td><a href="">amazon:auto-breaths</a></td>
<td>Not available</td>
</tr>
<tr>
<td>Newscaster speaking style (p. 128)</td>
<td>&lt;amazon:domain name=&quot;news&quot;&gt;</td>
<td>Select neural voices only</td>
</tr>
<tr>
<td>Adding Dynamic Range Compression (p. 129)</td>
<td>&lt;amazon:effect name=&quot;drc&quot;&gt;</td>
<td>Full availability</td>
</tr>
<tr>
<td>Speaking Softly (p. 130)</td>
<td>&lt;amazon:effect phonation=&quot;soft&quot;&gt;</td>
<td>Not available</td>
</tr>
<tr>
<td>Controlling Timbre (p. 131)</td>
<td>&lt;amazon:effect vocal-tract-length&gt;</td>
<td>Not available</td>
</tr>
<tr>
<td>Whispering (p. 132)</td>
<td>&lt;amazon: effect name=&quot;whispered&quot;&gt;</td>
<td>Not available</td>
</tr>
</tbody>
</table>

If you use unsupported SSML tags in either neural or standard format, you will get an error.

**Identifying SSML-Enhanced Text**

<say>

The `<say>` tag is the root element of all Amazon Polly SSML text. All SSML-enhanced text must be enclosed within a pair of `<say>` tags.

```
<say>Mary had a little lamb.</say>
```

This tag is supported by both neural and standard TTS formats.

**Adding a Pause**

<break>

This tag is supported by both neural and standard TTS formats.
To add a pause to your text, use the <break> tag. You can set a pause based on strength (equivalent to the pause after a comma, a sentence, or a paragraph), or you can set it to a specific length of time in seconds or milliseconds. If you don't specify an attribute to determine the pause length, Amazon Polly uses the default, which is <break strength="medium" />, which adds a pause the length of a pause after a comma.

**strength attribute values:**

- none: No pause. Use none to remove a normally occurring pause, such as after a period.
- x-weak: Has the same strength as none, no pause.
- weak: Sets a pause of the same duration as the pause after a comma.
- medium: Has the same strength as weak.
- strong: Sets a pause of the same duration as the pause after a sentence.
- x-strong: Sets a pause of the same duration as the pause after a paragraph.

**time attribute values:**

- [number]s: The duration of the pause, in seconds. The maximum duration is 10s.
- [number]ms: The duration of the pause, in milliseconds. The maximum duration is 10000ms.

For example:

```xml
<speak>
  Mary had a little lamb <break time="3s"/>Whose fleece was white as snow.
</speak>
```

If you don't use an attribute with the break tag, the result varies depending on text:

- If there is no other punctuation next to the break tag, it creates a <break strength="medium"/> (comma-length pause).
- If the tag is next to a comma, it upgrades the tag to a <break strength="strong"/> (sentence-length pause).
- If the tag is next to a period, it upgrades the tag to <break strength="x-strong"/> (paragraph-length pause).

### Emphasizing Words

**<emphasis>**

This tag is supported only by the standard TTS format.

To emphasize words, use the <emphasis> tag. Emphasizing words changes the speaking rate and volume. More emphasis makes Amazon Polly speak the text louder and slower. Less emphasis makes it speak quieter and faster. To specify the degree of emphasis, use the **level** attribute.

**level attribute values:**

- Strong: Increases the volume and slows the speaking rate so that the speech is louder and slower.
- Moderate: Increases the volume and slows the speaking rate, but less than strong. Moderate is the default.
- Reduced: Decreases the volume and speeds up the speaking rate. Speech is softer and faster.
Note
The normal speaking rate and volume for a voice falls between the moderate and reduced levels.

For example:

```xml
<speak>
  I already told you I &lt;emphasis level="strong">really like</emphasis> that person.
</speak>
```

## Specifying Another Language for Specific Words

`<lang>`

This tag is supported by both neural and standard TTS formats.

Specify another language for a specific word, phrase, or sentence with the `<lang>` tag. Foreign language words and phrases are generally spoken better when they are enclosed within a pair of `<lang>` tags. To specify the language, use the `xml:lang` attribute. For a complete list of available languages, see Languages Supported by Amazon Polly (p. 18).

Unless you apply the `<lang>` tag, all of the words in the input text are spoken in the language of the voice specified in the `voice-id`. If you apply the `<lang>` tag, the words are spoken in that language.

For example, if the `voice-id` is Joanna (who speaks US English), Amazon Polly speaks the following in the Joanna voice without a French accent:

```xml
<speak>
  Je ne parle pas français.
</speak>
```

If you use the Joanna voice with the `<lang>` tag, Amazon Polly speaks the sentence in the Joanna voice in American-accented French:

```xml
<speak>
  <lang xml:lang="fr-FR">Je ne parle pas français.</lang>.
</speak>
```

Because Joanna is not a native French voice, pronunciation is based on her native language, US English. For example, although perfect French pronunciation features an uvual trill /R/ in the word `français`, Joanna's US English voice pronounces this phoneme as the corresponding sound /r/.

If you use the `voice-id` of Giorgio, who speaks Italian, with the following text, Amazon Polly speaks the sentence in Giorgio's voice with an Italian pronunciation:

```xml
<speak>
  Mi piace Bruce Springsteen.
</speak>
```

If you use the same voice with the following `<lang>` tag, Amazon Polly pronounces Bruce Springsteen in Italian-accented English:
This tag can also be used as a substitute for the optional DefaultLangCode option when synthesizing speech. However, doing so requires that you format your text using SSML.

**Placing a Custom Tag in Your Text**

<mark>

This tag is supported by both neural and standard TTS formats.

To put a custom tag within the text, use the <mark> tag. Amazon Polly takes no action on the tag, but returns the location of the tag in the SSML metadata. This tag can be anything you want to call out, as long as it maintains the following format:

<mark name="tag_name"/>

For example, suppose that the tag name is "animal" and the input text is:

<speak>
    Mary had a little <mark name="animal"/>lamb.
</speak>

Amazon Polly might return the following SSML metadata:

{"time":767,"type":"ssml","start":25,"end":46,"value":"animal"}

**Adding a Pause Between Paragraphs**

<p>

This tag is supported by both neural and standard TTS formats.

To add a pause between paragraphs in your text, use the <p> tag. Using this tag provides a longer pause than native speakers usually place at commas or the end of a sentence. Use the <p> tag to enclose the paragraph:

<speak>
    <p>This is the first paragraph. There should be a pause after this text is spoken.</p>
    <p>This is the second paragraph.</p>
</speak>

This is equivalent to specifying a pause using <break strength="x-strong"/>.

**Using Phonetic Pronunciation**

<phoneme>

This tag is supported by both neural and standard TTS formats.
To make Amazon Polly use phonetic pronunciation for specific text, use the `<phoneme>` tag.

Two attributes are required with the `<phoneme>` tag. They indicate the phonetic alphabet Amazon Polly uses and the phonetic symbols of the corrected pronunciation:

- **alphabet**
  - *ipa*—Indicates that the International Phonetic Alphabet (IPA) will be used.
  - *x-sampa*—Indicates that the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) will be used.
- **ph**
  - Specifies the phonetic symbols for pronunciation. For more information, see *Phoneme and Viseme Tables for Supported Languages* (p. 19)

With the `<phoneme>` tag, Amazon Polly uses the pronunciation specified by the `ph` attribute instead of the standard pronunciation associated by default with the language used by the selected voice.

For instance, the word "pecan" can be pronounced two ways. In the following example, "pecan" is assigned a different pronunciation in each line. Amazon Polly pronounces pecan as specified in the `ph` attributes, instead of using the default pronunciation.

**International Phonetic Alphabet (IPA)**

```
<speak>
  You say, <phoneme alphabet="ipa" ph="p##k##n">pecan</phoneme>.
  I say, <phoneme alphabet="ipa" ph="#pi.kæn">pecan</phoneme>.
</speak>
```

**Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA)**

```
<speak>
  You say, <phoneme alphabet='x-sampa' ph='pI"kA:n'>pecan</phoneme>.
  I say, <phoneme alphabet='x-sampa' ph='"pi.k{n'>pecan</phoneme>.
</speak>
```

Additionally, Mandarin Chinese uses Pinyin for phonetic pronunciation.

**Pinyin**

```
<speak>
  #<phoneme alphabet='x-amazon-pinyin' ph='bo2'>#</phoneme>#
  #<phoneme alphabet='x-amazon-pinyin' ph='bao2'>#</phoneme>#
</speak>
```

---

### Controlling Volume, Speaking Rate, and Pitch

**<prosody>**

Prosody tag attributes are fully supported by the standard TTS voices. Neural voices support the `volume` and `rate` attributes, but don't support the `pitch` attribute.

To control the volume, rate, or pitch of your selected voice, use the `prosody` tag.

Volume, speech rate, and pitch are dependent on the specific voice selected. In addition to differences between voices for different languages, there are differences between individual voices speaking
the same language. Because of this, while attributes are similar across all languages, there are clear variations from language to language and no absolute value is available.

The prosody tag has three attributes, each of which has several available values to set the attribute. Each attribute uses the same syntax:

```
<prosody attribute="value"></prosody>
```

- **volume**
  - default: Resets volume to the default level for the current voice.
  - silent, x-soft, soft, medium, loud, x-loud: Sets the volume to a predefined value for the current voice.
  - +ndB, −ndB: Changes volume relative to the current level. A value of +0dB means no change, +6dB means approximately twice the current volume, and −6dB means approximately half the current volume.

For example, you could set the volume for a passage as follows:

```
<speak>
  Sometimes it can be useful to <prosody volume="loud">increase the volume for a specific speech.</prosody>
</speak>
```

Or you could set it this way:

```
<speak>
  And sometimes a lower volume <prosody volume="-6dB">is a more effective way of interacting with your audience.</prosody>
</speak>
```

- **rate**
  - x-slow, slow, medium, fast, x-fast. Sets the pitch to a predefined value for the selected voice.
  - n%: A non-negative percentage change in the speaking rate. For example, a value of 100% means no change in speaking rate, a value of 200% means a speaking rate twice the default rate, and a value of 50% means a speaking rate of half the default rate. This value has a range of 20-200%.

For example, you could set the speech rate for a passage as follows:

```
<speak>
  For dramatic purposes, you might wish to <prosody rate="slow">slow up the speaking rate of your text.</prosody>
</speak>
```

Or you could set it this way:

```
<speak>
  Although in some cases, it might help your audience to <prosody rate="85%">slow the speaking rate slightly to aid in comprehension.</prosody>
</speak>
```

- **pitch**
  - default: Resets pitch to the default level for the current voice.
  - x-low, low, medium, high, x-high: Sets the pitch to a predefined value for the current voice.
  - +n% or −n%: Adjusts pitch by a relative percentage. For example, a value of +0% means no baseline pitch change, +5% gives a little higher baseline pitch, and −5% results in a little lower baseline pitch.
For example, you could set the pitch for a passage as follows:

```xml
<speak>
  Do you like synthesized speech <prosody pitch="high">with a pitch that is higher than normal</prosody>?
</speak>
```

Or you could set it this way:

```xml
<speak>
  Or do you prefer your speech <prosody pitch="-10%">with a somewhat lower pitch</prosody>?
</speak>
```

The `<prosody>` tag must contain at least one attribute, but can include more within the same tag.

```xml
<speak>
  Each morning when I wake up, <prosody volume="loud" rate="x-slow">I speak quite slowly and deliberately until I have my coffee</prosody>.
</speak>
```

It can also be combined with nested tags, as follows:

```xml
<speak>
  <prosody rate="85%">Sometimes combining attributes <prosody pitch="-10%">can change the impression your audience has of a voice</prosody> as well.</prosody>
</speak>
```

### Setting a Maximum Duration for Synthesized Speech

**<prosody amazon:max-duration>**

This tag is currently supported only by the standard TTS format.

To control how long you want a speech to take when it is synthesized, use the `<prosody>` tag with the `amazon:max-duration` attribute.

The duration of synthesized speech varies slightly, depending on the voice you select. This can make it difficult to match synthesized speech with visuals or other activities that require precise timing. This issue is magnified for translation applications because the time it takes to say particular phrases can vary widely with different languages.

The `<prosody amazon:max-duration>` tag matches synthesized speech to the amount of time you want it to take (the duration).

This tag uses the following syntax:

```xml
<prosody amazon:max-duration="time duration">
```

With the `<prosody amazon:max-duration>` tag, you can specify duration in either seconds or milliseconds:

- `ns`: the maximum duration in seconds
• **ms**: the maximum duration in milliseconds

For example, the following spoken text has a maximum duration of 2 seconds:

```xml
<speak>
  <prosody amazon:max-duration="2s">
    Human speech is a powerful way to communicate.
  </prosody>
</speak>
```

Text placed within the tag, it doesn't exceed the specified duration. If the chosen voice or language would normally take longer than that duration, Amazon Polly speeds up the speech so that it fits into the specified duration.

If the specified duration is longer than it takes to read the text at a normal rate, Amazon Polly reads the speech normally. It doesn't slow down the speech or add silence, so the resulting audio is shorter than requested.

**Note**

Amazon Polly increases the speed no more than 5 times the normal rate. If text is spoken faster than this, it usually doesn't make sense. If a speech cannot fit within your specified duration even when speeded up to the maximum, the audio will be speeded up but will last longer than the specified duration.

You can include a single sentence or multiple sentences within a `<prosody amazon:max-duration>` tag, and you can use multiple `<prosody amazon:max-duration>` tags within your text.

For example:

```xml
<speak>
  <prosody amazon:max-duration="2400ms">
    Human speech is a powerful way to communicate.
  </prosody>
  <break strength="strong"/>
  <prosody amazon:max-duration="5100ms">
    Even a simple ‘Hello’ can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
  <break strength="strong"/>
  <prosody amazon:max-duration="8900ms">
    We naturally understand this information, which is why speech is ideal for creating applications where a screen isn’t practical or possible, or simply isn’t convenient.
  </prosody>
</speak>
```

Using the `<prosody amazon:max-duration>` tag can increase latency when Amazon Polly is returns synthesized speech. The degree of latency depends on the passage and its length. We recommend using text comprised of relatively short text passages.

**Limitations**

There are limitations both in how you use `<prosody amazon:max-duration>` tag and in how it works with other SSML tags:

- The text inside a `<prosody amazon:max-duration>` tag can't be longer than 1500 characters.
• You can't nest `<prosody amazon:max-duration>` tags. If you put one `<prosody amazon:max-duration>` tag inside another, Amazon Polly ignores the inner tag.

For example, in the following, the `<prosody amazon:max-duration="5s">` tag is ignored:

```xml
<speak>
  <prosody amazon:max-duration="16s">
    Human speech is a powerful way to communicate.
  </prosody>
  <prosody amazon:max-duration="5s">
    Even a simple ‘Hello’ can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
  We naturally understand this information, which is why speech is ideal for creating applications where a screen isn’t practical or possible, or simply isn’t convenient.
</speak>
```

• You can't use the `<prosody>` tags with the `rate` attribute within a `<prosody amazon:max-duration>` tag. This is because both affect the speed at which text is spoken.

In the following example, Amazon Polly ignores the `<prosody rate="2">` tag:

```xml
<speak>
  <prosody amazon:max-duration="7500ms">
    Human speech is a powerful way to communicate.
  </prosody>
  <prosody rate="2">
    Even a simple ‘Hello’ can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
</speak>
```

**Pauses and max-duration**

When using `max-duration` tag, you can still insert pauses within your text. However, Amazon Polly includes the length of the pause when calculating the maximum duration for speech. Additionally, Amazon Polly preserves the short pauses that occur where commas and periods are placed within a passage and includes in the maximum duration.

For example, in the following block, the 600 millisecond break and the breaks caused by the commas and periods occur within the 8-second speech:

```xml
<speak>
  <prosody amazon:max-duration="8s">
    Human speech is a powerful way to communicate.
    <break time="600ms"/>
    Even a simple ‘Hello’ can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
</speak>
```

**Adding a Pause Between Sentences**

<s>
This tag is supported by both neural and standard TTS formats.

To add a pause between lines or sentences in your text, use the `<s>` tag. Using this tag has the same effect as:

- Ending a sentence with a period (.)
- Specifying a pause with `<break strength="strong"/>`

Unlike the `<break>` tag, the `<s>` tag encloses the sentence. This is useful for synthesizing speech that is organized in lines, rather than sentence, such as poetry.

In the following example, the `<s>` tag creates a short pause after both the first and second sentences. The final sentence has no `<s>` tag, but it is also followed by a short pause because it ends with a period.

```
<speak>
<s>Mary had a little lamb</s>
<s>Whose fleece was white as snow</s>
   And everywhere that Mary went, the lamb was sure to go.
</speak>
```

### Controlling How Special Types of Words Are Spoken

**<say-as>**

Except for the "characters" or "spell-out" feature, this tag is supported by both neural and standard TTS formats. If this SSML code is encountered by Amazon Polly at run-time, the affected sentence will be synthesized using the related standard voice. Please note, however, that this sentence will still be billed as if it uses the neural voice.

Use the `<say-as>` tag with the `interpret-as` attribute to tell Amazon Polly how to say certain characters, words, and numbers. This enables you to provide additional context to eliminate any ambiguity on how Amazon Polly should render the text.

The `<say-as>` tag uses one attribute, `<interpret-as>`, which uses a number of possible available values. Each uses the same syntax:

```
<say-as interpret-as="value">[text to be interpreted]</say-as>
```

The following values are available with `interpret-as`:

- **characters** or **spell-out**: Spells out each letter of the text, as in a-b-c.

  **Note**
  
  This option is not currently supported using neural voices. If this SSML code is encountered by Amazon Polly at run-time, the affected sentence will be synthesized using the related standard voice. Please note, however, that this sentence will still be billed as if it uses the neural voice.

- **cardinal** or **number**: Interprets the numerical text as a cardinal number, as in 1,234.
- **ordinal**: Interprets the numerical text as an ordinal number, as in 1,234th.
- **digits**: Spells out each digit individually, as in 1-2-3-4.
- **fraction**: Interprets the numerical text as a fraction. This works for both common fractions such as 3/20, and mixed fractions, such as 2 ½. See below for more information.
- **unit**: Interprets a numerical text as a measurement. The value should be either a number or a fraction followed by a unit with no space in between as in 1/2 inch, or by just a unit, as in 1 meter.
Controlling How Special Types of Words Are Spoken

- **date**: Interprets the text as a date. The format of the date must be specified with the format attribute. See below for more information.
- **time**: Interprets the numerical text as duration, in minutes and seconds, as in 1'21".
- **address**: Interprets the text as part of a street address.
- **expletive**: "Beeps out" the content included within the tag.
- **telephone**: Interprets the numerical text as a 7-digit or 10-digit telephone number, as in 2025551212. You can also use this value for handle telephone extensions, as in 2025551212x345. See below for more information.

**Note**
Currently the telephone option is not available for all languages. However, it is available for voices speaking English language variants (en-AU, en-GB, en-IN, en-US, and en-GB-WLS), Spanish language variants (es-ES, es-MX, and es-US), French language variants (fr-FR and fr-CA), and Portuguese variants (pt-BR and pt-PT), as well as German (de-DE), Italian (it-IT), Japanese (ja-JP), and Russian (ru-RU). It should also be noted that in some cases, languages such as Arabic (arb) automatically handle the number set as a telephone number and so do not actually implement the telephone SSML tag.

**Fractions**
Amazon Polly interprets values within the `say-as` tag that have the interpret-as="fraction" attribute as common fractions. The following is the syntax for fractions:

- **Fraction**
  
  Syntax: `cardinal number/cardinal number`, such as 2/9.

  For example: `<say-as interpret-as="fraction">2/9</say-as>` is pronounced "two ninths."
  
- **Non-negative Mixed Number**

  Syntax: `cardinal number+cardinal number/cardinal number`, such as 3+1/2.

  For example, `<say-as interpret-as="fraction">3+1/2</say-as>` is pronounced "three and a half."

  **Note**
  There must be a + between the "3" and the "1/2". Amazon Polly doesn't support a mixed number without the +, such as "3 1/2".

**Dates**
When interpret-as is set to date, you also need to indicate the format of the date.

This uses the following syntax:

```
<say-as interpret-as="date" format="format">[date]</say-as>
```

For example:

```
<speak>
  I was born on <say-as interpret-as="date" format="mdy">12-31-1900</say-as>.
</speak>
```

The following formats can be used with the date attribute:

- **mdy**: Month-day-year.
• dmy: Day-month-year.
• ymd: Year-month-day.
• md: Month-day.
• dm: Day-month.
• ym: Year-month.
• my: Month-year.
• d: Day.
• m: Month.
• y: Year.
• yyyyymmdd: Year-month-day. If you use this format, you can make Amazon Polly skip parts of the date using question marks.

For example, Amazon Polly renders the following as "September 22nd":

```
<say-as interpret-as="date">????0922</say-as>
```

Format is not needed.

**Telephone**

Amazon Polly attempts to interpret the text you provide correctly based on the text's formatting even without the <say-as> tag. For example, if your text includes "202-555-1212," Amazon Polly interprets it as a 10-digit telephone number and says each digit individually, with a brief pause for each dash. In this case, you don't need to use <say-as interpret-as="telephone">. However, if you provide the text "2025551212" and want Amazon Polly to say it as a phone number, you would specify <say-as interpret-as="telephone">.

The logic for interpreting each element is language-specific. For example, US and UK English differ in how phone numbers are pronounced (in UK English, sequences of the same digit are grouped together, as in "double five" or "triple four"). To see the difference, test the following example with a US voice and with a UK voice:

```
<speak>
    Richard's number is <say-as interpret-as="telephone">2122241555</say-as>
</speak>
```

**Pronouncing Acronyms and Abbreviations**

<sub>

This tag is supported by both neural and standard TTS formats.

Use the <sub> tag with the alias attribute to substitute a different word (or pronunciation) for selected text such as an acronym or abbreviation.

This uses the syntax:

```
<sub alias="new word">abbreviation</sub>
```

In the following example, the name "Mercury" is substituted for the element's chemical symbol to make the audio content clearer.
Improving Pronunciation by Specifying Parts of Speech

This tag is supported by both neural and standard TTS formats.

You can use the `<w>` tag to customize the pronunciation of words by specifying the word's part of speech or alternate meaning. This is done using the `role` attribute.

This tag uses the following syntax:

```xml
<w role="attribute">text</w>
```

The following values can be used for the `role` attribute:

To specify the part of speech:

- `amazon:VB`: interprets the word as a verb (present simple).
- `amazon:VBD`: interprets the word as past tense verb.
- `amazon:DT`: interprets the word as a determiner.
- `amazon:IN`: interprets the word as a proposition.
- `amazon:JJ`: interprets the word as an adjective.
- `amazon:NN`: interprets the word as a noun.

For example, depending on its part of speech, the US English pronunciation of the word "read" varies based on the tag:

```xml
<speak>
  The word <say-as interpret-as="characters">read</say-as> may be interpreted as either the present simple form <w role="amazon:VB">read</w>, or the past participle form <w role="amazon:VBD">read</w>.
</speak>
```

To specify a specific meaning:

- `amazon:DEFAULT`: uses the default sense of the word.
- `amazon:SENSE_1`: uses the non-default sense of the word when present. For example, the noun "bass" is pronounced differently depending on its meaning. The default meaning is the lowest part of the musical range. The alternate meaning is a species of freshwater fish, also called "bass" but pronounced differently. Using `<w role="amazon:SENSE_1">bass</w>` renders the non-default pronunciation (freshwater fish) for the audio text.

This difference in pronunciation and meaning can be heard if you synthesize the following:

```xml
<speak>
</speak>
```
Depending on your meaning, the word `<say-as interpret-as="characters">bass</say-as>` may be interpreted as either a musical element: bass, or as its alternative meaning, a freshwater fish `<w role="amazon:SENSE_1">bass</w>`.

</speak>

**Note**
Some languages may have a different selection of supported parts of speech.

## Adding the Sound of Breathing

<amazon:breath> and <amazon:auto-breaths>

This tag is supported only by the standard TTS format.

Natural-sounding speech includes both correctly spoken words and breathing sounds. By adding breathing sounds to synthesized speech, you can make it sound more natural. The `<amazon:breath>` and `<amazon:auto-breaths>` tags provide breaths. You have the following options:

- Manual mode: you set the location, length, and volume of a breath sound within the text
- Automated mode: Amazon Polly automatically inserts breathing sounds into the speech output
- Mixed mode: both you and Amazon Polly add breathing sounds

### Manual Mode

In manual mode, you place the `<amazon:breath/>` tag in the input text where you want to locate a breath. You can customize the length and volume of breaths with the `duration` and `volume` attributes, respectively:

- **duration**: Controls the length of the breath. Valid values are: default, x-short, short, medium, long, x-long. The default value is medium.
- **volume**: Controls how loud breathing sounds. Valid values are: default, x-soft, soft, medium, loud, x-loud. The default value is medium.

**Note**
The exact length and volume of each attribute value is dependent on the specific Amazon Polly voice used.

To set a breath sound using the defaults, use `<amazon:breath/>` without attributes.

For example, to use attributes to set the duration and volume for a breath to medium, you would set the attributes as follows:

```xml
<speak>
  Sometimes you want to insert only <amazon:breath duration="medium" volume="x-loud"/> a single breath.
</speak>
```

To use the defaults, you would just use the tag:

```xml
<speak>
  Sometimes you need <amazon:breath/> to insert one or more average breathes <amazon:breath/> so that the text sounds correct.
</speak>
```
You can add individual breathing sounds within a passage, as follows:

```xml
<speak>
  <amazon:breath duration="long" volume="x-loud"/>
  <prosody rate="120%">
    Wow! <amazon:breath duration="long" volume="loud"/>
    That was quite fast
  </prosody>
  <amazon:breath duration="medium" volume="x-loud"/>
  I almost beat my personal best time on this track.
</speak>
```

Automated Mode

In automated mode, you use the `<amazon:auto-breaths>` tag to tell Amazon Polly to automatically create breathing noises at appropriate intervals. You can set the frequency of the intervals, their volume, and their duration. Place the `<amazon:auto-breaths>` tag at the beginning of the text that you want to apply automated breathing to and then close the tag at the end.

**Note**

Unlike the manual mode tag, `<amazon:breath/>`, the `<amazon:auto-breaths>` tag requires a closing tag (`</amazon:auto-breaths>`).

You can use the following optional attributes with the `<amazon:auto-breaths>` tag:

- **volume**: Controls how loud the breathing sounds. Valid values are: default, x-soft, soft, medium, loud, x-loud. The default value is medium.

- **frequency**: Controls how often breathing sounds occur in the text. Valid values are: default, x-low, low, medium, high, x-high. The default value is medium.

- **duration**: Controls the length of the breath. Valid values are: default, x-short, short, medium, long, x-long. The default value is medium.

By default, the frequency of breathing sounds depends on the input text. However, breathing sounds often occur after commas and periods.

The following examples show how to use the `<amazon:auto-breaths>` tag. To decide which options to use for your content, copy the applicable examples to the Amazon Polly console and listen to the differences.

- Using automated mode without optional parameters.

```xml
<speak>
  <amazon:auto-breaths>Amazon Polly is a service that turns text into lifelike speech, allowing you to create applications that talk and build entirely new categories of speech-enabled products. Amazon Polly is a text-to-speech service that uses advanced deep learning technologies to synthesize speech that sounds like a human voice. With dozens of lifelike voices across a variety of languages, you can select the ideal voice and build speech-enabled applications that work in many different countries.</amazon:auto-breaths>
</speak>
```

- Using automated mode with volume control. The unspecified parameters (duration and frequency) are set to the default values (medium).

```xml
<speak>
  <amazon:auto-breaths volume="x-soft">Amazon Polly is a service that turns text into lifelike
```
speech, allowing you to create applications that talk and build entirely new
categories of
speech-enabled products. Amazon Polly is a text-to-speech service, that uses
advanced deep
learning technologies to synthesize speech that sounds like a human voice. With
dozens of
lifelike voices across a variety of languages, you can select the ideal voice and
build speech-
enabled applications that work in many different countries.</amazon:auto-breaths>
</speak>

• Using automated mode with frequency control. The unspecified parameters (duration and volume)
are set to the default values (medium).

<speak>
<amazon:auto-breaths frequency="x-low">Amazon Polly is a service that turns text
into lifelike
speech, allowing you to create applications that talk and build entirely new
categories of
speech-enabled products. Amazon Polly is a text-to-speech service, that uses
advanced deep
learning technologies to synthesize speech that sounds like a human voice. With
dozens of
lifelike voices across a variety of languages, you can select the ideal voice and
build speech-
enabled applications that work in many different countries.</amazon:auto-breaths>
</speak>

• Using automated mode with multiple parameters. For the unspecified Duration parameter, Amazon
Polly uses the default value (medium).

<speak>
<amazon:auto-breaths volume="x-loud" frequency="x-low">Amazon Polly is a service that turns
text into lifelike speech, allowing you to create applications that talk and build
entirely new
categories of speech-enabled products. Amazon Polly is a text-to-speech service,
that uses
advanced deep learning technologies to synthesize speech that sounds like a human
voice. With
dozens of lifelike voices across a variety of languages, you can select the ideal
voice and build
speech-enabled applications that work in many different countries.</amazon:auto-breaths>
</speak>

Newscaster speaking style

<amazon:domain name="news"/>

The newscaster style is available only for the Matthew or Joanna voices, which are available only in
American English (en-US), Lupe, in US Spanish (es-US) and Amy, in British English (en-GB). It is only
supported when using Neural format.

To use the newscaster style, you use SSML tags and the following syntax:

<amazon:domain name="news">text</amazon:domain>

For example, you might use the newscaster style with the Amy voice as follows:
Adding Dynamic Range Compression

This tag is supported by both neural and standard TTS formats.

Depending on the text, language, and voice used in an audio file, the sounds range from soft to loud. Environmental sounds, such as the sound of a moving vehicle, can often mask the softer sounds, which makes the audio track difficult to hear clearly. To enhance the volume of certain sounds in your audio file, use the dynamic range compression (drc) tag.

The drc tag sets a midrange "loudness" threshold for your audio, and increases the volume (the gain) of the sounds around that threshold. It applies the greatest gain increase closest to the threshold, and the gain increase is lessened farther away from the threshold.

This makes the middle-range sounds easier to hear in a noisy environment, which makes the entire audio file clearer.

The drc tag is a Boolean parameter (it's either present or it isn't). It uses the syntax: <amazon:effect name="drc"> and is closed with </amazon:effect>.

You can use the drc tag with any voice or language supported by Amazon Polly. You can apply it to an entire section of the recording, or for only a few words. For example:

Some audio is difficult to hear in a moving vehicle, but <amazon:effect name="drc"> this audio is less difficult to hear in a moving vehicle. </amazon:effect>

Note
When you use "drc" in the amazon:effect syntax, it is case-sensitive.

Using drc with the prosody volume Tag
As the following graphic shows, the prosody volume tag evenly increases the volume of an entire audio file from the original level (dotted line) to an adjusted level (solid line). To further increase the volume of certain parts of the file, use the drc tag with the prosody volume tag. Combining tags doesn't affect the settings of the prosody volume tag.

When you use the drc and prosody volume tags together, Amazon Polly applies the drc tag first, increasing the middle-range sounds (those near the threshold). It then applies the prosody volume tag and further increases the volume of the entire audio track evenly.

To use the tags together, nest one inside the other. For example:

```xml
<speak>
  <prosody volume="loud">This text needs to be understandable and loud. <amazon:effect name="drc">
    This text also needs to be more understandable in a moving car.</amazon:effect></prosody>
</speak>
```

In this text, the prosody volume tag increases the volume of the entire passage to "loud." The drc tag enhances the volume of the middle-range values in the second sentence.

**Note**

When using the drc and prosody volume tags together, use standard XML practices for nesting tags.

### Speaking Softly

```xml
<amazon:effect phonation="soft"/>
```

This tag is currently supported only by the standard TTS format.

To specify that input text should be spoken in a softer-than-normal voice, use the <amazon:effect phonation="soft"> tag.

This uses the syntax:

```xml
<amazon:effect phonation="soft">text</amazon:effect>
```

For example, you might use this tag with the Matthew voice as follows:
<speak>
This is Matthew speaking in my normal voice. <amazon:effect phonation="soft">This is Matthew speaking in my softer voice.</amazon:effect>
</speak>

Controlling Timbre

<amazon:effect vocal-tract-length>
This tag is currently supported only by the standard TTS format.

Timbre is the tonal quality of a voice that helps you tell the difference between voices, even when they have the same pitch and loudness. One of the most important physiological features that contributes to speech timbre is the length of the vocal tract. The vocal tract is a cavity of air that spans from the top of the vocal folds up to the edge of the lips.

To control the timbre of output speech in Amazon Polly, use the vocal-tract-length tag. This tag has the effect of changing the length of the speaker's vocal tract, which sounds like a change in the speaker's size. When you increase the vocal-tract-length, the speaker sounds physically bigger. When you decrease it, the speaker sounds smaller. You can use this tag with any of the voices in the Amazon Polly Text-to-Speech portfolio.

To change timbre, use the following values:

- +n% or -n%: Adjusts the vocal tract length by a relative percentage change in the current voice. For example, +4% or -2%. Valid values range from +100% to -50%. Values outside this range are clipped. For example, +111% sounds like +100% and -60% sounds like -50%.
- n%: Changes the vocal tract length to an absolute percentage of the tract length of the current voice. For example, 110% or 75%. An absolute value of 110% is equivalent to a relative value of +10%. An absolute value of 100% is the same as the default value for the current voice.

The following example shows how to change the vocal tract length to change timbre:

<speak>
This is my original voice, without any modifications. <amazon:effect vocal-tract-length="+15%"
    > Now, imagine that I am much bigger. </amazon:effect> <amazon:effect vocal-tract-length="-15%"
    > Or, perhaps you prefer my voice when I'm very small. </amazon:effect> You can also control the timbre of my voice by making minor adjustments. <amazon:effect vocal-tract-length="+10%"
    > For example, by making me sound just a little bigger. </amazon:effect><amazon:effect vocal-tract-length="-10%">
    Or, making me sound only somewhat smaller. </amazon:effect>
</speak>

Combining Multiple Tags

You can combine the vocal-tract-length tag with any other SSML tag that is supported by Amazon Polly. Because timbre (vocal tract length) and pitch are closely connected, you might get the best results by using both the vocal-tract-length and the <prosody pitch> tags. To produce the most realistic voice, we recommend that you use different percentages of change for the two tags. Experiment with various combinations to get the results you want.

The following example shows how to combine tags.
<speak>
The pitch and timbre of a person's voice are connected in human speech.
</speak>

<amazon:effect vocal-tract-length="-15%"> If you are going to reduce the vocal tract length,
</amazon:effect><amazon:effect vocal-tract-length="-15%"> <prosody pitch="+20%"> you might consider increasing the pitch, too. </prosody></amazon:effect>
<amazon:effect vocal-tract-length="+15%"> If you choose to lengthen the vocal tract,
</amazon:effect> <amazon:effect vocal-tract-length="+15%"> <prosody pitch="-10%"> you might also want to lower the pitch. </prosody></amazon:effect>

Whispering

<amazon:effect name="whispered">
This tag is currently supported only by the standard TTS format.

This tag indicates that the input text should be spoken in a whispered voice rather than as normal speech. This can be used with any of the voices in the Amazon Polly Text-to-Speech portfolio.

This uses the following syntax:

<amazon:effect name="whispered">text</amazon:effect>

For example:

<speak>
<amazon:effect name="whispered">If you make any noise, </amazon:effect>
she said, <amazon:effect name="whispered">they will hear us.</amazon:effect>
</speak>

In this case, the synthesized speech spoken by the character is whispered, but the phrase "she said" is spoken in the normal synthesized speech of the selected Amazon Polly voice.

You can enhance the "whispered" effect by slowing down the prosody rate by up to 10%, depending on the effect you want.

For example:

<speak>
When any voice is made to whisper, <amazon:effect name="whispered">
<prosody rate="-10%">the sound is slower and quieter than normal speech
</prosody></amazon:effect>
</speak>

When generating speech marks for a whispered voice, the audio stream must also include the whispered voice to ensure that the speech marks match the audio stream.
Managing Lexicons

Pronunciation lexicons enable you to customize the pronunciation of words. Amazon Polly provides API operations that you can use to store lexicons in an AWS region. Those lexicons are then specific to that particular region. You can use one or more of the lexicons from that region when synthesizing the text by using the `SynthesizeSpeech` operation. This applies the specified lexicon to the input text before the synthesis begins. For more information, see `SynthesizeSpeech` (p. 270).

**Note**
These lexicons must conform with the Pronunciation Lexicon Specification (PLS) W3C recommendation. For more information, see Pronunciation Lexicon Specification (PLS) Version 1.0 on the W3C website.

The following are examples of ways to use lexicons with speech synthesis engines:

- Common words are sometimes stylized with numbers taking the place of letters, as with "g3t sm4rt" (get smart). Humans can read these words correctly. However, a Text-to-Speech (TTS) engine reads the text literally, pronouncing the name exactly as it is spelled. This is where you can leverage lexicons to customize the synthesized speech by using Amazon Polly. In this example, you can specify an alias (get smart) for the word "g3t sm4rt" in the lexicon.
- Your text might include an acronym, such as W3C. You can use a lexicon to define an alias for the word W3C so that it is read in the full, expanded form (World Wide Web Consortium).

Lexicons give you additional control over how Amazon Polly pronounces words uncommon to the selected language. For example, you can specify the pronunciation using a phonetic alphabet. For more information, see Pronunciation Lexicon Specification (PLS) Version 1.0 on the W3C website.

**Topics**
- Applying Multiple Lexicons (p. 133)
- Managing Lexicons Using the Amazon Polly Console (p. 134)
- Managing Lexicons Using the AWS CLI (p. 137)

Applying Multiple Lexicons

You can apply up to five lexicons to your text. If the same grapheme appears in more than one lexicon that you apply to your text, the order in which they are applied can make a difference in the resulting speech. For example, given the following text, "Hello, my name is Bob." and two lexemes in different lexicons that both use the grapheme Bob.

**LexA**

```
<lexeme>
  <grapheme>Bob</grapheme>
  <alias>Robert</alias>
</lexeme>
```

**LexB**

```
<lexeme>
```

133
<grapheme>Bob</grapheme> <alias>Bobby</alias> </lexeme>

If the lexicons are listed in the order LexA and then LexB, the synthesized speech will be "Hello, my name is Robert." If they are listed in the order LexB and then LexA, the synthesized speech is "Hello, my name is Bobby."

**Example – Applying LexA Before LexB**

```bash
aws polly synthesize-speech \
--lexicon-names LexA LexB \
--output-format mp3 \
--text 'Hello, my name is Bob' \
--voice-id Justin \
bobAB.mp3
```

**Speech output:** "Hello, my name is Robert."

**Example – Applying LexB before LexA**

```bash
aws polly synthesize-speech \
--lexicon-names LexB LexA \
--output-format mp3 \
--text 'Hello, my name is Bob' \
--voice-id Justin \
bobBA.mp3
```

**Speech output:** "Hello, my name is Bobby."

For information about applying lexicons using the Amazon Polly console, see [Applying Lexicons Using the Console (Synthesize Speech)](p. 135).

---

### Managing Lexicons Using the Amazon Polly Console

You can use the Amazon Polly console to upload, download, apply, filter, and delete lexicons. The following procedures demonstrate each of these processes.

#### Uploading Lexicons Using the Console

To use a pronunciation lexicon, you must first upload it. There are two locations on the console from which you can upload a lexicon, the **Text-to-Speech** tab and the **Lexicons** tab.

The following processes describe how to add lexicons that you can use to customize how words and phrases uncommon to the chosen language are pronounced.

**To add a lexicon from the Lexicons tab**

1. Sign in to the AWS Management Console and open the Amazon Polly console at [https://console.aws.amazon.com/polly/](https://console.aws.amazon.com/polly/).
2. Choose the **Lexicons** tab.
3. Choose Upload lexicon.
4. Provide a name for the lexicon and then use Choose a lexicon file to find the lexicon to upload. You can only upload PLS files with .pls or .xml extensions.
5. Choose Upload lexicon. If a lexicon by the same name (whether a .pls or .xml file) already exists, uploading the lexicon overwrites the existing lexicon.

To add a lexicon from the Text-to-Speech tab

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Text-to-Speech tab.
3. Expand Additional settings, turn on Customize pronunciation, and then choose Upload lexicon.
4. Provide a name for the lexicon and then use Choose a lexicon file to find the lexicon to upload. You can only use PLS files with .pls or .xml extensions.
5. Choose Upload lexicon. If a lexicon with the same name (whether a .pls or .xml file) already exists, uploading the lexicon overwrites the existing lexicon.

Applying Lexicons Using the Console (Synthesize Speech)

The following procedure demonstrates how to apply a lexicon to your input text by applying the W3c.pls lexicon to substitute "World Wide Web Consortium" for "W3C". If you apply multiple lexicons to your text they are applied in a top-down order with the first match taking precedence over later matches. A lexicon is applied to the text only if the language specified in the lexicon is the same as the language chosen.

You can apply a lexicon to plain text or SSML input.

Example – Applying the W3C.pls Lexicon

To create the lexicon you'll need for this exercise, see Using the PutLexicon Operation (p. 137). Use a plain text editor to create the W3C.pls lexicon shown at the top of the topic. Remember where you save this file.

To apply the W3C.pls lexicon to your input

In this example we introduce a lexicon to substitute "World Wide Web Consortium" for "W3C". Compare the results of this exercise with that of Using SSML (Console) (p. 107) for both US English and another language.

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Do one of the following:
   • Turn off SSML and then type or paste this text into the text input box.
     
     He was caught up in the game.  
     In the middle of the 10/3/2014 W3C meeting he shouted, "Score!" quite loudly.

   • Turn on SSML and then type or paste this text into the text input box.
     
     <speak>He wasn't paying attention.<break time="1s"/></speak>
In the middle of the 10/3/2014 W3C meeting he shouted, “Score!” quite loudly.<speak>

3. From the Language list, choose English, US, then choose the voice you want to use for this text.
4. Expand Additional settings and turn on Customize pronunciation.
5. From the list of lexicons, choose W3C (English, US).

   If the W3C (English, US) lexicon is not listed, choose Upload lexicon and upload it, then choose it from the list. To create this lexicon, see Using the PutLexicon Operation (p. 137).
6. To listen to the speech immediately, choose Listen.
7. To save the speech to a file,
   a. Choose Download.
   b. To change to a different file format, turn on Speech file format settings, choose the file format you want, and then choose Download.

Repeat the previous steps, but choose a different language and notice the difference in the output.

Filtering the Lexicon List Using the Console

The following procedure describes how to filter the lexicons list so that only lexicons of a chosen language are displayed.

To filter the lexicons listed by language
1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Lexicons tab.
3. Choose Any language.
4. From the list of languages, choose the language you want to filter on.

   The list displays only the lexicons for the chosen language.

Downloading Lexicons Using the Console

The following process describes how to download one or more lexicons. You can add, remove, or modify lexicon entries in the file and then upload it again to keep your lexicon up-to-date.

To download one or more lexicons
1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Lexicons tab.
3. Choose the lexicon or lexicons you want to download.
   a. To download a single lexicon, choose its name from the list.
   b. To download multiple lexicons as a single compressed archive file, select the check box next to each entry in the list that you want to download.
4. Choose Download.
5. Open the folder where you want to download the lexicon.
6. Choose Save.
Deleting a Lexicon Using the Console

To delete a lexicon

The following process describes how to delete a lexicon. After deleting the lexicon, you must add it back before you can use it again. You can delete one or more lexicons at the same time by selecting the check boxes next to individual lexicons.

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Lexicons tab.
3. Choose one or more lexicons that you want to delete from the list.
4. Choose Delete.
5. Enter confirmation text and then choose Delete to remove the lexicon from the Region or Cancel to keep it.

Managing Lexicons Using the AWS CLI

The following topics cover the AWS CLI commands needed to manage your pronunciation lexicons.

Topics
- Using the PutLexicon Operation (p. 137)
- Using the GetLexicon Operation (p. 142)
- Using the ListLexicons Operations (p. 142)
- Using the DeleteLexicon Operation (p. 143)

Using the PutLexicon Operation

With Amazon Polly, you can use PutLexicon (p. 262) to store pronunciation lexicons in a specific AWS Region for your account. Then, you can specify one or more of these stored lexicons in your SynthesizeSpeech (p. 270) request that you want to apply before the service starts synthesizing the text. For more information, see Managing Lexicons (p. 133).

This section provides example lexicons and step-by-step instructions for storing and testing them.

Note

These lexicons must conform to the Pronunciation Lexicon Specification (PLS) W3C recommendation. For more information, see Pronunciation Lexicon Specification (PLS) Version 1.0 on the W3C website.

Example 1: Lexicon with One Lexeme

Consider the following W3C PLS-compliant lexicon.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
    xmlns=http://www.w3.org/2005/01/pronunciation-lexicon
    xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
    xsi:schemaLocation=http://www.w3.org/2005/01/pronunciation-lexicon
                     http://www.w3.org/TR/2007/CR-pronunciation-lexicon-20071212/pls.xsd"
    alphabet="ipa"
```
Note the following:

- The two attributes specified in the `<lexicon>` element:
  - The `xml:lang` attribute specifies the language code, `en-US`, to which the lexicon applies. Amazon Polly can use this example lexicon if the voice you specify in the `SynthesizeSpeech` call has the same language code (en-US).
    
    **Note**
    You can use the `DescribeVoices` operation to find the language code associated with a voice.

  - The `alphabet` attribute specifies IPA, which means that the International Phonetic Alphabet (IPA) alphabet is used for pronunciations. IPA is one of the alphabets for writing pronunciations. Amazon Polly also supports the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA).

- The `<lexeme>` element describes the mapping between `<grapheme>` (that is, a textual representation of the word) and `<alias>`.

To test this lexicon, do the following:

1. Save the lexicon as `example.pls`.
2. Run the `put-lexicon` AWS CLI command to store the lexicon (with the name `w3c`), in the `us-east-2` region.
   
   ```bash
   aws polly put-lexicon \
   --name w3c \
   --content file://example.pls
   ```
3. Run the `synthesize-speech` command to synthesize sample text to an audio stream (speech.mp3), and specify the optional `lexicon-name` parameter.
   
   ```bash
   aws polly synthesize-speech \
   --text 'W3C is a Consortium' \
   --voice-id Joanna \
   --output-format mp3 \
   --lexicon-names="w3c" \
   speech.mp3
   ```
4. Play the resulting `speech.mp3`, and notice that the word W3C in the text is replaced by World Wide Web Consortium.

The preceding example lexicon uses an alias. The IPA alphabet mentioned in the lexicon is not used. The following lexicon specifies a phonetic pronunciation using the `<phoneme>` element with the IPA alphabet.
Follow the same steps to test this lexicon. Make sure you specify input text that has word “pecan” (for example, “Pecan pie is delicious”).

Example 2: Lexicon with Multiple Lexemes

In this example, the lexeme that you specify in the lexicon applies exclusively to the input text for the synthesis. Consider the following lexicon:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
    xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.w3.org/2005/01/pronunciation-lexicon
    http://www.w3.org/TR/2007/CR-pronunciation-lexicon-20071212/pls.xsd"
    alphabet="ipa" xml:lang="en-US">
    <lexeme>
        <grapheme>W3C</grapheme>
        <alias>World Wide Web Consortium</alias>
    </lexeme>
    <lexeme>
        <grapheme>W3C</grapheme>
        <alias>WWW Consortium</alias>
    </lexeme>
    <lexeme>
        <grapheme>Consortium</grapheme>
        <alias>Community</alias>
    </lexeme>
</lexicon>
```

The lexicon specifies three lexemes, two of which define an alias for the grapheme W3C as follows:

- The second `<lexeme>` defines an alternative alias (WWW Consortium).

Amazon Polly uses the first replacement for any given grapheme in a lexicon.

The third `<lexeme>` defines a replacement (Community) for the word Consortium.

First, let’s test this lexicon. Suppose you want to synthesize the following sample text to an audio file (speech.mp3), and you specify the lexicon in a call to `SynthesizeSpeech`.

```
The W3C is a Consortium
```

`SynthesizeSpeech` first applies the lexicon as follows:

- As per the first lexeme, the word W3C is revised as World Wide Web Consortium. The revised text appears as follows:
The World Wide Web Consortium is a Consortium

• The alias defined in the third lexeme applies only to the word Consortium that was part of the original text, resulting in the following text:

The World Wide Web Consortium is a Community.

You can test this using the AWS CLI as follows:

1. Save the lexicon as example.pls.
2. Run the put-lexicon command to store the lexicon with name w3c in the us-east-2 region.

```bash
aws polly put-lexicon \
--name w3c \
--content file://example.pls
```

3. Run the list-lexicons command to verify that the w3c lexicon is in the list of lexicons returned.

```bash
aws polly list-lexicons
```

4. Run the synthesize-speech command to synthesize sample text to an audio file (speech.mp3), and specify the optional lexicon-name parameter.

```bash
aws polly synthesize-speech \
--text 'W3C is a Consortium' \
--voice-id Joanna \
--output-format mp3 \
--lexicon-names="w3c" \
speech.mp3
```

5. Play the resulting speech.mp3 file to verify that the synthesized speech reflects the text changes.

**Example 3: Specifying Multiple Lexicons**

In a call to SynthesizeSpeech, you can specify multiple lexicons. In this case, the first lexicon specified (in order from left to right) overrides any preceding lexicons.

Consider the following two lexicons. Note that each lexicon describes different aliases for the same grapheme W3C.

• Lexicon 1: w3c.pls

```xml
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
   xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   alphabet="ipa" xml:lang="en-US">
   <lexeme>
   <grapheme>W3C</grapheme>
   <alias>World Wide Web Consortium</alias>
   </lexeme>
</lexicon>
```
• Lexicon 2: w3cAlternate.pls

```xml
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
   xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.w3.org/2005/01/pronunciation-lexicon
   http://www.w3.org/TR/2007/CR-pronunciation-lexicon-20071212/pls.xsd"
   alphabet="ipa" xml:lang="en-US">
<lexeme>
  <grapheme>W3C</grapheme>
  <alias>WWW Consortium</alias>
</lexeme>
</lexicon>
```

Suppose you store these lexicons as w3c and w3cAlternate respectively. If you specify lexicons in order (w3c followed by w3cAlternate) in a SynthesizeSpeech call, the alias for W3C defined in the first lexicon has precedence over the second. To test the lexicons, do the following:

1. Save the lexicons locally in files called w3c.pls and w3cAlternate.pls.
2. Upload these lexicons using the put-lexicon AWS CLI command.
   - Upload the w3c.pls lexicon and store it as w3c.
     ```bash
     aws polly put-lexicon \
     --name w3c \
     --content file://w3c.pls
     ```
   - Upload the w3cAlternate.pls lexicon on the service as w3cAlternate.
     ```bash
     aws polly put-lexicon \
     --name w3cAlternate \
     --content file://w3cAlternate.pls
     ```
3. Run the synthesize-speech command to synthesize sample text to an audio stream (speech.mp3), and specify both lexicons using the lexicon-name parameter.
   ```bash
   aws polly synthesize-speech \
   --text 'PLS is a W3C recommendation' \
   --voice-id Joanna \
   --output-format mp3 \
   --lexicon-names '["w3c","w3cAlternate"]' \
   speech.mp3
   ```
4. Test the resulting speech.mp3. It should read as follows:
   ```plaintext
   PLS is a World Wide Web Consortium recommendation
   ```

### Additional Code Samples for the PutLexicon API

- **Java Sample**: PutLexicon (p. 152)
- **Python (Boto3) Sample**: PutLexicon (p. 158)
Using the GetLexicon Operation

Amazon Polly provides the GetLexicon (p. 254) API operation to retrieve the content of a pronunciation lexicon you stored in your account in a specific region.

The following `get-lexicon` AWS CLI command retrieves the content of the example lexicon.

```
aws polly get-lexicon \
--name example
```

If you don’t already have a lexicon stored in your account, you can use the PutLexicon operation to store one. For more information, see Using the PutLexicon Operation (p. 137).

The following is a sample response. In addition to the lexicon content, the response returns the metadata, such as the language code to which the lexicon applies, number of lexemes defined in the lexicon, the Amazon Resource Name (ARN) of the resource, and the size of the lexicon in bytes. The `LastModified` value is a Unix timestamp.

```
{
    "Lexicon": {
        "Content": "lexicon content in plain text PLS format",
        "Name": "example"
    },
    "LexiconAttributes": {
        "LanguageCode": "en-US",
        "LastModified": 1474222543.989,
        "Alphabet": "ipa",
        "LexemesCount": 1,
        "LexiconArn": "arn:aws:polly:us-east-2:account-id:lexicon/example",
        "Size": 495
    }
}
```

Additional Code Samples for the GetLexicon API

- Java Sample: GetLexicon (p. 151)
- Python (Boto3) Sample: GetLexicon (p. 156)

Using the ListLexicons Operations

Amazon Polly provides the ListLexicons (p. 258) API operation that you can use to get the list of pronunciation lexicons in your account in a specific AWS Region. The following AWS CLI call lists the lexicons in your account in the us-east-2 region.

```
aws polly list-lexicons
```

The following is an example response, showing two lexicons named w3c and tomato. For each lexicon, the response returns metadata such as the language code to which the lexicon applies, the number of lexemes defined in the lexicon, the size in bytes, and so on. The language code describes a language and locale to which the lexemes defined in the lexicon apply.

```
{
    "Lexicons": [
        {
```
Additional Code Samples for the ListLexicon API

- Java Sample: ListLexicons (p. 151)
- Python (Boto3) Sample: ListLexicon (p. 157)

Using the DeleteLexicon Operation

Amazon Polly provides the DeleteLexicon (p. 249) API operation to delete a pronunciation lexicon from a specific AWS Region in your account. The following AWS CLI deletes the specified lexicon.

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (’) for interior tags.

```bash
aws polly delete-lexicon \
  --name example
```

Additional Code Samples for the DeleteLexicon API

- Java Sample: DeleteLexicon (p. 149)
- Python (Boto3) Sample: DeleteLexicon (p. 156)
Creating Long Audio Files

To create TTS files for large passages of text, use Amazon Polly's asynchronous synthesis functionality. This uses the three SpeechSynthesisTask APIs:

- **StartSpeechSynthesisTask**: starts a new synthesis task.
- **GetSpeechSynthesisTask**: returns details about a previously submitted synthesis task.
- **ListSpeechSynthesisTasks**: lists all submitted synthesis tasks.

The SynthesizeSpeech operation produces audio in near-real time, with relatively little latency in most cases. To do this, the operation can only synthesize 3000 characters.

Amazon Polly's Asynchronous Synthesis feature overcomes the challenge of processing a larger text document by changing the way the document is both synthesized and returned. When a synthesis request is made by submitting input text using the `StartSpeechSynthesisTask`, Amazon Polly queues the requests, and then asynchronously processes them in the background as soon as the system resources are available. Amazon Polly then uploads the resulting speech or speech marks stream directly to your (required) Amazon Simple Storage Service (Amazon S3) bucket, and notifies you about the completed file's availability through your (optional) SNS topic.

In this way, all of the functionality except near-real time processing is available for texts of up to 100,000 billable characters (or 200,000 total characters) in length.

To synthesize a document using this method, you must have an Amazon S3 bucket that is writable to which the audio file can be saved. You can be notified when the synthesized audio is ready by providing an optional SNS Topic identifier. When the synthesis task is complete, Amazon Polly will publish a message on that topic. This message may also contain useful error information in cases where the synthesis task didn't succeed. To do this, make sure that the user creating the synthesis task can also publish to the SNS Topic. See the Amazon SNS documentation for more information on how to create and subscribe to an SNS Topic.

**Encryption**

You can store the output file in an encrypted form in your S3 bucket if desired. To do this, you enable Amazon S3 bucket encryption, which use one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256).

**Topics**

- Setting Up the IAM Policy for Asynchronous Synthesis (p. 144)
- Creating Long Audio Files (Console) (p. 145)
- Creating Long Audio Files (CLI) (p. 146)

Setting Up the IAM Policy for Asynchronous Synthesis

In order to use the asynchronous synthesis functionality, you will need an IAM policy that allows the following:
Creating Long Audio Files (Console)

You can use the Amazon Polly console to create long speeches using asynchronous synthesis with the same functionality as you can use with the AWS CLI. This is done using the Text-to-Speech tab much like any other synthesis.

The other asynchronous synthesis functionality is also available via the console. The S3 synthesis tasks tab reflects the ListSpeechSynthesisTasks functionality, displaying all tasks saved to the S3 bucket and enabling you to filter them if you want. Clicking on a specific single task shows its details, reflecting GetSpeechSynthesisTask functionality.

To synthesize a large text using the Amazon Polly Console

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. Choose the Text-to-Speech tab.
3. With SSML on or off, type or past your text into the input box.
4. Choose the language, region, and voice for your text.
5. Choose Save to S3.

   **Note**
   Both the Download and Listen options are greyed out if the text length is above the limits for the real-time SynthesizeSpeech operation.
6. The console opens a form so that you can choose where to store the output file.
a. Fill in the name of the destination Amazon S3 bucket.
b. Optionally, fill in the prefix key of the output.

**Note**
The output S3 bucket must be writable.

c. If you want to be notified when the synthesis task is complete, provide an optional SNS topic identifier.

**Note**
The SNS must be open for publication by the current console user to use this option.
For more information, see Amazon Simple Notification Service (SNS)

d. Choose Save to S3.

To retrieve information on your speech synthesis tasks

1. In the console, choose the **S3 Synthesis Tasks** tab.
2. The tasks are displayed in date order. To filter the tasks, by status, choose All statuses and then choose the status to use.
3. To view the details of a specific task, choose the linked Task ID.

Creating Long Audio Files (CLI)

Amazon Polly asynchronous synthesis functionality uses three SpeechSynthesisTask APIs to work with large amounts of text:

- **StartSpeechSynthesisTask**: starts a new synthesis task.
- **GetSpeechSynthesisTask**: returns details about a previously submitted synthesis task.
- **ListSpeechSynthesisTasks**: lists all submitted synthesis tasks.

**Synthesizing large amounts of text (StartSpeechSynthesisTask)**

When you want to create an audio file larger than one that you can create with the real-time SynthesizeSpeech, use the StartSpeechSynthesisTask operation. In addition to the arguments needed for the SynthesizeSpeech operation, StartSpeechSynthesisTask also requires the name of an Amazon S3 bucket. Two other optional arguments are also available: a key prefix for the output file and the ARN for an SNS Topic if you want to receive status notification about the task.

- **OutputS3BucketName**: The name of the Amazon S3 bucket where the synthesis should be uploaded. This bucket should be in the same region as the Amazon Polly service. Additionally, the IAM user being used to make the call should have access to the bucket. [Required]
- **OutputS3KeyPrefix**: Key prefix for the output file. Use this parameter if you want to save the output speech file in a custom directory-like key in your bucket. [Optional]
- **SnsTopicArn**: The SNS topic ARN to use if you want to receive notification about status of the task. This SNS topic should be in the same region as the Amazon Polly service. Additionally, the IAM user being used to make the call should have access to the topic. [Optional]

For example, the following example can be used to run the start-speech-synthesis-task AWS CLI command in the US East (Ohio) region:

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (') for interior tags.
Creating Long Audio Files (CLI)

```bash
aws polly start-speech-synthesis-task \\
  --region us-east-2 \\
  --endpoint-url "https://polly.us-east-2.amazonaws.com/" \\
  --output-format mp3 \\
  --output-s3-bucket-name your-bucket-name \\
  --output-s3-key-prefix optional/prefix/path/file \\
  --voice-id Joanna \\
  --text file://text_file.txt
```

This will result in a response that looks similar to this:

```json
"SynthesisTask":
{
  "OutputFormat": "mp3",
  "TextType": "text",
  "CreationTime": [..],
  "RequestCharacters": [..],
  "TaskStatus": "scheduled",
  "TaskId": [task_id],
  "VoiceId": "Joanna"
}
```

The `start-speech-synthesis-task` operation returns several new fields:

- **OutputUri**: the location of your output speech file.
- **TaskId**: a unique identifier for the speech synthesis task generated by Amazon Polly.
- **CreationTime**: a timestamp for when the task was initially submitted.
- **RequestCharacters**: the number of billable characters in the task.
- **TaskStatus**: provides information on the status of the submitted task.

When your task is submitted, the initial status will show **scheduled**. When Amazon Polly starts processing the task, the status will change to **inProgress** and later, to **completed** or **failed**. If the task fails, an error message will be returned when calling either the `GetSpeechSynthesisTask` or `ListSpeechSynthesisTasks` operation.

When the task is completed, the speech file is available at the location specified in `OutputUri`.

### Retrieving information on your speech synthesis task

You can get information on a task, such as errors, status, and so on, using the `GetSpeechSynthesisTask` operation. To do this, you will need the `task-id` returned by the `StartSpeechSynthesisTask`.

For example, the following example can be used to run the `get-speech-synthesis-task` AWS CLI command:

```bash
aws polly get-speech-synthesis-task \\
  --region us-east-2 \\
  --endpoint-url "https://polly.us-east-2.amazonaws.com/" \\
  --task-id task_identifier
```

You can also list all speech synthesis tasks that you've run in the current region using the `ListSpeechSynthesisTasks` operation.
For example, the following example can be used to run the `list-speech-synthesis-tasks` AWS CLI command:

```
aws polly list-speech-synthesis-tasks \
--region us-east-2 \
--endpoint-url "https://polly.us-east-2.amazonaws.com/
```
Code and Application Examples

This section provides code samples and example applications that you can use to explore Amazon Polly.

Topics
- Sample Code (p. 149)
- Example Applications (p. 194)

The **Sample Code** topic contains snippets of code organized by programming language and separated into examples for different Amazon Polly functionality. The **Example Application** topic contains applications organized by programming language that can be used independently to explore Amazon Polly.

Before you start using these examples, we recommend that you first read How Amazon Polly Works (p. 3) and follow the steps described in Getting Started with Amazon Polly (p. 4).

Sample Code

This topic contains code samples for various functionality which can be used to explore Amazon Polly.

**Sample Code by Programming Language**
- Java Samples (p. 149)
- Python Samples (p. 155)

Java Samples

The following code samples show how to use Java-based applications to accomplish various tasks with Amazon Polly. These samples are not full examples, but can be included in larger Java applications that use the AWS SDK for Java.

**Code Snippets**
- DeleteLexicon (p. 149)
- DescribeVoices (p. 150)
- GetLexicon (p. 151)
- ListLexicons (p. 151)
- PutLexicon (p. 152)
- StartSpeechSynthesisTask (p. 153)
- Speech Marks (p. 154)
- SynthesizeSpeech (p. 155)

**DeleteLexicon**

The following Java code sample show how to use Java-based applications to delete a specific lexicon stored in an AWS Region. A lexicon which has been deleted is not available for speech synthesis, nor can it be retrieved using either the GetLexicon or ListLexicon APIs.

For more information on this operation, see the reference for the `DeleteLexicon` API.

```java
package com.amazonaws.polly.samples;
```
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
import com.amazonaws.services.polly.model.DeleteLexiconRequest;

public class DeleteLexiconSample {
    private String LEXICON_NAME = "SampleLexicon";
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();

    public void deleteLexicon() {
        DeleteLexiconRequest deleteLexiconRequest = new DeleteLexiconRequest().withName(LEXICON_NAME);
        try {
            client.deleteLexicon(deleteLexiconRequest);
        } catch (Exception e) {
            System.err.println("Exception caught: " + e);
        }
    }
}

DescribeVoices

The following Java code sample show how to use Java-based applications to produce a list of the voices that are available for use when requesting speech synthesis. You can optionally specify a language code to filter the available voices. For example, if you specify en-US, the operation returns a list of all available US English voices.

For more information on this operation, see the reference for the DescribeVoices API.

package com.amazonaws.polly.samples;
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
import com.amazonaws.services.polly.model.DescribeVoicesRequest;
import com.amazonaws.services.polly.model.DescribeVoicesResult;

public class DescribeVoicesSample {
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();

    public void describeVoices() {
        DescribeVoicesRequest allVoicesRequest = new DescribeVoicesRequest();
        DescribeVoicesRequest enUsVoicesRequest = new DescribeVoicesRequest().withLanguageCode("en-US");

        try {
            String nextToken;
            do {
                DescribeVoicesResult allVoicesResult = client.describeVoices(allVoicesRequest);
                nextToken = allVoicesResult.getNextToken();
                allVoicesRequest.setNextToken(nextToken);
                System.out.println("All voices: " + allVoicesResult.getVoices());
            } while (nextToken != null);

            do {
                DescribeVoicesResult enUsVoicesResult = client.describeVoices(enUsVoicesRequest);
                nextToken = enUsVoicesResult.getNextToken();
                enUsVoicesRequest.setNextToken(nextToken);
                System.out.println("en-US voices: " + enUsVoicesResult.getVoices());
            } while (nextToken != null);
        }
    }
}
GetLexicon

The following Java code sample show how to use Java-based applications to produce the content of a specific pronunciation lexicon stored in a AWS Region.

For more information on this operation, see the reference for the GetLexicon API.

ListLexicons

The following Java code sample shows how to use Java-based applications to produce a list of pronunciation lexicons stored in an AWS Region.

For more information on this operation, see the reference for the ListLexicons API.
String nextToken;
    do {
        ListLexiconsResult listLexiconsResult =
            client.listLexicons(listLexiconsRequest);
        nextToken = listLexiconsResult.getNextToken();
        listLexiconsRequest.setNextToken(nextToken);

        for (LexiconDescription lexiconDescription :
            listLexiconsResult.getLexicons()) {
            LexiconAttributes attributes = lexiconDescription.getAttributes();
        }
    } while (nextToken != null);
} catch (Exception e) {
    System.err.println("Exception caught: " + e);
}
}

PutLexicon

The following Java code sample show how to use Java-based applications to store a pronunciation lexicon in an AWS Region.

For more information on this operation, see the reference for the PutLexicon API.

```java
package com.amazonaws.polly.samples;
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
import com.amazonaws.services.polly.model.PutLexiconRequest;

public class PutLexiconSample {

    private String LEXICON_NAME = "SampleLexicon";

    public void putLexicon() {
        PutLexiconRequest putLexiconRequest = new PutLexiconRequest()
            .withContent(LEXICON_CONTENT)
            .withName(LEXICON_NAME);

        try {
            client.putLexicon(putLexiconRequest);
        } catch (Exception e) {
            System.err.println("Exception caught: " + e);
        }
    }
}
```
StartSpeechSynthesisTask

The following Java code sample show how to use Java-based applications to synthesize a long speech (up to 100,000 billed characters) and store it directly in an Amazon S3 bucket.

For more information, see the reference for StartSpeechSynthesisTask API.

```java
package com.amazonaws.parrot.service.tests.speech.task;

import com.amazonaws.parrot.service.tests.AbstractParrotServiceTest;
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.model.*;
import org.awaitility.Duration;
import java.util.concurrent.TimeUnit;
import static org.awaitility.Awaitility.await;

public class StartSpeechSynthesisTaskSample {
    private static final int SYNTHESIS_TASK_TIMEOUT_SECONDS = 300;
    private static final AmazonPolly AMAZON_POLLY_CLIENT =
        AmazonPollyClientBuilder.defaultClient();
    private static final String PLAIN_TEXT = "This is a sample text to be synthesized.";
    private static final String OUTPUT_FORMAT_MP3 = OutputFormat.Mp3.toString();
    private static final String OUTPUT_BUCKET = "synth-books-buckets";
    private static final Duration SYNTHESIS_TASK_POLL_INTERVAL = Duration.FIVE_SECONDS;
    private static final Duration SYNTHESIS_TASK_POLL_DELAY = Duration.TEN_SECONDS;

    public static void main(String... args) {
        StartSpeechSynthesisTaskRequest request = new StartSpeechSynthesisTaskRequest()
            .withOutputFormat(OUTPUT_FORMAT_MP3)
            .withText(PLAIN_TEXT)
            .withTextType(TextType.Text)
            .withVoiceId(VoiceId.Amy)
            .withOutputS3BucketName(OUTPUT_BUCKET)
            .withSnsTopicArn(SNS_TOPIC_ARN)
            .withEngine("neural");
        StartSpeechSynthesisTaskResult result =
            AMAZON_POLLY_CLIENT.startSpeechSynthesisTask(request);
        String taskId = result.getSynthesisTask().getTaskId();
        await().with()
            .pollInterval(SYNTHESIS_TASK_POLL_INTERVAL)
            .pollDelay(SYNTHESIS_TASK_POLL_DELAY)
            .atMost(SYNTHESIS_TASK_TIMEOUT_SECONDS, TimeUnit.SECONDS)
            .until()
                () ->
            getSynthesisTaskStatus(taskId).equals(TaskStatus.Completed.toString());

        private static SynthesisTask getSynthesisTask(String taskId) {
            GetSpeechSynthesisTaskRequest request = new GetSpeechSynthesisTaskRequest()
                .withTaskId(taskId);
            GetSpeechSynthesisTaskResult result =
                AMAZON_POLLY_CLIENT.getSpeechSynthesisTask(request);
            return result.getSynthesisTask();
        }
    }
```
private static String getSynthesisTaskStatus(String taskId) {
    GetSpeechSynthesisTaskRequest getSpeechSynthesisTaskRequest = new GetSpeechSynthesisTaskRequest()
        .withTaskId(taskId);
    GetSpeechSynthesisTaskResult result = AMAZON_POLLY_CLIENT.getSpeechSynthesisTask(getSpeechSynthesisTaskRequest);
    return result.getSynthesisTask().getTaskStatus();
}

Speech Marks

The following code sample shows how to use Java-based applications to synthesize speech marks for inputed text. This functionality uses the SynthesizeSpeech API.

For more information on this functionality, see Speech Marks (p. 99).

For more information on the API, see the reference for SynthesizeSpeech API.

package com.amazonaws.polly.samples;
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
import com.amazonaws.services.polly.model.OutputFormat;
import com.amazonaws.services.polly.model.SpeechMarkType;
import com.amazonaws.services.polly.model.SynthesizeSpeechRequest;
import com.amazonaws.services.polly.model.SynthesizeSpeechResult;
import com.amazonaws.services.polly.model.VoiceId;
import java.io.File;
import java.io.FileOutputStream;
import java.io.InputStream;
public class SynthesizeSpeechMarksSample {
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();

    public void synthesizeSpeechMarks() {
        String outputFileName = "/tmp/speechMarks.json";
        SynthesizeSpeechRequest synthesizeSpeechRequest = new SynthesizeSpeechRequest()
            .withOutputFormat(OutputFormat.Json)
            .withSpeechMarkTypes(SpeechMarkType.Viseme, SpeechMarkType.Word)
            .withVoiceId(VoiceId.Joanna)
            .withText("This is a sample text to be synthesized.");

        try (FileOutputStream outputStream = new FileOutputStream(new File(outputFileName))) {
            SynthesizeSpeechResult synthesizeSpeechResult = client.synthesizeSpeech(synthesizeSpeechRequest);
            byte[] buffer = new byte[2 * 1024];
            int readBytes;

            try (InputStream in = synthesizeSpeechResult.getAudioStream()){
                while ((readBytes = in.read(buffer)) > 0) {
                    outputStream.write(buffer, 0, readBytes);
                }
            }
            catch (Exception e) {
                System.err.println("Exception caught: " + e);
            }
        }
    }
The following Java code sample show how to use Java-based applications to synthesize speech with shorter texts for near-real time processing.

For more information, see the reference for `SynthesizeSpeech API`.

```java
package com.amazonaws.polly.samples;

import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
import com.amazonaws.services.polly.model.OutputFormat;
import com.amazonaws.services.polly.model.SynthesizeSpeechRequest;
import com.amazonaws.services.polly.model.SynthesizeSpeechResult;
import com.amazonaws.services.polly.model.VoiceId;
import java.io.File;
import java.io.FileOutputStream;
import java.io.InputStream;
public class SynthesizeSpeechSample {
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();
    public void synthesizeSpeech() {
        String outputFileName = "/tmp/speech.mp3";
        SynthesizeSpeechRequest synthesizeSpeechRequest = new SynthesizeSpeechRequest()
                .withOutputFormat(OutputFormat.Mp3)
                .withVoiceId(VoiceId.Joanna)
                .withText("This is a sample text to be synthesized.")
                .withEngine("neural");

        try (FileOutputStream outputStream = new FileOutputStream(new File(outputFileName))) {
            SynthesizeSpeechResult synthesizeSpeechResult =
                    client.synthesizeSpeech(synthesizeSpeechRequest);
            byte[] buffer = new byte[2 * 1024];
            int readBytes;
            try (InputStream in = synthesizeSpeechResult.getAudioStream()){
                while ((readBytes = in.read(buffer)) > 0) {
                    outputStream.write(buffer, 0, readBytes);
                }
            }
            catch (Exception e) {
                System.err.println("Exception caught: " + e);
            }
        }
    }
}
```

**Python Samples**

The following code samples show how to use Python (boto3)-based applications to accomplish various tasks with Amazon Polly. These samples are not intended to be full examples, but can be included in larger Python applications that use the AWS SDK for Python (Boto).

**Code Snippets**
- `DeleteLexicon (p. 156)`
DeleteLexicon

The following Python code example uses the AWS SDK for Python (Boto) to delete a lexicon in the region specified in your local AWS configuration. The example deletes only the specified lexicon. It asks you to confirm that you want to proceed before actually deleting the lexicon.

The following code example uses default credentials stored in the AWS SDK configuration file. For information about creating the configuration file, see Step 3.1: Set Up the AWS Command Line Interface (AWS CLI) (p. 7).

For more information on this operation, see the reference for the `DeleteLexicon` API.

```python
from argparse import ArgumentParser
from sys import version_info
from boto3 import Session
from botocore.exceptions import BotoCoreError, ClientError

# Define and parse the command line arguments
cli = ArgumentParser(description="DeleteLexicon example")
cli.add_argument("name", type=str, metavar="LEXICON_NAME")
arguments = cli.parse_args()

# Create a client using the credentials and region defined in the adminuser
# section of the AWS credentials and configuration files
session = Session(profile_name="adminuser")
polly = session.client("polly")

# Request confirmation
prompt = input if version_info >= (3, 0) else raw_input
proceed = prompt((u"This will delete the "{0}" lexicon,"
" do you want to proceed? [y,n]: ").format(arguments.name))

if proceed in ("y", "Y"):
    print(u"Deleting {0}...".format(arguments.name))
    try:
        # Request deletion of a lexicon by name
        response = polly.delete_lexicon(Name=arguments.name)
    except (BotoCoreError, ClientError) as error:
        # The service returned an error, exit gracefully
        cli.error(error)
    print("Done."
else:
    print("Cancelled."

GetLexicon

The following Python code uses the AWS SDK for Python (Boto) to retrieve all lexicons stored in an AWS Region. The example accepts a lexicon name as a command line parameter and fetches that lexicon only, printing out the tmp path where it has been saved locally.

```
The following code example uses default credentials stored in the AWS SDK configuration file. For information about creating the configuration file, see Step 3.1: Set Up the AWS Command Line Interface (AWS CLI) (p. 7).

For more information on this operation, see the reference for the GetLexicon API.

```python
from argparse import ArgumentParser
from os import path
from tempfile import gettempdir
from boto3 import Session
from botocore.exceptions import BotoCoreError, ClientError

# Define and parse the command line arguments
cli = ArgumentParser(description="GetLexicon example")
cli.add_argument("name", type=str, metavar="LEXICON_NAME")
arguments = cli.parse_args()

# Create a client using the credentials and region defined in the adminuser
# section of the AWS credentials and configuration files
session = Session(profile_name="adminuser")
polly = session.client("polly")

print(u"Fetching {0}...".format(arguments.name))

try:
    # Fetch lexicon by name
    response = polly.get_lexicon(Name=arguments.name)
except (BotoCoreError, ClientError) as error:
    # The service returned an error, exit gracefully
    cli.error(error)

# Get the lexicon data from the response
lexicon = response.get("Lexicon", {})

# Access the lexicon's content
if "Content" in lexicon:
    output = path.join(gettempdir(), u"%s.pls" % arguments.name)
    print(u"Saving to %s..." % output)

    try:
        # Save the lexicon contents to a local file
        with open(output, "w") as pls_file:
            pls_file.write(lexicon["Content"])
        except IOError as error:
            # Could not write to file, exit gracefully
            cli.error(error)
    except IOError as error:
        # The response didn't contain lexicon data, exit gracefully
        cli.error("Could not fetch lexicons contents")

print("Done.")
```

**ListLexicon**

The following Python code example uses the AWS SDK for Python (Boto) to list the lexicons in your account in the region specified in your local AWS configuration. For information about creating the configuration file, see Step 3.1: Set Up the AWS Command Line Interface (AWS CLI) (p. 7).

For more information on this operation, see the reference for the ListLexicons API.

```python
import sys
```
from boto3 import Session
from botocore.exceptions import BotoCoreError, ClientError

# Create a client using the credentials and region defined in the adminuser
# section of the AWS credentials and configuration files
session = Session(profile_name="adminuser")
polly = session.client("polly")

try:
    # Request the list of available lexicons
    response = polly.list_lexicons()
except (BotoCoreError, ClientError) as error:
    # The service returned an error, exit gracefully
    print(error)
sys.exit(-1)

# Get the list of lexicons in the response
lexicons = response.get("Lexicons", [])
print("{0} lexicon(s) found".format(len(lexicons)))

# Output a formatted list of lexicons with some of the attributes
for lexicon in lexicons:
    print((u" - {Name} ({Attributes[LanguageCode]}), ",{Attributes[LexemesCount]} lexeme(s)").format(**lexicon))

PutLexicon

The following code sample show how to use Python (boto3)-based applications to store a pronunciation lexicon in an AWS Region.

For more information on this operation, see the reference for the PutLexicon API.

Note the following:

- You need to update the code by providing a local lexicon file name and a stored lexicon name.
- The example assumes you have lexicon files created in a subdirectory called pls. You need to update the path as appropriate.

The following code example uses default credentials stored in the AWS SDK configuration file. For information about creating the configuration file, see Step 3.1: Set Up the AWS Command Line Interface (AWS CLI) (p. 7).

For more information on this operation, see the reference for the PutLexicon API.

from argparse import ArgumentParser
from boto3 import Session
from botocore.exceptions import BotoCoreError, ClientError

# Define and parse the command line arguments
cli = ArgumentParser(description="PutLexicon example")
cli.add_argument("path", type=str, metavar="FILE_PATH")
cli.add_argument("-n", "--name", type=str, required=True,
                     metavar="LEXICON_NAME", dest="name")
arguments = cli.parse_args()

# Create a client using the credentials and region defined in the adminuser
# section of the AWS credentials and configuration files
session = Session(profile_name="adminuser")
polly = session.client("polly")

# Open the PLS lexicon file for reading
try:
    with open(arguments.path, "r") as lexicon_file:
        # Read the pls file contents
        lexicon_data = lexicon_file.read()

        # Store the PLS lexicon on the service.
        # If a lexicon with that name already exists,
        # its contents will be updated
        response = polly.put_lexicon(Name=arguments.name,
                                Content=lexicon_data)
except (IOError, BotoCoreError, ClientError) as error:
    # Could not open/read the file or the service returned an error,
    # exit gracefully
    cli.error(error)

print(u"The "{0}" lexicon is now available for use.".format(arguments.name))

### StartSpeechSynthesisTask

The following Python code example uses the AWS SDK for Python (Boto) to list the lexicons in your account in the region specified in your local AWS configuration. For information about creating the configuration file, see Step 3.1: Set Up the AWS Command Line Interface (AWS CLI) (p. 7).

For more information, see the reference for StartSpeechSynthesisTask API.

```python
import boto3
import time

polly_client = boto3.Session(  
    aws_access_key_id='',
    aws_secret_access_key='',
    region_name='eu-west-2').client('polly')

response = polly_client.start_speech_synthesis_task(VoiceId='Joanna',
                                OutputS3BucketName='synth-books-buckets',
                                OutputS3KeyPrefix='key',
                                OutputFormat='mp3',
                                Text='This is a sample text to be synthesized.',
                                Engine='neural')

taskId = response['SynthesisTask']['TaskId']

print( "Task id is {} ".format(taskId))

task_status = polly_client.get_speech_synthesis_task(TaskId = taskId)

print(task_status)
```

### SynthesizeSpeech

The following Python code example uses the AWS SDK for Python (Boto) synthesize speech with shorter texts for near real-time processing. For more information, see the reference for the

```
Synthesizes UTF-8 input, plain text or SSML, to a stream of bytes. SSML input must be valid, well-formed SSML. Some alphabets might not be available with all the voices (for example, Cyrillic might not be read at all by English voices) unless phoneme mapping is used. For more information, see How it Works.

```

159
Request Syntax

POST /v1/speech HTTP/1.1

Content-type: application/json

{
    "Engine": "string",
    "LanguageCode": "string",
    "LexiconNames": [ "string" ],
    "OutputFormat": "string",
    "SampleRate": "string",
    "SpeechMarkTypes": [ "string" ],
    "Text": "string",
    "TextType": "string",
    "VoiceId": "string"
}

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

Engine (p. 270)

Specifies the engine (standard or neural) for Amazon Polly to use when processing input text for speech synthesis. For information on Amazon Polly voices and which voices are available in standard-only, NTTS-only, and both standard and NTTS formats, see Available Voices.

NTTS-only voices

When using NTTS-only voices such as Kevin (en-US), this parameter is required and must be set to neural. If the engine is not specified, or is set to standard, this will result in an error.

Type: String

Valid Values: standard | neural

Required: Yes

Standard voices

For standard voices, this is not required; the engine parameter defaults to standard. If the engine is not specified, or is set to standard and an NTTS-only voice is selected, this will result in an error.

Type: String
If a bilingual voice is used and no language code is specified, Amazon Polly uses the default language of the bilingual voice. The default language for any voice is the one returned by the DescribeVoices operation for the LanguageCode parameter. For example, if no language code is specified, Aditi will use Indian English rather than Hindi.

Type: String


Required: No

LexiconNames (p. 270)

List of one or more pronunciation lexicon names you want the service to apply during synthesis. Lexicons are applied only if the language of the lexicon is the same as the language of the voice. For information about storing lexicons, see PutLexicon.

Type: Array of strings

Array Members: Maximum number of 5 items.

Pattern: [0-9A-Za-z]{1,20}

Required: No

OutputFormat (p. 270)

The format in which the returned output will be encoded. For audio stream, this will be mp3, ogg_vorbis, or pcm. For speech marks, this will be json.

When pcm is used, the content returned is audio pcm in a signed 16-bit, 1 channel (mono), little-endian format.

Type: String

Valid Values: json | mp3 | ogg_vorbis | pcm

Required: Yes

SampleRate (p. 270)

The audio frequency specified in Hz.

The valid values for mp3 and ogg_vorbis are "8000", "16000", "22050", and "24000". The default value for standard voices is "22050". The default value for neural voices is "24000".

Valid values for pcm are "8000" and "16000" The default value is "16000".

Type: String

Required: No

SpeechMarkTypes (p. 270)

The type of speech marks returned for the input text.

Type: Array of strings
Array Members: Maximum number of 4 items.

Valid Values: sentence | ssml | viseme | word

Required: No

Text (p. 270)

Input text to synthesize. If you specify ssml as the TextType, follow the SSML format for the input text.

Type: String

Required: Yes

TextType (p. 270)

Specifies whether the input text is plain text or SSML. The default value is plain text. For more information, see Using SSML.

Type: String

Valid Values: ssml | text

Required: No

VoiceId (p. 270)

Voice ID to use for the synthesis. You can get a list of available voice IDs by calling the DescribeVoices operation.

Type: String

Valid Values: Aditi | Amy | Astrid | Bianca | Brian | Camila | Carla | Carmen | Celine | Chantal | Conchita | Cristiano | Dora | Emma | Enrique | Ewa | Filiz | Gabrielle | Geraint | Giorgio | Gwyneth | Hans | Ines | Ivy | Jacek | Jan | Joanna | Joey | Justin | Karl | Kendra | Kevin | Kimberly | Lea | Liv | Lotte | Lucia | Lupe | Mads | Maja | Marlene | Mathieu | Matthew | Maxim | Mia | Miguel | Mizuki | Naja | Nicole | Olivia | Penelope | Raveena | Ricardo | Ruben | Russell | Salli | Seoyeon | Takumi | Tatyana | Vicki | Vitoria | Zeina | Zhiyu | Aria | Ayanda

Required: Yes

Response Syntax

<table>
<thead>
<tr>
<th>HTTP/1.1 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content-Type: ContentType</td>
</tr>
<tr>
<td>x-amzn-RequestCharacters: RequestCharacters</td>
</tr>
</tbody>
</table>

AudioStream

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.
**ContentType (p. 272)**

Specifies the type audio stream. This should reflect the `OutputFormat` parameter in your request.
- If you request `mp3` as the `OutputFormat`, the `ContentType` returned is `audio/mpeg`.
- If you request `ogg_vorbis` as the `OutputFormat`, the `ContentType` returned is `audio/ogg`.
- If you request `pcm` as the `OutputFormat`, the `ContentType` returned is `audio/pcm` in a signed 16-bit, 1 channel (mono), little-endian format.
- If you request `json` as the `OutputFormat`, the `ContentType` returned is `audio/json`.

**RequestCharacters (p. 272)**

Number of characters synthesized.
The response returns the following as the HTTP body.

**AudioStream (p. 272)**

Stream containing the synthesized speech.

**Errors**

**EngineNotSupportedException**

This engine is not compatible with the voice that you have designated. Choose a new voice that is compatible with the engine or change the engine and restart the operation.

HTTP Status Code: 400

**InvalidSampleRateException**

The specified sample rate is not valid.

HTTP Status Code: 400

**InvalidSsmIException**

The SSML you provided is invalid. Verify the SSML syntax, spelling of tags and values, and then try again.

HTTP Status Code: 400

**LanguageNotSupportedException**

The language specified is not currently supported by Amazon Polly in this capacity.

HTTP Status Code: 400

**LexiconNotFoundException**

Amazon Polly can't find the specified lexicon. This could be caused by a lexicon that is missing, its name is misspelled or specifying a lexicon that is in a different region.

Verify that the lexicon exists, is in the region (see `ListLexicons` (p. 258)) and that you spelled its name is spelled correctly. Then try again.

HTTP Status Code: 404
MarksNotSupportedForFormatException

Speech marks are not supported for the OutputFormat selected. Speech marks are only available for content in json format.

HTTP Status Code: 400

ServiceFailureException

An unknown condition has caused a service failure.

HTTP Status Code: 500

SsmiMarksNotSupportedForTextTypeException

SSML speech marks are not supported for plain text-type input.

HTTP Status Code: 400

TextLengthExceedededException

The value of the "Text" parameter is longer than the accepted limits. For the SynthesizeSpeech API, the limit for input text is a maximum of 6000 characters total, of which no more than 3000 can be billed characters. For the StartSpeechSynthesisTask API, the maximum is 200,000 characters, of which no more than 100,000 can be billed characters. SSML tags are not counted as billed characters.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS Command Line Interface

• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3

(p. 270) operation.

This example uses a short string of plain text. You can use SSML text for more control over the output. For more information, see

You can use Amazon Polly to generate speech from either plain text or from documents marked up with Speech Synthesis Markup Language (SSML). Using SSML-enhanced text gives you additional control over how Amazon Polly generates speech from the text you provide.

For example, you can include a long pause within your text, or change the speech rate or pitch. Other options include:
• emphasizing specific words or phrases

• using phonetic pronunciation
• including breathing sounds
• whispering
• using the Newscaster speaking style.

For complete details on the SSML tags supported by Amazon Polly and how to use them, see Supported SSML Tags (p. 112)

When using SSML, there are several reserved characters that require special treatment. This is because SSML uses these characters as part of its code. In order to use them, you use a specific entity to escape them. For more information, see Reserved Characters in SSML (p. 105)

Amazon Polly provides these types of control with a subset of the SSML markup tags that are defined by Speech Synthesis Markup Language (SSML) Version 1.1, W3C Recommendation.

You can use SSML within the Amazon Polly console or by using the AWS CLI. The following topics show you how you can use SSML to generate speech and control the output so that it precisely fits your needs.

Topics
• Reserved Characters in SSML (p. 105)
• Using SSML (Console) (p. 107)
• Using SSML (AWS CLI) (p. 108)
• Supported SSML Tags (p. 112)

Reserved Characters in SSML

There are five predefined characters that can't normally be used within an SSML statement. These entities are reserved by the language specification. These characters are

• quotation mark
• (double quotation mark)
• &amp;
• &apos;
• &lt;
• than
Because SSML uses these characters as part of its code, to use these symbols in SSML, you must escape the character when you use it. You use the escape code instead of the actual character so it displays properly while still creating a valid SSML document. For example, the following sentence

We're using the lawyer at Peabody & Chambers, attorneys-at-law.

would be rendered in SSML as

```xml
<speak>We're using the lawyer at Peabody & Chambers, attorneys-at-law.</speak>
```

In this case, the special characters for the apostrophe and ampersand are escaped so the SSML document remains valid.

For the &, <, and > symbols, escape codes are always necessary when you use SSML. Additionally, when you use the apostrophe/single quotation mark (’) as an apostrophe, you must also use the escape code.

However, when you use the double quotation mark ("), or the apostrophe/single quotation mark (’) as a quotation mark, then whether or not you use the escape code is dependent on context.

**Double quotation marks**

- Must be escaped when in a attribute value delimited by double quotes. For example, in the following AWS CLI code

```bash
--text "Pete "Maverick" Mitchell"
```

- Do not need to be escaped when in textual context. For example, in the following

```plaintext
He said, "Turn right at the corner."
```

- Do not need to be escaped when in a attribute value delimited by single quotes. For example, in the following AWS CLI code

```bash
--text 'Pete "Maverick" Mitchell'
```

**Single quotation marks**

- Must be escaped when used as an apostrophe. For example, in the following

```plaintext
We've got to leave quickly.
```

- Do not need to be escaped when in textual context. For example, in the following

```plaintext
We've got to leave quickly.
```
Using SSML (Console)

With SSML tags, you can customize and control aspects of speech such as pronunciation, volume, and speech rate. In the AWS Management Console, the SSML-enhanced text that you want to convert to audio is entered on the SSML tab of the Text-to-Speech page. Although text entered in plain text relies on default settings for the language and voice you've chosen, text enhanced with SSML tells Amazon Polly not only what you want to say, but how you want to say it. Except for the added SSML tags, Amazon Polly synthesizes SSML-enhanced text in the same way as it synthesizes plain text. See Exercise 2: Synthesizing Speech with Plain Text Input (Console) (p. 6) for more information.

When using SSML, you enclose the entire text in a `<speak>` tag to let Amazon Polly know that you're using SSML. For example:

```
<speak>Hi! My name is Joanna. I will read any text you type here.</speak>
```

You then use specific SSML tags on the text inside the `<speak>` tags to customize the way you want the text to sound. You can add a pause, change the pace of the speech, lower or raise the volume of the voice, or add many other customizations so that the text sounds right for you. For a full list of the SSML tags that you can use, see Supported SSML Tags (p. 112).

In the following example, you use an SSML tag to tell Amazon Polly to substitute "World Wide Web Consortium" for "W3C" when it speaks a short paragraph. You also use tags to introduce a pause and whisper a word. Compare the results of this exercise with that of Applying Lexicons Using the Console (Synthesize Speech) (p. 135).

For more information on SSML, with examples, see Supported SSML Tags (p. 112).

**To synthesize speech from SSML-enhanced text (console)**

1. Sign in to the AWS Management Console and open the Amazon Polly console at https://console.aws.amazon.com/polly/.
2. If it isn't already displayed, choose the Text-to-Speech tab.
3. Turn on SSML.
4. Type or paste the following text in the text box:

   ```
   <speak>
   He was caught up in the game. In the middle of the 10/3/2014 <sub alias="World Wide Web Consortium">W3C</sub> meeting, he shouted, "Nice job!" quite loudly. When his boss stared at him, he repeated
   "Nice job," in a whispered voice.
   </speak>
   ```

The SSML tags tell Amazon Polly how to render the text:
• `<break time="1s"/>` tells Amazon Polly to pause 1 second between the first two sentences.
• `<sub alias="World Wide Web Consortium">W3C</sub>` tells Amazon Polly to substitute World Wide Web Consortium for the acronym W3C.
• `<amazon:effect name="whispered">Nice job</amazon:effect>` tells Amazon Polly to whisper the second instance of "Nice job."

Note
When you use the AWS CLI, you enclose the input text in quotation marks to differentiate it from the surrounding code. The Amazon Polly console doesn't show you code, so you don't enclose input text in quotation marks when you use it.

5. For Language, choose English, US, then choose a voice.
6. To listen to the speech, choose Listen.
7. To save the speech file, choose Download. If you want to save it in a different format, expand Additional settings, turn on Speech file format settings and choose the format that you want, then choose Download.

Using SSML (AWS CLI)

You can use the AWS CLI to synthesize SSML input text. The following examples show how to perform common tasks using the AWS CLI.

Topics

• Using SSML With the Synthesize-Speech Command (p. 108)
• Synthesizing an SSML-enhanced Document (p. 109)
• Using SSML for Common Amazon Polly Tasks (p. 110)

Using SSML With the Synthesize-Speech Command

This example shows how to use the `synthesize-speech` command with an SSML string. When you use the `synthesize-speech` command, you typically provide the following:

• The input text (required)

• Opening and closing tags (required)
• The output format
• A voice
In this example, you specify a simple text string in quotation marks along with the required opening and closing `<speak></speak>` tags.

Important
Although you don’t use quotation marks around input text in the Amazon Polly console, you must use them in use the AWS CLI. It’s also important that you differentiate between the quotation marks around input text and quotations required for individual tags.
For example, you can use standard quotation marks (") to enclose the input text, and single quotation marks (’) for interior tags, or vice versa. Either option works for
Unix, Linux, and macOS. However, with Windows you must enclose the input text in standard quotations marks and use single quotation marks for the tags.

For all operating systems, you can use standard quotation marks (") to enclose the input text, and single quotation marks (') for interior tags. For example:

```
--text "<speak>Hello <break time='300ms'/> World</speak>"
```

For Unix, Linux, and macOS, you can also use the reverse, with single quotation marks (') enclosing the input text and standard quotation marks (") for interior tags:

```
--text '<speak>Hello <break time="300ms"/> World</speak>'
```

The following AWS CLI example is formatted for Unix, Linux, and macOS. For Windows, replace the backslash (\) Unix continuation character at the end of each line with a caret (^) and use full quotation marks (") around the input text with single quotes (') for interior tags.

```
aws polly synthesize-speech \
  --text-type ssml \
  --text '<speak>Hello world</speak>' \
  --output-format mp3 \
  --voice-id Joanna \
  speech.mp3
```

To hear the synthesized speech, play the resulting `speech.mp3` file using any audio player.

## Synthesizing an SSML-enhanced Document

For longer input text, you may find it easier to save your SSML content to a file and simply specify the file name in the `synthesize-speech` command. For example you could save the following to a file called `example.xml`:

```
<?xml version="1.0"?>
<speak version="1.1"
  xmlns="http://www.w3.org/2001/10/synthesis"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/10/synthesis http://www.w3.org/TR/speech-synthesis11/synthesis.xsd"
  xml:lang="en-US">Hello World</speak>
```

The `xml:lang` attribute specifies `en-US` (US English) as the language of the input text. For information about how the language of the input text and the language of the chosen voice affect the `SynthesizeSpeech` operation, see Improving the Pronunciation of Foreign Words (p. 112).

### To run an SSML-enhanced file

1. Save the SSML to a file (for example, `example.xml`).
2. Run the following `synthesize-speech` command from the path where the XML file is stored and specify the SSML file as input by substituting `file:\example.xml` for the input text. Because this command points to a file instead of containing the actual input text, you don't use quotation marks.

    ```
    aws polly synthesize-speech \ 
    --text-type ssml \ 
    --text 'file:\example.xml' \ 
    --output-format mp3 \ 
    --voice-id Joanna \ 
    speech.mp3
    ```
Using SSML for Common Amazon Polly Tasks

The following examples show how to use SSML tags to complete common Amazon Polly tasks. For more SSML tags, see Supported SSML Tags (p. 112).

To test the following examples, use the following `synthesize-speech` command with the appropriate SSML-enhanced text:

```bash
aws polly synthesize-speech \
--text-type ssml \
--text '<speak>Hello <break time="300ms"/> World</speak>' \
--output-format mp3 \
--voice-id Joanna \
speech.mp3
```

Adding a Pause

To add a pause between words, use the `<break>` element. The following SSML `synthesize-speech` command uses the `<break>` element to add a 300-millisecond delay between the words "Hello" and "World."

```xml
<speak>
  Hello <break time="300ms"/> World.
</speak>
```

Controlling Volume, Pitch, and Speed

To control pitch, speaking rate, and speech volume, use the `<prosody>` element.

- The following `synthesize-speech` command uses the `<prosody>` element to control volume:

```xml
<speak>
  <prosody volume="+20dB">Hello world</prosody>
</speak>
```

- The following `synthesize-speech` command uses the `<prosody>` element to control pitch:

```xml
<speak>
</speak>
```
• The following `synthesize-speech` command uses the `<prosody>` element to specify the speech rate (speaking speed):

```xml
<speak>
  <prosody rate="x-fast">Hello world.</prosody>
</speak>
```

• You can specify multiple attributes in a `<prosody>` element, as shown in the following examples:

```xml
<speak>
  <prosody volume="x-loud" pitch="x-high" rate="x-fast">Hello world.</prosody>
</speak>
```

**Whispering**

To whisper words, use the `<amazon:effect name="whispered">` element. In the following example, the `<amazon:effect name="whispered">` element tells Amazon Polly to whisper "little lamb":

```xml
<speak>
  Mary has a <amazon:effect name="whispered">little lamb.</amazon:effect>
</speak>
```

To enhance this effect, use the `<prosody>` element to slightly slow down the whispered speech.

**Emphasizing Words**

To stress a word or phrase, use the `<emphasis>` element.

```xml
<speak>
  <emphasis level="strong">Hello</emphasis> world how are you?
</speak>
```

**Specifying How to Say Certain Words**

To provide information about the type of text to be spoken, use the `<say-as>` element.

For instance, in the following SSML, `<say-as>` indicates that the text 4/6 should be interpreted as a date. The attribute `interpret-as="date" format="dm"` indicates that it should be spoken as a date with the format month/day.

You can also use the `<say-as>` element to tell Amazon Polly to say numbers as fractions, telephone numbers, measurement units, and more.

```xml
<speak>
  Today is <say-as interpret-as="date" format="md">4/6</say-as>
</speak>
```

The resulting speech is "Today is June 4th." The `<say-as>` tag describes how the text should be interpreted by providing additional context with the `interpret-as` attribute.
To verify the accuracy of the synthesized speech, play the resulting speech.mp3 file.

For more information on this element, see Controlling How Special Types of Words Are Spoken (p. 122).

**Improving the Pronunciation of Foreign Words**

Amazon Polly assumes that the input text is in the same language as the language spoken by the voice you choose. To improve the pronunciation of foreign words within input text, in the `synthesize-speech` call. Specify the target language with the `xml:lang` attribute. This tells Amazon Polly to apply different pronunciation rules for the foreign words that you tag.

The following examples show how to use different combinations of languages in the input text, and how to specify voices and the pronunciation of foreign words. For a complete list of available languages, see Languages Supported by Amazon Polly (p. 18).

In the following example, the voice (Joanna) is a US English voice. By default, Amazon Polly assumes that the input text is in the same language as the voice (in this case, US English). When you use the `xml:lang` tag, Amazon Polly interprets the text as Spanish and the text is spoken as the selected voice would pronounce Spanish words, according to the pronunciation rules of the foreign language. Without this tag, the text is spoken using the pronunciation rules of the selected voice.

```xml
<speak>
  That restaurant is terrific. <lang xml:lang="es-ES">Mucho gusto.</lang>
</speak>
```

Because the language of the input text is English, Amazon Polly maps the Spanish phonemes to the closest English phonemes. As a result, Joanna speaks the text as a native US speaker who pronounces the works correctly in Spanish, but with a US English accent.

**Note**

Some languages are more similar than others, and so some language combinations work better than others.

### Supported SSML Tags

Amazon Polly supports the following SSML tags:

<table>
<thead>
<tr>
<th>Tag</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding a Pause (p. 113)</td>
<td><code>&lt;break&gt;</code></td>
</tr>
<tr>
<td>Emphasizing Words (p. 114)</td>
<td><code>&lt;emphasis&gt;</code></td>
</tr>
<tr>
<td>Specifying Another Language for Specific Words (p. 115)</td>
<td><code>&lt;lang&gt;</code></td>
</tr>
<tr>
<td>Placing a Custom Tag in Your Text (p. 116)</td>
<td><code>&lt;mark&gt;</code></td>
</tr>
<tr>
<td>Adding a Pause Between Paragraphs (p. 116)</td>
<td><code>&lt;p&gt;</code></td>
</tr>
</tbody>
</table>

Full availability
### Using Phonetic Pronunciation (p. 116)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;phoneme&gt;</td>
<td>Full availability</td>
</tr>
</tbody>
</table>

### Controlling Volume, Speaking Rate, and Pitch (p. 117)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;prosody&gt;</td>
<td>Partial availability</td>
</tr>
</tbody>
</table>

### Setting a Maximum Duration for Synthesized Speech (p. 119)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;prosody amazon:max-duration&gt;</td>
<td>Not available</td>
</tr>
</tbody>
</table>

### Adding a Pause Between Sentences (p. 121)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;s&gt;</td>
<td>Full availability</td>
</tr>
</tbody>
</table>

### Controlling How Special Types of Words Are Spoken (p. 122)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;say-as&gt;</td>
<td>Partial availability</td>
</tr>
</tbody>
</table>

### Identifying SSML-Enhanced Text (p. 113)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;speak&gt;</td>
<td>Full availability</td>
</tr>
</tbody>
</table>

### Pronouncing Acronyms and Abbreviations (p. 124)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;sub&gt;</td>
<td>Full availability</td>
</tr>
</tbody>
</table>

### Improving Pronunciation by Specifying Parts of Speech (p. 125)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;w&gt;</td>
<td>Full availability</td>
</tr>
</tbody>
</table>

### Adding the Sound of Breathing (p. 126)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="">amazon:auto-breaths</a></td>
<td>Not available</td>
</tr>
</tbody>
</table>

### Newscaster speaking style (p. 128)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;amazon:domain name=&quot;news&quot;&gt;</td>
<td>Select neural voices only</td>
</tr>
</tbody>
</table>

### Adding Dynamic Range Compression (p. 129)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;amazon:effect name=&quot;drc&quot;&gt;</td>
<td>Full availability</td>
</tr>
</tbody>
</table>

### Speaking Softly (p. 130)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;amazon:effect phonation=&quot;soft&quot;&gt;</td>
<td>Not available</td>
</tr>
</tbody>
</table>

### Controlling Timbre (p. 131)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;amazon:effect vocal-tract-length&gt;</td>
<td>Not available</td>
</tr>
</tbody>
</table>

### Whispering (p. 132)

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;amazon: effect name=&quot;whispered&quot;&gt;</td>
<td>Not available</td>
</tr>
</tbody>
</table>

### Action | SSML Tag | Availability with Neural Voices |
|--------|----------|--------------------------------|

<table>
<thead>
<tr>
<th>Action</th>
<th>SSML Tag</th>
<th>Availability with Neural Voices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you use unsupported SSML tags in either neural or standard format, you will get an error.

## Identifying SSML-Enhanced Text

### <speak>

The `<speak>` tag is the root element of all Amazon Polly SSML text. All SSML-enhanced text must be enclosed within a pair of `<speak>` tags.

```xml
<speak>Mary had a little lamb.</speak>
```

This tag is supported by both neural and standard TTS formats.
Adding a Pause

<break>

This tag is supported by both neural and standard TTS formats.

To add a pause to your text, use the <break> tag. You can set a pause based on strength (equivalent to the pause after a comma, a sentence, or a paragraph), or you can set it to a specific length of time in seconds or milliseconds. If you don't specify an attribute to determine the pause length, Amazon Polly uses the default, which is <break strength="medium" />, which adds a pause the length of a pause after a comma.

strength attribute values:

• none: No pause. Use none to remove a normally occurring pause, such as after a period.

• x-weak: Has the same strength as none, no pause.
• weak: Sets a pause of the same duration as the pause after a comma.
• medium: Has the same strength as weak.
• strong: Sets a pause of the same duration as the pause after a sentence.
• x-strong: Sets a pause of the same duration as the pause after a paragraph.

time attribute values:

• [number]s: The duration of the pause, in seconds. The maximum duration is 10s.

• [number]ms: The duration of the pause, in milliseconds. The maximum duration is 10000ms.

For example:

```
<speak>
  Mary had a little lamb <break time="3s"/>Whose fleece was white as snow.
</speak>
```

If you don't use an attribute with the break tag, the result varies depending on text:

• If there is no other punctuation next to the break tag, it creates a <break strength="medium"/> (comma-length pause).

• If the tag is next to a comma, it upgrades the tag to a <break strength="strong"/> (sentence-length pause).
• If the tag is next to a period, it upgrades the tag to <break strength="x-strong"/> (paragraph-length pause).

Emphasizing Words

<emphasis>

This tag is supported only by the standard TTS format.

To emphasize words, use the <emphasis> tag. Emphasizing words changes the speaking rate and volume. More emphasis makes Amazon Polly speak the text louder and slower.
Less emphasis makes it speak quieter and faster. To specify the degree of emphasis, use the `level` attribute.

level attribute values:

- **Strong**: Increases the volume and slows the speaking rate so that the speech is louder and slower.

- **Moderate**: Increases the volume and slows the speaking rate, but less than strong. Moderate is the default.

- **Reduced**: Decreases the volume and speeds up the speaking rate. Speech is softer and faster.

  **Note**
  The normal speaking rate and volume for a voice falls between the moderate and reduced levels.

For example:

```xml
<speak>
  I already told you I <emphasis level="strong">really like</emphasis> that person.
</speak>
```

Specifying Another Language for Specific Words

`<lang>`

This tag is supported by both neural and standard TTS formats.

Specify another language for a specific word, phrase, or sentence with the `<lang>` tag. Foreign language words and phrases are generally spoken better when they are enclosed within a pair of `<lang>` tags. To specify the language, use the `xml:lang` attribute. For a complete list of available languages, see Languages Supported by Amazon Polly (p. 18).

Unless you apply the `<lang>` tag, all of the words in the input text are spoken in the language of the voice specified in the `voice-id`. If you apply the `<lang>` tag, the words are spoken in that language.

For example, if the `voice-id` is Joanna (who speaks US English), Amazon Polly speaks the following in the Joanna voice without a French accent:

```xml
<speak>
  Je ne parle pas français.
</speak>
```

If you use the Joanna voice with the `<lang>` tag, Amazon Polly speaks the sentence in the Joanna voice in American-accented French:

```xml
<speak>
  <lang xml:lang="fr-FR">Je ne parle pas français.</lang>.
</speak>
```
Because Joanna is not a native French voice, pronunciation is based on her native language, US English. For example, although perfect French pronunciation features an uvular trill /R/ in the word  français, Joanna’s US English voice pronounces this phoneme as the corresponding sound /r/.

If you use the voice-id of Giorgio, who speaks Italian, with the following text, Amazon Polly speaks the sentence in Giorgio’s voice with an Italian pronunciation:

```xml
<speak>
   Mi piace Bruce Springsteen.
</speak>
```

If you use the same voice with the following `<lang>` tag, Amazon Polly pronounces Bruce Springsteen in Italian-accented English:

```xml
<speak>
   Mi piace <lang xml:lang="en-US">Bruce Springsteen.</lang>
</speak>
```

This tag can also be used as a substitute for the optional DefaultLangCode option when synthesizing speech. However, doing so requires that you format your text using SSML.

### Placing a Custom Tag in Your Text

This tag is supported by both neural and standard TTS formats.

To put a custom tag within the text, use the `<mark>` tag. Amazon Polly takes no action on the tag, but returns the location of the tag in the SSML metadata. This tag can be anything you want to call out, as long as it maintains the following format:

```xml
<mark name="tag_name"/>
```

For example, suppose that the tag name is "animal" and the input text is:

```xml
<speak>
   Mary had a little <mark name="animal"/>lamb.
</speak>
```

Amazon Polly might return the following SSML metadata:

```json
{"time":767,"type":"ssml","start":25,"end":46,"value":"animal"}
```

### Adding a Pause Between Paragraphs

```xml
<p>
</p>
```
This tag is supported by both neural and standard TTS formats.

To add a pause between paragraphs in your text, use the `<p>` tag. Using this tag provides a longer pause than native speakers usually place at commas or the end of a sentence. Use the `<p>` tag to enclose the paragraph:

```xml
<speak>
  <p>This is the first paragraph. There should be a pause after this text is spoken.</p>
  <p>This is the second paragraph.</p>
</speak>
```

This is equivalent to specifying a pause using `<break strength="x-strong"/>`.

### Using Phonetic Pronunciation

The `<phoneme>` tag is supported by both neural and standard TTS formats.

To make Amazon Polly use phonetic pronunciation for specific text, use the `<phoneme>` tag. Two attributes are required with the `<phoneme>` tag. They indicate the phonetic alphabet Amazon Polly uses and the phonetic symbols of the corrected pronunciation:

- **alphabet**
  - `ipa` — Indicates that the International Phonetic Alphabet (IPA) will be used.
  - `x-sampa` — Indicates that the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) will be used.
- **ph**
  - Specifies the phonetic symbols for pronunciation. For more information, see Phoneme and Viseme Tables for Supported Languages (p. 19)

With the `<phoneme>` tag, Amazon Polly uses the pronunciation specified by the `ph` attribute instead of the standard pronunciation associated by default with the language used by the selected voice.

For instance, the word "pecan" can be pronounced two ways. In the following example, "pecan" is assigned a different pronunciation in each line. Amazon Polly pronounces pecan as specified in the `ph` attributes, instead of using the default pronunciation.

#### International Phonetic Alphabet (IPA)

```xml
<speak>
  You say, <phoneme alphabet="ipa" ph="p##k##n">pecan</phoneme>.

  I say, <phoneme alphabet="ipa" ph="#pi.kæn">pecan</phoneme>.
</speak>
```

#### Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA)

```xml
<speak>
  You say, <phoneme alphabet='x-sampa' ph='pI"kA:n'>pecan</phoneme>.

  I say, <phoneme alphabet='x-sampa' ph='"pi.k{n'>pecan</phoneme>.
</speak>
```
Additionally, Mandarin Chinese uses Pinyin for phonetic pronunciation.

Pinyin

```xml
<speak>
  ## <phoneme alphabet="x-amazon-pinyin" ph="bo2">#</phoneme>#
  ## <phoneme alphabet="x-amazon-pinyin" ph="bao2">#</phoneme>#
</speak>
```

**Controlling Volume, Speaking Rate, and Pitch**

```xml
<prosody>

Prosody tag attributes are fully supported by the standard TTS voices. Neural voices support the volume and rate attributes, but don't support the pitch attribute.

To control the volume, rate, or pitch of your selected voice, use the prosody tag.

Volume, speech rate, and pitch are dependent on the specific voice selected. In addition to differences between voices for different languages, there are differences between individual voices speaking the same language. Because of this, while attributes are similar across all languages, there are clear variations from language to language and no absolute value is available.

The prosody tag has three attributes, each of which has several available values to set the attribute. Each attribute uses the same syntax:

```xml
<prosody attribute="value"/>
```

- **volume**
  - **default**: Resets volume to the default level for the current voice.
  - **silent, x-soft, soft, medium, loud, x-loud**: Sets the volume to a predefined value for the current voice.
  - **+n dB, -n dB**: Changes volume relative to the current level. A value of 0 dB means no change, +6 dB means approximately twice the current volume, and -6 dB means approximately half the current volume.

For example, you could set the volume for a passage as follows:

```xml
<speak>
  Sometimes it can be useful to <prosody volume="loud">increase the volume for a specific speech.</prosody>
</speak>
```

Or you could set it this way:

```xml
<speak>
  And sometimes a lower volume <prosody volume="-6dB">is a more effective way of interacting with your audience.</prosody>
</speak>
```

- **rate**
  - **x-slow, slow, medium, fast, x-fast**: Sets the pitch to a predefined value for the selected voice.
• **rate**
  - A non-negative percentage change in the speaking rate. For example, a value of 100% means no change in speaking rate, a value of 200% means a speaking rate twice the default rate, and a value of 50% means a speaking rate of half the default rate. This value has a range of 20-200%.
  
  For example, you could set the speech rate for a passage as follows:
  
  ```xml
  <speak>
    For dramatic purposes, you might wish to <prosody rate="slow">slow up the speaking rate of your text.</prosody>
  </speak>
  
  Or you could set it this way:
  
  ```xml
  <speak>
    Although in some cases, it might help your audience to <prosody rate="85%">slow the speaking rate slightly to aid in comprehension.</prosody>
  </speak>
  
  • **pitch**
  - **default**: Resets pitch to the default level for the current voice.
  - **x-low, low, medium, high, x-high**: Sets the pitch to a predefined value for the current voice.
  - **+n% or -n%**: Adjusts pitch by a relative percentage. For example, a value of +0% means no baseline pitch change, +5% gives a little higher baseline pitch, and -5% results in a little lower baseline pitch.
  
  For example, you could set the pitch for a passage as follows:
  
  ```xml
  <speak>
    Do you like synthesized speech <prosody pitch="high">with a pitch that is higher than normal?</prosody>
  </speak>
  
  Or you could set it this way:
  
  ```xml
  <speak>
    Or do you prefer your speech <prosody pitch="-10%">with a somewhat lower pitch?</prosody>
  </speak>
  
  The `<prosody>` tag must contain at least one attribute, but can include more within the same tag.
  
  ```xml
  <speak>
    Each morning when I wake up, <prosody volume="loud" rate="x-slow">I speak quite slowly and deliberately until I have my coffee.</prosody>
  </speak>
  
  It can also be combined with nested tags, as follows:
  
  ```xml
  <speak>
    For dramatic purposes, you might wish to <prosody rate="slow">slow up the speaking rate of your text.</prosody>
  </speak>
Sometimes combining attributes can change the impression your audience has of a voice as well.

**Setting a Maximum Duration for Synthesized Speech**

This tag is currently supported only by the standard TTS format.

To control how long you want a speech to take when it is synthesized, use the `<prosody>` tag with the `amazon:max-duration` attribute.

The duration of synthesized speech varies slightly, depending on the voice you select. This can make it difficult to match synthesized speech with visuals or other activities that require precise timing. This issue is magnified for translation applications because the time it takes to say particular phrases can vary widely with different languages.

The `<prosody amazon:max-duration>` tag matches synthesized speech to the amount of time you want it to take (the duration).

This tag uses the following syntax:

```
<prosody amazon:max-duration="time duration">
```

With the `<prosody amazon:max-duration>` tag, you can specify duration in either seconds or milliseconds:

- **ns**: the maximum duration in seconds
- **nms**: the maximum duration in milliseconds

For example, the following spoken text has a maximum duration of 2 seconds:

```xml
<speak>

    <prosody amazon:max-duration="2s">
        Human speech is a powerful way to communicate.
    </prosody>

</speak>
```

Text placed within the tag, it doesn't exceed the specified duration. If the chosen voice or language would normally take longer than that duration, Amazon Polly speeds up the speech so that it fits into the specified duration.

If the specified duration is longer than it takes to read the text at a normal rate, Amazon Polly reads the speech normally. It doesn't slow down the speech or add silence, so the resulting audio is shorter than requested.

**Note**

Amazon Polly increases the speed no more than 5 times the normal rate. If text is spoken faster than this, it usually doesn't make sense. If a speech cannot fit within your specified duration even when speeded up to the maximum, the audio will be speeded up but will last longer than the specified duration.
You can include a single sentence or multiple sentences within a `<prosody amazon:max-duration>` tag, and you can use multiple `<prosody amazon:max-duration>` tags within your text.

For example:

```xml
<speak>
  <prosody amazon:max-duration="2400ms">
    Human speech is a powerful way to communicate.
  </prosody>
  <break strength="strong"/>
  <prosody amazon:max-duration="5100ms">
    Even a simple 'Hello' can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
  <break strength="strong"/>
  <prosody amazon:max-duration="8900ms">
    We naturally understand this information, which is why speech is ideal for creating applications where a screen isn’t practical or possible, or simply isn’t convenient.
  </prosody>
</speak>
```

Using the `<prosody amazon:max-duration>` tag can increase latency when Amazon Polly returns synthesized speech. The degree of latency depends on the passage and its length. We recommend using text comprised of relatively short text passages.

**Limitations**

There are limitations both in how you use `<prosody amazon:max-duration>` tag and in how it works with other SSML tags:

- The text inside a `<prosody amazon:max-duration>` tag can’t be longer than 1500 characters.
- You can’t nest `<prosody amazon:max-duration>` tags. If you put one `<prosody amazon:max-duration>` tag inside another, Amazon Polly ignores the inner tag. For example, in the following, the `<prosody amazon:max-duration="5s">` tag is ignored:

```xml
<speak>
  <prosody amazon:max-duration="16s">
    Human speech is a powerful way to communicate.
  </prosody>
  <prosody amazon:max-duration="5s">
    Even a simple 'Hello' can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
  We naturally understand this information, which is why speech is ideal for creating applications where a screen isn’t practical or possible, or simply isn’t convenient.
</speak>
```
In the following example, Amazon Polly ignores the $\texttt{<prosody rate="2">}$ tag:

```xml
<speak>
  <prosody amazon:max-duration="7500ms">
    Human speech is a powerful way to communicate.
    <prosody rate="2">
      Even a simple ‘Hello’ can convey a lot of information depending on the pitch, intonation, and tempo.
    </prosody>
  </prosody>
</speak>
```

**Pauses and max-duration**

When using $\texttt{max-duration}$ tag, you can still insert pauses within your text. However, Amazon Polly includes the length of the pause when calculating the maximum duration for speech. Additionally, Amazon Polly preserves the short pauses that occur where commas and periods are placed within a passage and includes in the maximum duration.

For example, in the following block, the 600 millisecond break and the breaks caused by the commas and periods occur within the 8-second speech:

```xml
<speak>
  <prosody amazon:max-duration="8s">
    Human speech is a powerful way to communicate.
    <break time="600ms"/>
    Even a simple ‘Hello’ can convey a lot of information depending on the pitch, intonation, and tempo.
  </prosody>
</speak>
```

**Adding a Pause Between Sentences**

$\texttt{<s>}$

This tag is supported by both neural and standard TTS formats.

To add a pause between lines or sentences in your text, use the $\texttt{<s>}$ tag. Using this tag has the same effect as:

- Ending a sentence with a period ($.$)

- Specifying a pause with $\texttt{<break strength="strong"/>}$

Unlike the $\texttt{<break>}$ tag, the $\texttt{<s>}$ tag encloses the sentence. This is useful for synthesizing speech that is organized in lines, rather than sentence, such as poetry.

In the following example, the $\texttt{<s>}$ tag creates a short pause after both the first and second sentences. The final sentence has no $\texttt{<s>}$ tag, but it is also followed by a short pause because it ends with a period.

```xml
<speak>
  <s>Mary had a little lamb</s>
  <s>Whose fleece was white as snow</s>
</speak>
```
Controlling How Special Types of Words Are Spoken

<say-as>
Except for the "characters" or "spell-out" feature, this tag is supported by both neural and standard TTS formats. If this SSML code is encountered by Amazon Polly at run-time, the affected sentence will be synthesized using the related standard voice. Please note, however, that this sentence will still be billed as if it uses the neural voice.

Use the <say-as> tag with the interpret-as attribute to tell Amazon Polly how to say certain characters, words, and numbers. This enables you to provide additional context to eliminate any ambiguity on how Amazon Polly should render the text.

The say-as tag uses one attribute, <interpret-as>, which uses a number of possible available values. Each uses the same syntax:

<say-as interpret-as="value">[text to be interpreted]</say-as>

The following values are available with interpret-as:

- characters or spell-out: Spells out each letter of the text, as in a-b-c.
  
  **Note**
  
  This option is not currently supported using neural voices. If this SSML code is encountered by Amazon Polly at run-time, the affected sentence will be synthesized using the related standard voice. Please note, however, that this sentence will still be billed as if it uses the neural voice.

- cardinal or number: Interprets the numerical text as a cardinal number, as in 1,234.
- ordinal: Interprets the numerical text as an ordinal number, as in 1,234th.
- digits: Spells out each digit individually, as in 1-2-3-4.
- fraction: Interprets the numerical text as a fraction. This works for both common fractions such as 3/20, and mixed fractions, such as 2 ½. See below for more information.
- unit: Interprets a numerical text as a measurement. The value should be either a number or a fraction followed by a unit with no space in between as in 1/2 inch, or by just a unit, as in 1 meter.
- date: Interprets the text as a date. The format of the date must be specified with the format attribute. See below for more information.
- time: Interprets the numerical text as duration, in minutes and seconds, as in 1'21".
- address: Interprets the text as part of a street address.
- expletive: "Beeps out" the content included within the tag.
- telephone: Interprets the numerical text as a 7-digit or 10-digit telephone number, as in 2025551212. You can also use this value for handle telephone extensions, as in 2025551212x345. See below for more information.

  **Note**
  
  Currently the telephone option is not available for all languages. However, it is available for voices speaking English language variants (en-AU, en-GB, en-IN, en-US, and en-GB-WLS), Spanish language variants (es-ES, es-MX, and es-US), French language variants (fr-FR and fr-CA), and Portuguese variants (pt-BR and pt-PT), as well as German (de-DE), Italian (it-IT), Japanese (ja-JP), and Russian
It should also be noted that in some cases, languages such as Arabic (ru-RU) automatically handle the number set as a telephone number and so do not actually implement the telephone SSML tag.

**Fractions**

Amazon Polly interprets values within the `say-as` tag that have the `interpret-as="fraction"` attribute as common fractions. The following is the syntax for fractions:

- **Fraction**

  Syntax: `cardinal number/cardinal number`, such as 2/9.
  For example: `<say-as interpret-as="fraction">2/9</say-as>` is pronounced “two ninths.”

- **Non-negative Mixed Number**

  Syntax: `cardinal number+cardinal number/cardinal number`, such as 3+1/2.
  For example, `<say-as interpret-as="fraction">3+1/2</say-as>` is pronounced “three and a half.”

  **Note**

  There must be a `+` between the "3" and the "1/2". Amazon Polly doesn't support a mixed number without the `+`, such as "3 1/2".

**Dates**

When `interpret-as` is set to `date`, you also need to indicate the format of the date. This uses the following syntax:

```
<say-as interpret-as="date" format="format">[date]</say-as>
```

For example:

```xml
<speak>
  I was born on <say-as interpret-as="date" format="mdy">12-31-1900</say-as>.
</speak>
```

The following formats can be used with the `date` attribute.

- `mdy`: Month-day-year.
- `dmy`: Day-month-year.
- `ymd`: Year-month-day.
- `md`: Month-day.
- `dm`: Day-month.
- `ym`: Year-month.
- `my`: Month-year.
- `d`: Day.
- `m`: Month.
- `y`: Year.
- `yyyymmdd`: Year-month-day. If you use this format, you can make Amazon Polly skip parts of the date using question marks.
For example, Amazon Polly renders the following as "September 22nd":

```xml
<say-as interpret-as="date">0922</say-as>
```

_Format is not needed._

**Telephone**

Amazon Polly attempts to interpret the text you provide correctly based on the text's formatting even without the `<say-as>` tag. For example, if your text includes "202-555-1212," Amazon Polly interprets it as a 10-digit telephone number and says each digit individually, with a brief pause for each dash. In this case, you don't need to use `<say-as interpret-as="telephone">`. However, if you provide the text “2025551212” and want Amazon Polly to say it as a phone number, you would specify `<say-as interpret-as="telephone">`.

The logic for interpreting each element is language-specific. For example, US and UK English differ in how phone numbers are pronounced (in UK English, sequences of the same digit are grouped together, as in "double five" or "triple four"). To see the difference, test the following example with a US voice and with a UK voice:

```xml
<speak>
    Richard's number is <say-as interpret-as="telephone">2122241555</say-as>
</speak>
```

**Pronouncing Acronyms and Abbreviations**

 `<sub>`

This tag is supported by both neural and standard TTS formats.

Use the `<sub>` tag with the `alias` attribute to substitute a different word (or pronunciation) for selected text such as an acronym or abbreviation.

This uses the syntax:

```xml
<sub alias="new word">abbreviation</sub>
```

In the following example, the name "Mercury" is substituted for the element's chemical symbol to make the audio content clearer.

```xml
<speak>
    My favorite chemical element is <sub alias="Mercury">Hg</sub>, because it looks so shiny.
</speak>
```

**Improving Pronunciation by Specifying Parts of Speech**

 `<w>`
This tag is supported by both neural and standard TTS formats.

You can use the `<w>` tag to customize the pronunciation of words by specifying the word’s part of speech or alternate meaning. This is done using the `role` attribute.

This tag uses the following syntax:

```xml
<w role="attribute">text</w>
```

The following values can be used for the `role` attribute:

To specify the part of speech:

- `amazon:VB`: interprets the word as a verb (present simple).
- `amazon:VBD`: interprets the word as past tense verb.
- `amazon:DT`: interprets the word as a determiner.
- `amazon:IN`: interprets the word as a proposition.
- `amazon:JJ`: interprets the word as an adjective.
- `amazon:NN`: interprets the word as a noun.

For example, depending on its part of speech, the US English pronunciation of the word "read" varies based on the tag:

```xml
<speak>
  The word <say-as interpret-as="characters">read</say-as> may be interpreted
  as either the present simple form <w role="amazon:VB">read</w>, or the past
  participle form <w role="amazon:VBD">read</w>.
</speak>
```

To specify a specific meaning:

- `amazon:DEFAULT`: uses the default sense of the word.
- `amazon:SENSE_1`: uses the non-default sense of the word when present. For example, the noun "bass" is pronounced differently depending on its meaning. The default meaning is the lowest part of the musical range. The alternate meaning is a species of freshwater fish, also called "bass" but pronounced differently. Using `<w role="amazon:SENSE_1">bass</w>` renders the non-default pronunciation (freshwater fish) for the audio text.

This difference in pronunciation and meaning can be heard if you synthesize the following:

```xml
<speak>
  Depending on your meaning, the word <say-as interpret-as="characters">bass</say-as>
  may be interpreted as either a musical element: bass, or as its alternative
  meaning,
  a freshwater fish <w role="amazon:SENSE_1">bass</w>.
</speak>
```

**Note**

Some languages may have a different selection of supported parts of speech.
This tag is supported only by the standard TTS format.

Natural-sounding speech includes both correctly spoken words and breathing sounds. By adding breathing sounds to synthesized speech, you can make it sound more natural. The `<amazon:breath>` and `<amazon:auto-breaths>` tags provide breaths. You have the following options:

- Manual mode: you set the location, length, and volume of a breath sound within the text
- Automated mode: Amazon Polly automatically inserts breathing sounds into the speech output
- Mixed mode: both you and Amazon Polly add breathing sounds

**Manual Mode**

In manual mode, you place the `<amazon:breath/>` tag in the input text where you want to locate a breath. You can customize the length and volume of breaths with the `duration` and `volume` attributes, respectively:

- **duration**: Controls the length of the breath. Valid values are: `default`, `x-short`, `short`, `medium`, `long`, `x-long`. The default value is `medium`.
- **volume**: Controls how loud breathing sounds. Valid values are: `default`, `x-soft`, `soft`, `medium`, `loud`, `x-loud`. The default value is `medium`.

**Note**

The exact length and volume of each attribute value is dependent on the specific Amazon Polly voice used.

To set a breath sound using the defaults, use `<amazon:breath/>` without attributes.

For example, to use attributes to set the duration and volume for a breath to medium, you would set the attributes as follows:

```
<speak>
    Sometimes you want to insert only <amazon:breath duration="medium" volume="x-loud"]/a single breath.
</speak>
```

To use the defaults, you would just use the tag:

```
<speak>
    Sometimes you need <amazon:breath/> to insert one or more average breathes <amazon:breath/> so that the text sounds correct.
</speak>
```

You can add individual breathing sounds within a passage, as follows:

```
<speak>
    <amazon:breath duration="long" volume="x-loud"/> <prosody rate="120%">
        Wow! <amazon:breath duration="long" volume="loud"/> <prosody volume="loud">
            That was quite fast <amazon:breath duration="medium" volume="x-loud"/>.</prosody>
    </prosody>
</speak>
```
Automated Mode

In automated mode, you use the `<amazon:auto-breaths>` tag to tell Amazon Polly to automatically create breathing noises at appropriate intervals. You can set the frequency of the intervals, their volume, and their duration. Place the `</amazon:auto-breaths>` tag at the beginning of the text that you want to apply automated breathing to and then close the tag at the end.

**Note**

Unlike the manual mode tag, `<amazon:breath/>`, the `<amazon:auto-breaths>` tag requires a closing tag (`</amazon:auto-breaths>`).

You can use the following optional attributes with the `<amazon:auto-breaths>` tag:

- **volume**: Controls how loud the breathing sounds. Valid values are: default, x-soft, soft, medium, loud, x-loud. The default value is medium.

- **frequency**: Controls how often breathing sounds occur in the text. Valid values are: default, x-low, low, medium, high, x-high. The default value is medium.

- **duration**: Controls the length of the breath. Valid values are: default, x-short, short, medium, long, x-long. The default value is medium.

By default, the frequency of breathing sounds depends on the input text. However, breathing sounds often occur after commas and periods.

The following examples show how to use the `<amazon:auto-breaths>` tag. To decide which options to use for your content, copy the applicable examples to the Amazon Polly console and listen to the differences.

- Using automated mode without optional parameters.

  ```xml
  <speak>
  <amazon:auto-breaths>Amazon Polly is a service that turns text into lifelike speech, allowing you to create applications that talk and build entirely new categories of speech-enabled products. Amazon Polly is a text-to-speech service that uses advanced deep learning technologies to synthesize speech that sounds like a human voice. With dozens of lifelike voices across a variety of languages, you can select the ideal voice and build speech-enabled applications that work in many different countries.</amazon:auto-breaths>
  </speak>
  ```

- Using automated mode with volume control. The unspecified parameters (duration and frequency) are set to the default values (medium).

  ```xml
  <speak>
  <amazon:auto-breaths volume="x-soft">Amazon Polly is a service that turns text into lifelike speech, allowing you to create applications that talk and build entirely new categories of speech-enabled products. Amazon Polly is a text-to-speech service, that uses advanced deep learning technologies to synthesize speech that sounds like a human voice. With dozens of lifelike voices across a variety of languages, you can select the ideal voice and build speech-enabled applications that work in many different countries.</amazon:auto-breaths>
  </speak>
  ```
Newscaster speaking style

<amazon:domain name="news">

The newscaster style is available only for the Matthew or Joanna voices, which are available only in American English (en-US), Lupe, in US Spanish (es-US) and Amy, in British English (en-GB). It is only supported when using Neural format.

To use the newscaster style, you use SSML tags and the following syntax:

<amazon:domain name="news">text</amazon:domain>

For example, you might use the newscaster style with the Amy voice as follows:

<amazon:domain name="news">

From the Tuesday, April 16th, 1912 edition of The Guardian newspaper:

The maiden voyage of the White Star liner Titanic, the largest ship ever launched, has ended in disaster.

</amazon:domain>
The Titanic started her trip from Southampton for New York on Wednesday. Late on
Sunday night she struck
an iceberg off the Grand Banks of Newfoundland. By wireless telegraphy she sent
out signals of distress,
and several liners were near enough to catch and respond to the call.
</speak>

Adding Dynamic Range Compression

<amazon:effect name="drc">
This tag is supported by both neural and standard TTS formats.

Depending on the text, language, and voice used in an audio file, the sounds range from soft
to loud. Environmental sounds, such as the sound of a moving vehicle, can often mask the
softer sounds, which makes the audio track difficult to hear clearly. To enhance the volume of
certain sounds in your audio file, use the dynamic range compression (drc) tag.

The drc tag sets a midrange "loudness" threshold for your audio, and increases the volume
(the gain) of the sounds around that threshold. It applies the greatest gain increase closest to
the threshold, and the gain increase is lessened farther away from the threshold.

This makes the middle-range sounds easier to hear in a noisy environment, which makes the
entire audio file clearer.

The drc tag is a Boolean parameter (it's either present or it isn't). It uses the syntax:
<amazon:effect name="drc"> and is closed with </amazon:effect>.

You can use the drc tag with any voice or language supported by Amazon Polly. You can
apply it to an entire section of the recording, or for only a few words. For example:

<speak>

Some audio is difficult to hear in a moving vehicle, but <amazon:effect name="drc">
this audio

is less difficult to hear in a moving vehicle.</amazon:effect>
</speak>

Note
When you use "drc" in theamazon:effect syntax, it is case-sensitive.

Using drc with the prosody volume Tag
When you use the `drc` and `prosody volume` tags together, Amazon Polly applies the `drc` tag first, increasing the middle-range sounds (those near the threshold). It then applies the `prosody volume` tag and further increases the volume of the entire audio track evenly.

To use the tags together, nest one inside the other. For example:

```xml
<speak>
  <prosody volume="loud">This text needs to be understandable and loud.
  <amazon:effect name="drc">
    This text also needs to be more understandable in a moving car.
  </amazon:effect>
</prosody>
</speak>
```

In this text, the `prosody volume` tag increases the volume of the entire passage to "loud." The `drc` tag enhances the volume of the middle-range values in the second sentence.

**Note**
When using the `drc` and `prosody volume` tags together, use standard XML practices for nesting tags.

### Speaking Softly

```xml
<amazon:effect phonation="soft">
  This tag is currently supported only by the standard TTS format.

  To specify that input text should be spoken in a softer-than-normal voice, use the <amazon:effect phonation="soft"> tag.

  This uses the syntax:

  ```xml
  <amazon:effect phonation="soft">text</amazon:effect>
  ```

  For example, you might use this tag with the Matthew voice as follows:

  ```xml
  <speak>
    This is Matthew speaking in my normal voice. <amazon:effect phonation="soft">
      
  ```
```
```xml
Controlling Timbre

This tag is currently supported only by the standard TTS format.

Timbre is the tonal quality of a voice that helps you tell the difference between voices, even when they have the same pitch and loudness. One of the most important physiological features that contributes to speech timbre is the length of the vocal tract. The vocal tract is a cavity of air that spans from the top of the vocal folds up to the edge of the lips.

To control the timbre of output speech in Amazon Polly, use the `vocal-tract-length` tag.

To change timbre, use the following values:

- `+n%` or `-n%`: Adjusts the vocal tract length by a relative percentage change in the current voice. For example, +4% or -2%. Valid values range from +100% to -50%. Values outside this range are clipped. For example, +111% sounds like +100% and -60% sounds like -50%.

- `n%`: Changes the vocal tract length to an absolute percentage of the tract length of the current voice. For example, 110% or 75%. An absolute value of 110% is equivalent to a relative value of +10%. An absolute value of 100% is the same as the default value for the current voice.

The following example shows how to change the vocal tract length to change timbre:

```
<speak>
  This is my original voice, without any modifications. <amazon:effect vocal-tract-length="+15%">
    Now, imagine that I am much bigger. </amazon:effect> <amazon:effect vocal-tract-length="-15%">
      Or, perhaps you prefer my voice when I'm very small. </amazon:effect> You can also control the timbre of my voice by making minor adjustments. <amazon:effect vocal-tract-length="+10%">
        For example, by making me sound just a little bigger. </amazon:effect>
  </amazon:effect>
  Or, making me sound only somewhat smaller. </amazon:effect>
</speak>
```
The following example shows how to combine tags.

```xml
<speak>
    The pitch and timbre of a person's voice are connected in human speech.
    
    <amazon:effect vocal-tract-length="-15%">
        If you are going to reduce the vocal tract length,
    </amazon:effect>

    <amazon:effect vocal-tract-length="-15%">
        <prosody pitch="+20%">
            you might consider increasing the pitch, too.
        </prosody>
    </amazon:effect>

    <amazon:effect vocal-tract-length="+15%">
        If you choose to lengthen the vocal tract,
    </amazon:effect>

    <amazon:effect vocal-tract-length="+15%">
        <prosody pitch="-10%">
            you might also want to lower the pitch.
        </prosody>
    </amazon:effect>
</speak>
```

**Whispering**

```xml
<amazon:effect name="whispered">
This tag is currently supported only by the standard TTS format.

This tag indicates that the input text should be spoken in a whispered voice rather than as normal speech. This can be used with any of the voices in the Amazon Polly Text-to-Speech portfolio.

This uses the following syntax:

```xml
<amazon:effect name="whispered">
    text
</amazon:effect>
```

For example:

```xml
<speak>
    <amazon:effect name="whispered">If you make any noise, </amazon:effect>

    she said, <amazon:effect name="whispered">they will hear us.</amazon:effect>
</speak>
```

In this case, the synthesized speech spoken by the character is whispered, but the phrase "she said" is spoken in the normal synthesized speech of the selected Amazon Polly voice.

You can enhance the "whispered" effect by slowing down the prosody rate by up to 10%, depending on the effect you want.

For example:

```xml
<speak>
    <amazon:effect name="whispered">
        <prosody rate="-10%">
            the sound is slower and quieter than normal speech
        </prosody>
    </amazon:effect>
</speak>
```
Example Applications

This section contains additional examples, in the form of example applications which can be used to explore Amazon Polly.

Example Applications by Programming Language

- Python Example (HTML5 Client and Python Server) (p. 194)
- Java Example (p. 204)
- iOS Example (p. 208)
- Android Example (p. 210)

Python Example (HTML5 Client and Python Server)

This example application consists of the following:

- An HTTP 1.1 server using the HTTP chunked transfer coding (see Chunked Transfer Coding)
- A simple HTML5 user interface that interacts with the HTTP 1.1 server (shown below):

![HTML5 User Interface](image)

```python
polly_client = boto3.Session(
    aws_access_key_id=,
    aws_secret_access_key=,
    region_name='us-west-2').client('polly')

response = polly_client.synthesize_speech(VoiceId='Joanna',
                                          OutputFormat='mp3',
                                          Text = 'This is a sample text to be synthesized.',
                                          Engine = 'neural')

file = open('speech.mp3', 'wb')
file.write(response['AudioStream'].read())
file.close()
```
The goal of this example is to show how to use Amazon Polly to stream speech from a browser-based HTML5 application. Consuming the audio stream produced by Amazon Polly as the text gets synthesized is the recommended approach for use cases where responsiveness is an important factor (for example, dialog systems, screen readers, etc.).

To run this example application you need the following:

- Web browser compliant with the HTML5 and EcmaScript5 standards (for example, Chrome 23.0 or higher, Firefox 21.0 or higher, Internet Explorer 9.0, or higher)
- Python version greater than 3.0

To test the application

1. Save the server code as server.py. For the code, see Python Example: Python Server Code (server.py) (p. 198).
2. Save the HTML5 client code as index.html. For the code, see Python Example: HTML5 User Interface (index.html) (p. 195).
3. Run the following command from the path where you saved server.py to start the application (on some systems you might need to use python3 instead of python when running the command).

   ```bash
   $ python server.py
   ```

   After the application starts, a URL appears on the terminal.

4. Open the URL shown in the terminal in a web browser.

5. To listen to speech, choose a voice from the list, type some text, and then choose Read. The speech starts playing as soon as Amazon Polly transfers the first usable chunk of audio data.
6. To stop the Python server when you’re finished testing the application, press Ctrl+C in the terminal where the server is running.

   **Note**
   The server creates a Boto3 client using the AWS SDK for Python (Boto). The client uses the credentials stored in the AWS config file on your computer to sign and authenticate the requests to Amazon Polly. For more information on how to create the AWS config file and store credentials, see Configuring the AWS Command Line Interface in the AWS Command Line Interface User Guide.

**Python Example: HTML5 User Interface (index.html)**

This section provides the code for the HTML5 client described in Python Example (HTML5 Client and Python Server) (p. 194).

```html
<html>
<head>
  <title>Text-to-Speech Example Application</title>
  <script>
    /*
     * This sample code requires a web browser with support for both the
     * HTML5 and EcmaScript 5 standards; the following is a non-comprehensive
     * list of compliant browsers and their minimum version:
     * - Chrome 23.0+
     */
  </script>
</head>
<body>
```
* - Firefox 21.0+
* - Internet Explorer 9.0+
* - Edge 12.0+
* - Opera 15.0+
* - Safari 6.1+
* - Android (stock web browser) 4.4+
* - Chrome for Android 51.0+
* - Firefox for Android 48.0+
* - Opera Mobile 37.0+
* - iOS (Safari Mobile and Chrome) 3.2+
* - Internet Explorer Mobile 10.0+
* - Blackberry Browser 10.0+

// Mapping of the OutputFormat parameter of the SynthesizeSpeech API
// and the audio format strings understood by the browser
var AUDIO_FORMATS = {
'ogg_vorbis': 'audio/ogg',
'mp3': 'audio/mpeg',
'pcm': 'audio/wave; codecs=1'
};

/**
* Handles fetching JSON over HTTP
*/
function fetchJSON(method, url, onSuccess, onError) {
var request = new XMLHttpRequest();
request.open(method, url, true);
request.onload = function () {
  // If loading is complete
  if (request.readyState === 4) {
    // if the request was successful
    if (request.status === 200) {
      var data;
      try {
        data = JSON.parse(request.responseText);
      } catch (error) {
        onError(request.status, error.toString());
      }
      onSuccess(data);
    } else {
      onError(request.status, request.responseText)
    }
  }
}
request.send();

/**
* Returns a list of audio formats supported by the browser
*/
function getSupportedAudioFormats(player) {
  return Object.keys(AUDIO_FORMATS).filter(function (format) {
    var supported = player.canPlayType(AUDIO_FORMATS[format]);
    return supported === 'probably' || supported === 'maybe';
  });
}

// Initialize the application when the DOM is loaded and ready to be
// manipulated
document.addEventListener("DOMContentLoaded", function () {
var input = document.getElementById('input'),
voiceMenu = document.getElementById('voice'),
text = document.getElementById('text'),
player = document.getElementById('player'),
supportFormat = getSupportedAudioFormats(player);

// Display a message and don't allow submitting the form if the
// browser doesn't support any of the available audio formats
if (supportedFormats.length === 0) {
  submit.disabled = true;
  alert('The web browser in use does not support any of the ' +
    'available audio formats. Please try with a different ' +
    'one.');
}

// Play the audio stream when the form is submitted successfully
input.addEventListener('submit', function (event) {
  // Validate the fields in the form, display a message if
  // unexpected values are encountered
  if (voiceMenu.selectedIndex <= 0 || text.value.length === 0) {
    alert('Please fill in all the fields.' );
  } else {
    var selectedVoice = voiceMenu
      .options[voiceMenu.selectedIndex].value;

    // Point the player to the streaming server
    player.src = '/read?voiceId=' +
      encodeURIComponent(selectedVoice) +
      '&text=' + encodeURIComponent(text.value) +
      '&outputFormat=' + supportedFormats[0];
    player.play();
  }
  // Stop the form from submitting,
  // Submitting the form is allowed only if the browser doesn't
  // support Javascript to ensure functionality in such a case
  event.preventDefault();
});

// Load the list of available voices and display them in a menu
fetchJSON('GET', '/voices',
  // If the request succeeds
  function (voices) {
    var container = document.createDocumentFragment();

    // Build the list of options for the menu
    voices.forEach(function (voice) {
      var option = document.createElement('option');
      option.value = voice['Id'];
      option.innerHTML = voice['Name'] + ' (' +
        voice['Gender'] + ', ' +
        voice['LanguageName'] + ')';
      container.appendChild(option);
    });

    // Add the options to the menu and enable the form field
    voiceMenu.appendChild(container);
    voiceMenu.disabled = false;
  },
  // If the request fails
  function (status, response) {
    // Display a message in case loading data from the server
    // fails
    alert(status + ' - ' + response);
  });
Python Example: Python Server Code (server.py)

This section provides the code for the Python server described in Python Example (HTML5 Client and Python Server) (p. 194).

```
Example Python 2.7+/3.3+ Application

This application consists of a HTTP 1.1 server using the HTTP chunked transfer coding (https://tools.ietf.org/html/rfc2616#section-3.6.1) and a minimal HTML5 user interface that interacts with it.

The goal of this example is to start streaming the speech to the client (the HTML5 web UI) as soon as the first consumable chunk of speech is returned in order to start playing the audio as soon as possible. For use cases where low latency and responsiveness are strong requirements,
```
this is the recommended approach.

The service documentation contains examples for non-streaming use cases where
waiting for the speech synthesis to complete and fetching the whole audio stream
at once are an option.

To test the application, run 'python server.py' and then open the URL
displayed in the terminal in a web browser (see index.html for a list of
supported browsers). The address and port for the server can be passed as
parameters to server.py. For more information, run: 'python server.py -h'

```python
from argparse import ArgumentParser
from collections import namedtuple
from contextlib import closing
from io import BytesIO
from json import dumps as json_encode
import os
import sys
if sys.version_info >= (3, 0):
    from http.server import BaseHTTPRequestHandler, HTTPServer
    from socketserver import ThreadingMixIn
    from urllib.parse import parse_qs
else:
    from BaseHTTPServer import BaseHTTPRequestHandler, HTTPServer
    from SocketServer import ThreadingMixIn
    from urlparse import parse_qs
from boto3 import Session
from botocore.exceptions import BotoCoreError, ClientError
ResponseStatus = namedtuple("HTTPStatus",
    ["code", "message"])
ResponseData = namedtuple("ResponseData",
    ["status", "content_type", "data_stream"])
# Mapping the output format used in the client to the content type for the
# response
AUDIO_FORMATS = {"ogg_vorbis": "audio/ogg",
    "mp3": "audio/mpeg",
    "pcm": "audio/wave; codecs=1"
} CHUNK_SIZE = 1024
HTTP_STATUS = {"OK": ResponseStatus(code=200, message="OK"),
    "BAD_REQUEST": ResponseStatus(code=400, message="Bad request"),
    "NOT_FOUND": ResponseStatus(code=404, message="Not found"),
    "INTERNAL_SERVER_ERROR": ResponseStatus(code=500, message="Internal server
    error")
} PROTOCOL = "http"
ROUTE_INDEX = "/index.html"
ROUTE_VOICES = "/voices"
ROUTE_READ = "/read"

# Create a client using the credentials and region defined in the adminuser
# section of the AWS credentials and configuration files
session = Session(profile_name="adminuser")
polly = session.client("polly")
```

```
```
(code, message, description), see http.server.HTTPStatus
'''
super(HTTPStatusError, self).__init__()
self.code = status.code
def message = status.message
def explain = description

class ThreadedHTTPServer(ThreadingMixIn, HTTPServer):
    """An HTTP Server that handle each request in a new thread""
    daemon_threads = True

class ChunkedHTTPRequestHandler(BaseHTTPRequestHandler):
    """HTTP 1.1 Chunked encoding request handler"
    # Use HTTP 1.1 as 1.0 doesn't support chunked encoding
    protocol_version = "HTTP/1.1"

def query_get(self, queryData, key, default=""):  
    """Helper for getting values from a pre-parsed query string""
    return queryData.get(key, [default])[0]

def do_GET(self):
    """Handles GET requests""
    
    # Extract values from the query string
    path, _, query_string = self.path.partition('?')
    query = parse_qs(query_string)
    response = None

    print(u"[START]: Received GET for %s with query: %s" % (path, query))
    try:
        # Handle the possible request paths
        if path == ROUTE_INDEX:
            response = self.route_index(path, query)
        elif path == ROUTE_VOICES:
            response = self.route_voices(path, query)
        elif path == ROUTE_READ:
            response = self.route_read(path, query)
        else:
            response = self.route_not_found(path, query)

        self.send_headers(response.status, response.content_type)
        self.stream_data(response.data_stream)

    except HTTPStatusError as err:
        # Respond with an error and log debug
        # information
        if sys.version_info >= (3, 0):
            self.send_error(err.code, err.message, err.explain)
        else:
            self.send_error(err.code, err.message)

        self.log_error(u"%s %s %s - [%d] %s", self.client_address[0],
                      self.command, self.path, err.code, err.explain)

    print("[END]")

def route_not_found(self, path, query):
    """Handles routing for unexpected paths""
    raise HTTPStatusError(HTTP_STATUS["NOT_FOUND"], "Page not found")

def route_index(self, path, query):
    """Handles routing for the application's entry point"""
try:
    return ResponseData(status=HTTP_STATUS["OK"], content_type="text_html",
                        # Open a binary stream for reading the index
                        # HTML file
data_stream=open(os.path.join(sys.path[0],
                                           path[1:]), "rb"))
except IOError as err:
    # Couldn't open the stream
    raise HTTPStatusError(HTTP_STATUS["INTERNAL_SERVER_ERROR"],
                           str(err))

def route_voices(self, path, query):
    """Handles routing for listing available voices""
    params = {}
    voices = []

    while True:
        try:
            # Request list of available voices, if a continuation token
            # was returned by the previous call then use it to continue
            # listing
            response = polly.describe_voices(**params)
        except (BotoCoreError, ClientError) as err:
            # The service returned an error
            raise HTTPStatusError(HTTP_STATUS["INTERNAL_SERVER_ERROR"],
                                   str(err))

        # Collect all the voices
        voices.extend(response.get("Voices", []))

        # If a continuation token was returned continue, stop iterating
        # otherwise
        if "NextToken" in response:
            params = {"NextToken": response["NextToken"]}
        else:
            break

    json_data = json_encode(voices)
    bytes_data = bytes(json_data, "utf-8") if sys.version_info >= (3, 0) else bytes(json_data)
    return ResponseData(status=HTTP_STATUS["OK"],
                         content_type="application/json",
                         # Create a binary stream for the JSON data
data_stream=BytesIO(bytes_data))

def route_read(self, path, query):
    """Handles routing for reading text (speech synthesis)""
    # Get the parameters from the query string
    text = self.query_get(query, "text")
    voiceId = self.query_get(query, "voiceId")
    outputFormat = self.query_get(query, "outputFormat")

    # Validate the parameters, set error flag in case of unexpected
    # values
    if len(text) == 0 or len(voiceId) == 0 or 
       outputFormat not in AUDIO_FORMATS:
        raise HTTPStatusError(HTTP_STATUS["BAD_REQUEST"],
                               "Wrong parameters")
    else:
        try:
            # Request speech synthesis
            response = polly.synthesize_speech(Text=text,
                                                 VoiceId=voiceId,
                                                 OutputFormat=outputFormat,
                                                 Engine="neural")

            # Check if an error occurred
            if response["ResponseMetadata"]['HTTPStatusCode'] != 201:
                raise HTTPStatusError(HTTP_STATUS["INTERNAL_SERVER_ERROR"],
                                       str(response["ResponseMetadata"]['HTTPMessage']))

            # Create a binary stream for the speech
            audio_stream = BytesIO(response["AudioStream"])
            return ResponseData(status=HTTP_STATUS["OK"],
                                 content_type="audio_mp3",
                                 # Create a binary stream for the audio
                                 data_stream=audio_stream)
        except (BotoCoreError, ClientError) as err:
            # The service returned an error
            raise HTTPStatusError(HTTP_STATUS["INTERNAL_SERVER_ERROR"],
                                   str(err))
except (BotoCoreError, ClientError) as err:
    # The service returned an error
    raise HTTPStatusError(HTTP_STATUS["INTERNAL_SERVER_ERROR"],
                          str(err))

return ResponseData(status=HTTP_STATUS["OK"],
                    content_type=AUDIO_FORMATS[outputFormat],
                    # Access the audio stream in the response
                    data_stream=response.get("AudioStream"),
                    # scope.
                    )

def send_headers(self, status, content_type):
    """Send out the group of headers for a successful request""
    # Send HTTP headers
    self.send_response(status.code, status.message)
    self.send_header('Content-type', content_type)
    self.send_header('Transfer-Encoding', 'chunked')
    self.send_header('Connection', 'close')
    self.end_headers()

def stream_data(self, stream):
    """Consumes a stream in chunks to produce the response's output'""
    print("Streaming started...")

    if stream:
        # Note: Closing the stream is important as the service throttles on
        # the number of parallel connections. Here we are using
        # contextlib.closing to ensure the close method of the stream object
        # will be called automatically at the end of the with statement's
        # scope.
        with closing(stream) as managed_stream:
            # Push out the stream's content in chunks
            while True:
                data = managed_stream.read(CHUNK_SIZE)
                self.wfile.write(b"%X\r\n%s\r\n" % (len(data), data))

                # If there's no more data to read, stop streaming
                if not data:
                    break

                # Ensure any buffered output has been transmitted and close the
                # stream
                self.wfile.flush()

        print("Streaming completed.")
    else:
        # The stream passed in is empty
        self.wfile.write(b"0\r\n\r\n")
        print("Nothing to stream.")

# Define and parse the command line arguments
cli = ArgumentParser(description='Example Python Application')
cli.add_argument("-p", "--port", type=int, metavar="PORT", dest="port", default=8000)
cli.add_argument("--host", type=str, metavar="HOST", dest="host", default="localhost")
arguments = cli.parse_args()

# If the module is invoked directly, initialize the application
if __name__ == '__main__':
    # Create and configure the HTTP server instance
    server = ThreadedHTTPServer((arguments.host, arguments.port),
                                 ChunkedHTTPRequestHandler)
    print("Starting server, use <Ctrl-C> to stop...")
    print(u"Open {0}:///{1}:{2}{3} in a web browser.".format(PROTOCOL,
                                                             arguments.host,
                                                             arguments.port,)
try:
    # Listen for requests indefinitely
    server.serve_forever()
except KeyboardInterrupt:
    # A request to terminate has been received, stop the server
    print("Shutting down...")
    server.socket.close()
Java Example

This example shows how to use Amazon Polly to stream speech from a Java-based application. The example uses the AWS SDK for Java to read the specified text using a voice selected from a list.

The code shown covers major tasks, but does only minimal error checking. If Amazon Polly encounters an error, the application terminates.

To run this example application, you need the following:

- Java 8 Java Development Kit (JDK)
- AWS SDK for Java
- Apache Maven

To test the application

1. Ensure that the JAVA_HOME environment variable is set for the JDK.

   For example, if you installed JDK 1.8.0_121 on Windows at `C:\Program Files\Java\jdk1.8.0_121`, you would type the following at the command prompt:

   ```
   set JAVA_HOME="C:\Program Files\Java\jdk1.8.0_121"
   ```

   If you installed JDK 1.8.0_121 in Linux at `/usr/lib/jvm/java8-openjdk-amd64`, you would type the following at the command prompt:

   ```
   export JAVA_HOME=/usr/lib/jvm/java8-openjdk-amd64
   ```

2. Set the Maven environment variables to run Maven from the command line.

   For example, if you installed Maven 3.3.9 on Windows at `C:\Program Files\apache-maven-3.3.9`, you would type the following:

   ```
   set M2_HOME="C:\Program Files\apache-maven-3.3.9"
   set M2=%M2_HOME%\bin
   set PATH=%M2%\bin;%PATH%
   ```

   If you installed Maven 3.3.9 on Linux at `/home/ec2-user/opt/apache-maven-3.3.9`, you would type the following:

   ```
   export M2_HOME=/home/ec2-user/opt/apache-maven-3.3.9
   export M2=$M2_HOME/bin
   export PATH=$M2:$PATH
   ```

3. Create a new directory called `polly-java-demo`.

4. In the `polly-java-demo` directory, create a new file called `pom.xml`, and paste the following code into it:

   ```xml
   <project xmlns="http://maven.apache.org/POM/4.0.0"
     <modelVersion>4.0.0</modelVersion>
     <groupId>com.amazonaws.polly</groupId>
     <artifactId>java-demo</artifactId>
     <version>0.0.1-SNAPSHOT</version>
   </project>
   ```
5. Create a new directory called `polly` at `src/main/java/com/amazonaws/demos`.

6. In the `polly` directory, create a new Java source file called `PollyDemo.java`, and paste in the following code:

```java
package com.amazonaws.demos.polly;

import java.io.IOException;
import java.io.InputStream;
import com.amazonaws.ClientConfiguration;
import com.amazonaws.auth.DefaultAWSCredentialsProviderChain;
import com.amazonaws.regions.Region;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.polly.AmazonPollyClient;
import com.amazonaws.services.polly.model.DescribeVoicesRequest;
import com.amazonaws.services.polly.model.DescribeVoicesResult;
import com.amazonaws.services.polly.model.OutputFormat;
import com.amazonaws.services.polly.model.SynthesizeSpeechRequest;
import com.amazonaws.services.polly.model.SynthesizeSpeechResult;
import com.amazonaws.services.polly.model.Voice;
import javazoom.jl.player.advanced.AdvancedPlayer;
import javazoom.jl.player.advanced.PlaybackEvent;
import javazoom.jl.player.advanced.PlaybackListener;

public class PollyDemo {
```
private final AmazonPollyClient polly;
private final Voice voice;
private static final String SAMPLE = "Congratulations. You have successfully built this working demo of Amazon Polly in Java. Have fun building voice enabled apps with Amazon Polly (that's me!), and always look at the AWS website for tips and tricks on using Amazon Polly and other great services from AWS";

public PollyDemo(Region region) {
    // create an Amazon Polly client in a specific region
    polly = new AmazonPollyClient(new DefaultAWSCredentialsProviderChain(),
        new ClientConfiguration());
    polly.setRegion(region);
    // Create describe voices request.
    DescribeVoicesRequest describeVoicesRequest = new DescribeVoicesRequest();
    DescribeVoicesResult describeVoicesResult = polly.describeVoices(describeVoicesRequest);
    // Synchronously ask Amazon Polly to describe available TTS voices.
    DescribeVoicesResult describeVoicesResult = polly.describeVoices(describeVoicesRequest);
    voice = describeVoicesResult.getVoices().get(0);
}

public InputStream synthesize(String text, OutputFormat format) throws IOException {
    SynthesizeSpeechRequest synthReq =
        new SynthesizeSpeechRequest().withText(text).withVoiceId(voice.getId()).
            .withOutputFormat(format).withEngine("neural");
    SynthesizeSpeechResult synthRes = polly.synthesizeSpeech(synthReq);
    return synthRes.getAudioStream();
}

public static void main(String args[]) throws Exception {
    //create the test class
    PollyDemo helloWorld = new PollyDemo(Region.getRegion(Regions.US_EAST_1));
    //get the audio stream
    InputStream speechStream = helloWorld.synthesize(SAMPLE, OutputFormat.Mp3);
    //create an MP3 player
    AdvancedPlayer player = new AdvancedPlayer(speechStream,
        javazoom.jl.player.FactoryRegistry.systemRegistry().createAudioDevice());

    player.setPlaybackListener(new PlaybackListener() {
        @Override
        public void playbackStarted(PlaybackEvent evt) {
            System.out.println("Playback started");
            System.out.println(SAMPLE);
        }

        @Override
        public void playbackFinished(PlaybackEvent evt) {
            System.out.println("Playback finished");
        }
    });

    // play it!
    player.play();
    }
}

7. Return to the polly-java-demo directory to clean, compile, and execute the demo:
mvn clean compile exec:java
iOS Example

The following example uses the iOS SDK for Amazon Polly to read the specified text using a voice selected from a list of voices.

The code shown here covers the major tasks but does not handle errors. For the complete code, see AWS Mobile SDK for iOS Amazon Polly demo.

Initialize

```swift
// Region of Amazon Polly.
let AwsRegion = AWSRegionType.usEast1

// Cognito pool ID. Pool needs to be an unauthenticated pool with
// Amazon Polly permissions.
let CognitoIdentityPoolId = "YourCognitoIdentityPoolId"

// Initialize the Amazon Cognito credentials provider.
let credentialProvider = AWSCognitoCredentialsProvider(regionType: AwsRegion, identityPoolId: CognitoIdentityPoolId)

// Create an audio player
var audioPlayer = AVPlayer()
```

Get List of Available Voices

```swift
// Use the configuration as default
AWSServiceManager.default().defaultServiceConfiguration = configuration

// Get all the voices (no parameters specified in input) from Amazon Polly
// This creates an async task.
let task = AWSPolly.default().describeVoices(AWSPollyDescribeVoicesInput())

// When the request is done, asynchronously do the following block
// (we ignore all the errors, but in a real-world scenario they need
// to be handled)
task.continue(successBlock: { (awsTask: AWSTask) -> Any? in
  // awsTask.result is an instance of AWSPollyDescribeVoicesOutput in
  // case of the "describeVoices" method
  let voices = (awsTask.result! as AWSPollyDescribeVoicesOutput).voices
  return nil
})
```

Synthesize Speech

```swift
// First, Amazon Polly requires an input, which we need to prepare.
// Again, we ignore the errors, however this should be handled in
// real applications. Here we are using the URL Builder Request,
// since in order to make the synthesis quicker we will pass the
// presigned URL to the system audio player.
let input = AWSPollySynthesizeSpeechURLBuilderRequest()

// Text to synthesize
input.text = "Sample text"

// We expect the output in MP3 format
input.outputFormat = AWSPollyOutputFormat.mp3
```
// Choose the voice ID
input.voiceId = AWSPitchVoiceId.joanna

// Create an task to synthesize speech using the given synthesis input
let builder = AWSPolIySynthesizeSpeechURLBuilder.default().getPreSignedURL(input)

// Request the URL for synthesis result
builder.continueOnSuccessWith(block: { (awsTask: AWSTask<NSURL>) -> Any? in
  // The result of getPresignedURL task is NSURL.
  // Again, we ignore the errors in the example.
  let url = awsTask.result!

  // Try playing the data using the system AVAudioPlayer
  self.audioPlayer.replaceCurrentItem(with: AVPlayerItem(url: url as URL))
  self.audioPlayer.play()

  return nil
})
Android Example

The following example uses the Android SDK for Amazon Polly to read the specified text using a voice selected from a list of voices.

The code shown here covers the major tasks but does not handle errors. For the complete code, see the AWS Mobile SDK for Android Amazon Polly demo.

Initialize

```
// Cognito pool ID. Pool needs to be unauthenticated pool with
// Amazon Polly permissions.
String COGNITO_POOL_ID = "YourCognitoIdentityPoolId";

// Region of Amazon Polly.
Regions MY_REGION = Regions.US_EAST_1;

// Initialize the Amazon Cognito credentials provider.
CognitoCachingCredentialsProvider credentialsProvider = new
CognitoCachingCredentialsProvider(
    getApplicationContext(),
    COGNITO_POOL_ID,
    MY_REGION
);

// Create a client that supports generation of presigned URLs.
AmazonPollyPresigningClient client = new AmazonPollyPresigningClient(credentialsProvider);
```

Get List of Available Voices

```
// Create describe voices request.
DescribeVoicesRequest describeVoicesRequest = new DescribeVoicesRequest();

// Synchronously ask Amazon Polly to describe available TTS voices.
DescribeVoicesResult describeVoicesResult = client.describeVoices(describeVoicesRequest);
List<Voice> voices = describeVoicesResult.getVoices();
```

Get URL for Audio Stream

```
// Create speech synthesis request.
SynthesizeSpeechPresignRequest synthesizeSpeechPresignRequest = new
    SynthesizeSpeechPresignRequest()
    // Set the text to synthesize.
    .withText("Hello world!")
    // Select voice for synthesis.
    .withVoiceId(voices.get(0).getId()) // "Joanna"
    // Set format to MP3.
    .withOutputFormat(OutputFormat.Mp3);

// Get the presigned URL for synthesized speech audio stream.
URL presignedSynthesizeSpeechUrl =
    client.getPresignedSynthesizeSpeechUrl(synthesizeSpeechPresignRequest);
```

Play Synthesized Speech

```
// Use MediaPlayer: https://developer.android.com/guide/topics/media/mediaplayer.html
```
// Create a media player to play the synthesized audio stream.
MediaPlayer mediaPlayer = new MediaPlayer();
mediaPlayer.setAudioStreamType(AudioManager.STREAM_MUSIC);

try {
    // Set media player's data source to previously obtained URL.
    mediaPlayer.setDataSource(presignedSynthesizeSpeechUrl.toString());
} catch (IOException e) {
    Log.e(TAG, "Unable to set data source for the media player! " + e.getMessage());
}

// Prepare the MediaPlayer asynchronously (since the data source is a network stream).
mediaPlayer.prepareAsync();

// Set the callback to start the MediaPlayer when it's prepared.
mediaPlayer.setOnPreparedListener(new MediaPlayer.OnPreparedListener() {
    @Override
    public void onPrepared(MediaPlayer mp) {
        mp.start();
    }
});

// Set the callback to release the MediaPlayer after playback is completed.
mediaPlayer.setOnCompletionListener(new MediaPlayer.OnCompletionListener() {
    @Override
    public void onCompletion(MediaPlayer mp) {
        mp.release();
    }
});
Amazon Polly for Windows SAPI Compliant Applications

With the Amazon Polly for Windows plugin, Amazon Polly provides greatly expanded speech capability for Speech Application Programming Interface (SAPI)-compliant Windows applications. SAPI is a Microsoft Windows API that enables desktop applications to generate speech.

Windows provides one male and one female voice that can be used with for the SAPI API. With the Amazon Polly for Windows plug-in, the Windows application becomes an AWS client, enabling you to use all Amazon Polly voices, including both standard and neural voices, and the newscaster speaking style. The Amazon Polly for Windows plugin only supports 64-bit SAPI applications.

Amazon Polly for Windows supports all Amazon Polly, features such as Speech Synthesis Markup Language (SSML). When using SSML with the plugin, the standard functionality is available and the standard limits apply. For more information about using SSML with Amazon Polly, see Generating Speech from SSML Documents (p. 105). For more information about limits, see Quotas in Amazon Polly (p. 225).

Topics
- Installing and Configuring Amazon Polly for Windows (SAPI) (p. 212)
- Using Amazon Polly in Applications (p. 214)

Installing and Configuring Amazon Polly for Windows (SAPI)

To install and configure the Amazon Polly for Windows (SAPI) plugin, you need an AWS account. If you don't have an account, see Step 1.1: Sign up for AWS (p. 4).

Topics
- Create an IAM User for the AWS Client (p. 212)
- Install the AWS CLI for Windows (p. 213)
- Create a Profile for the AWS Client (p. 213)
- Install the Amazon Polly for Windows Plugin (p. 213)

Create an IAM User for the AWS Client

Before connecting the AWS client to your AWS account, you need to create an Identity and Access Management (IAM) user for the client, then attach a permissions policy to that user. An IAM user is a person or application in an AWS account that can make API calls to AWS products.

The AWS client uses the Amazon Polly service, so it needs the AmazonPollyReadOnlyAccess permissions policy.
To create an IAM user

1. Sign in to the AWS Management Console and open the IAM console.
2. Choose Users.
3. Choose Add User.
4. For User Name, enter polly-windows-user.
5. For Access Type, choose Programmatic access, then choose Next: Permissions.
6. Choose Attach existing policies directly.
7. In the search box, enter polly.
8. Choose AmazonPollyReadOnlyAccess.
9. Choose Next: Tags.
10. Choose Next: Review.
11. Choose Create User.
12. Record the access key ID and secret access key. You need them to configure the AWS client.

Important
This is the only time you can access these keys, so be sure to record them.

Install the AWS CLI for Windows

The AWS Command Line Interface (AWS CLI) is an open source tool that enables you to interact with AWS services using commands from the Windows command prompt. To install the AWS CLI, see Installing the AWS CLI on Windows in the AWS Command Line Interface User’s Guide.

Create a Profile for the AWS Client

The Amazon Polly for Windows plugin requires an AWS profile called polly-windows. This profile ensures that the Amazon Polly engine use the correct account.

To create the polly-windows profile for the AWS client

1. In Windows, open a command prompt.
2. At the command prompt, run the aws configure --profile polly-windows command.
3. When prompted for the AWS Access Key ID and AWS Secret Access Key, enter the values that you saved when you created the IAM user.
4. For Default region, enter the name of the desired AWS Region in lower-case letters.

Note
Not all Amazon Polly voices and features are available in all AWS Regions. For more information, see Feature and Region Compatibility (p. 93).
5. For Default output format, press Enter.
6. At the command prompt, verify the profile by running the aws --profile polly-windows polly describe-voices command. If you have successfully created the profile, the AWS CLI displays a list the available Amazon Polly names.

Install the Amazon Polly for Windows Plugin

Install the Amazon Polly for Windows plugin by downloading it from Amazon Simple Storage Service (Amazon S3) and running the installer. After you install the plugin, you can use Amazon Polly voices in all Windows application that use the SAPI API.
Using Amazon Polly in Applications

After you've installed the Amazon Polly for Windows plugin, the Amazon Polly voices can be accessed by any Windows application that implements Windows SAPI. To use a voice, choose it from the list of Amazon Polly voices in the application. For a complete list of available voices, see Available Voices (p. 12).

For an example of how to use Amazon Polly voices in a Windows application (Adobe Captivate), see Using Amazon Polly in Windows Applications on the AWS Machine Learning Blog.

To use Amazon Polly voices in a Windows application, you can use PollyPlayer. PollyPlayer is a simple Windows speech synthesis application that's installed when you install the Amazon Polly for Windows plugin.

To use PollyPlayer

1. On the Windows Start menu, under Amazon Polly for Windows, choose PollyPlayer.
2. In the PollyPlayer application, choose a voice, then enter the text that you want to hear.
### Pick a Voice:

3. To hear the text, choose **Say It**.
The AWS for WordPress Plugin

With the AWS for WordPress plugin, you can set up several AWS services, including Amazon Polly. With Amazon Polly, you can provide visitors to your WordPress website audio recordings of your content. Use the plugin to create audio files in any of the voices and languages supported by Amazon Polly. Your visitors can stream the audio at their convenience using inline audio players and mobile applications.

You can configure the plugin to do the following:

• Automatically create audio recordings for new content upon publication, or choose to create recordings for individual posts
• Create audio recordings of your archived content
• Use the Amazon Pollycast RSS feed to podcast audio content

Installation Prerequisites

To use the AWS for WordPress plugin, you need an AWS account, an AWS Identity and Access Management (IAM) user, and a WordPress website.

Topics

• Installation Prerequisites (p. 217)
• Installing and Configuring the Plugin (p. 219)
• Customizing Your WordPress Page (p. 220)
• Podcasting with Amazon Pollycast (p. 222)
• Storing Audio Files (p. 223)

Creating an AWS Account

If you have an AWS account already, you can skip this section. Otherwise, create one.

To create an AWS account

2. Follow the online instructions.

Part of the sign-up procedure involves receiving a phone call and entering a verification code on the phone keypad.
Creating an IAM User

To use the AWS for WordPress plugin, you must create an IAM user for the plugin. An IAM user is a person or application under an AWS account that has permission to make API calls to AWS services.

**Note**
If you don't use WordPress.com and instead have a self-hosted WordPress website on Amazon EC2, you can use an IAM role instead of an IAM user. For more information, see IAM Roles for Amazon EC2 in the *Amazon EC2 User Guide*.

The following procedure contains the steps to create an IAM policy, and then attach it to the IAM user. An IAM policy is a document that defines the permissions that apply to the user.

**To create an IAM user**

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose **Policies**. Then choose **Create policy**.
3. Choose **JSON**.
4. Delete everything in the policy text box, and then paste or enter the following JSON policy into the text box:

   ```json
   {  
     "Version": "2012-10-17",  
     "Statement": [  
       {  
         "Effect": "Allow",  
         "Action": "acm:DeleteCertificate",  
         "Resource": "*",  
         "Condition": {  
           "StringEquals": {  
             "aws:RequestedRegion": "us-east-1"  
           }  
         }  
       }  
     ]  
   }
   ```

5. Choose **Review policy**.
6. On the **Review policy** page, do the following:
   a. For **Name**, enter AWSForWordPressDeleteCert.
   b. Choose **Create policy**.
7. In the navigation pane, choose **Users**. Then choose **Add user**.
8. On the **Set user details** page, do the following:
   a. For **User name**, enter AWSForWordPressPlugin.
   b. For **Access type**, choose **Programmatic access**.
   c. Choose **Next: Permissions**.
9. On the **Set permissions** page, do the following:
   a. Choose **Attach existing policies directly**.
   b. In the search box, enter **WordPress**, and then select the check boxes next to AWSForWordPressPolicy and AWSForWordPressDeleteCert. Make sure to select the check boxes for both WordPress policies.
Creating a WordPress Website

If you have a WordPress website already, you can skip ahead to Installing and Configuring the Plugin (p. 219).

If you don't have a WordPress website, you can create one using WordPress.com. To use the AWS for WordPress plugin, you need a WordPress.com Business or eCommerce plan.

You can also install the WordPress software on your own web server, using Amazon Lightsail, Amazon EC2, or another web hosting platform. Hosting your own WordPress website involves more steps than using WordPress.com, and requires the ability to configure and manage a web server, a load balancer, DNS records, and web server certificates.

Regardless of how you set up your WordPress website, you need the following before you can use the AWS for WordPress plugin:

- Your website must have its own domain name. A domain name, also known as a web address or a URL (uniform resource locator), is the address that visitors use to go to your website. For example, Amazon's domain name is amazon.com. In this topic, we use example.com as a generic example domain name, but you need a custom domain name for your website.
- Your website must work using HTTPS. This is a security best practice, and the plugin assumes that your website works using HTTPS. To check, go to your website's address using HTTPS (for example, https://example.com) and make sure that your website displays correctly.

When your website has a domain name and works using HTTPS, proceed to the following section.

Installing and Configuring the Plugin

Before you install the plugin, make sure to complete the prerequisites (p. 217).

To install the plugin

1. Log in to the admin dashboard for your WordPress website, also known as WP Admin.
2. Choose Plugins.
3. Choose Add New.
4. In the search box, enter AWS for WordPress.
5. Find the **AWS for WordPress** plugin. Choose **Install Now**, and then choose **Activate**.

## Configuring the Plugin

Once you've installed the **AWS for WordPress** plugin, configure it to enable podcasting, alternative storage locations, and other options.

**Note**

In the following procedure, command and field names might differ slightly from the names used in WordPress.

**To configure the plugin**

1. On the **WordPress Admin** page, choose **Settings**.
2. Configure the plugin using the following options:
   
   - **AWS access key and AWS secret key**—These AWS credentials allow the plugin to use Amazon Polly and Amazon Simple Storage Service (Amazon S3). Type the AWS access and secret keys that you created earlier. If you are hosting your WordPress site on Amazon EC2, you can use an IAM role instead of credentials. In that case, leave these two fields blank.
   
   - **Sample rate**—The sample rate for the audio files that will be generated, in Hz. Higher sampling rates produce higher-quality audio.
   
   - **Voice name**—The Amazon Polly voice to use in the audio file.
     
     **Note**
     
     Remember that the choice of voice is determined by the language you're using. For a complete set of available voices, see [Available Voices](#) (p. 12)

   - **Player position**—Where to position the audio player on the website. You can put it before or after the post, or not use it at all. If you want to make your files available as podcasts by using Amazon Pollycast, don’t display the audio player.
   
   - **New post default**—Specifies whether Amazon Polly should automatically create an audio file for all new posts. Choose this option if you want Amazon Polly to create an audio file for each new post.
   
   - **Autoplay**—Specifies whether the audio player automatically starts playing the audio for a post when a visitor visits it.
   
   - **Store audio in Amazon S3**—If you want to store audio files in an S3 bucket instead of on your web server, choose this option. Amazon Polly creates the bucket for you. For more information and pricing, see [Amazon S3](#).
   
   - **Amazon CloudFront (CDN) domain name**—If you want to broadcast your audio files with Amazon CloudFront, provide the name of your CloudFront domain. The plugin uses the domain to stream audio. If you don’t already have a domain, create one in Amazon CloudFront.
   
   - **iTunes category**—The category for your podcast. Choosing a category makes it easier for podcast users to find your podcast in the podcast catalog.
   
   - **iTunes explicit**—Specifies whether to enable Amazon Pollycast podcasting with the iTunes explicit setting.
   
   - **Bulk update all posts**—If you want to convert all your previous WordPress posts to use the **AWS for WordPress** plugin, choose this option.

3. Choose **Save Changes**.

## Customizing Your WordPress Page

You can use several options for customizing your WordPress content to make it work more effectively with the Amazon Polly WordPress plugin:
• Adjust the plugin settings to fine tune the audio files (p. 221)
• Use SSML in your content to modify how it will be spoken (p. 221)
• Use Audio Only and Word Only tags in your content (p. 221)

Adjust the plugin settings to fine tune the audio files

The AWS for WordPress plugin settings have three options that can help you customize the sound of your WordPress text for the audio file:

• Voice name: The voice name and language selected enables you to select the gender of the Amazon Polly voice. Specific voices are available for each language, and you have several options within each gender in many languages. For more information, see Voices in Amazon Polly (p. 12).

• Automated breaths: If enabled, Amazon Polly will automatically insert breathing sounds into your content at appropriate locations. Setting this option will only enable you to use automated breaths and not manually set them. For more information, see automated breathing.

• Audio speed: This option enables you to alter the speed of delivery of the audio version of your content, ranging from 20% to 200% of the default speed of the voice.

Use SSML in your content to modify how it will be spoken

Amazon Polly itself supports multiple SSML tags that enable you to control many aspects of how Amazon Polly generates speech from the text you provide. (For more information about SSML and Amazon Polly, see Supported SSML Tags (p. 112). A few of these tags are echoed in the plug-in settings when configuring your plug-in. Currently, however, only the <break> tag is available for direct use in the AWS for WordPress plugin.

The <break> tag enables you to add a pause into the spoken version of your text. You can adjust the length of this pause to meet your particular needs. The default length of the pause is equivalent to the pause following a comma. For more information on the <break> tag, see Supported SSML Tags (p. 112).

To use SSML tags to enhance your WordPress text, the Enable SSML support option must be selected in Amazon Polly Settings on the WordPress Admin page. The Store audio in Amazon S3 option must also be selected because audio files with SSML tags need to be stored in an S3 bucket.

Use Audio Only and Word Only tags in your content

Sometimes you want to add something to your audio podcast but don’t want it to display in the browser. Or you want something to show in the browser but not include it in the audio file. This can be done with the Audio Only and Word Only tags, which you place in the WordPress content to control which part of the text will be displayed or spoken.

To convert text to audio but not display it in the browser

1. Isolate the selected text on your WordPress page by placing an empty line above and below it.
2. On the line above the selected text, insert the following tag:
   ```html
   -AMAZONPOLLY-ONLYAUDIO-START-
   ```
3. On the line following the text, insert the following tag:
   ```html
   -AMAZONPOLLY-ONLYAUDIO-END-
   ```
Adding Translated Text to Your Post

The AWS for WordPress plugin uses Amazon Translate to create translated versions of your post in one or more other languages. In addition to English, four languages are available for this service: Spanish, French, German, and Portuguese. The languages to be used and the voices for those languages can be configured on the Settings page under Amazon Translate configuration.

To translate your WordPress post into other languages

2. On the same page, ensure that the Enable Amazon Polly option is chosen.
3. To see the approximate cost of creating audio files in the original languages plus any additionally selected languages, choose How much will this cost to convert? Choose OK to return to the Add New Post page.
4. Choose Translate

To set the languages into which you translate your post

1. On the Settings page, under Amazon Translate configuration, choose the language of your post from the Source language drop-down list.
2. Under Target languages, choose the languages into which you want to translate your post.
3. From the Voice dropdown list, choose the voice you want to use for each language selected.
4. Enter a label for the language choice.
5. Choose Save Changes.

Podcasting with Amazon Pollycast

With Amazon Pollycast feeds, your visitors can listen to your audio content using standard podcast applications. RSS 2.0-compliant Pollycast feeds provide the XML data needed to aggregate podcasts by popular mobile podcast applications, such as iTunes, and podcast directories.
When you install the **AWS for WordPress** plugin, choose the **iTunes explicit** option to automatically add Amazon Pollocast endpoints to all WordPress archive URLs. This lets you syndicate both site-wide or targeted podcasts. If you didn't choose the **iTunes explicit** option when you installed the plugin, follow these steps:

1. On the **WordPress Admin** page, choose **Settings**.
2. On the **Settings** page, choose **iTunes explicit**.

You can add Amazon Pollocast endpoints manually by adding `/amazon-pollcast/` to the URL for a page in a podcasting application. For example:

```
example.com/amazon-pollcast/
example.com/category/news/amazon-pollcast/
example.com/author/john/amazon-pollcast/
```

## Positioning the Player

When you install the **AWS for WordPress** plugin, it displays an audio player at the top of your WordPress website by default unless you choose to display it under your site’s text or to not display it.

You can reposition the player, remove it, or add it (if you chose not to display it) at any time.

**To reposition the player, remove it, or add it to your WordPress website**

1. On the **WordPress Admin** page, choose **Settings**.
2. On the **Amazon Polly Settings** page, for **Player position**, choose the appropriate option.

For more information about setting configuration options, see [Configuring the Plugin](#) (p. 220).

## Storing Audio Files

When you publish content on your site, it’s sent to Amazon Polly for synthesis. By default, Amazon Polly stores new audio files on your web server. You can also store the files on Amazon Simple Storage Service (Amazon S3) or on Amazon CloudFront, which is a global content delivery network (CDN).

Users have the same listening experience regardless of where you store your audio files. Only the broadcast location changes:

1. For audio files stored on the WordPress server, files are broadcast directly from the server.
2. For files stored in an S3 bucket, files are broadcast from the bucket.
3. If you use CloudFront, the files are stored on Amazon S3 and are broadcast with CloudFront.
You can choose where to store your files when you install the Amazon Polly plugin.
Quotas in Amazon Polly

The following are limits to be aware of when using Amazon Polly.

Supported Regions

For a list of AWS Regions where Amazon Polly is available, see Amazon Polly Endpoints and Quotas in the Amazon Web Services General Reference.

Neural voices aren't available in all AWS Regions. For the Regions that support neural voices, see the section called “Feature and Region Compatibility” (p. 93) for neural TTS.

Throttling

- Throttle rate per account: 100 transactions (requests or operations) per second (tps) with a burst limit of 120 tps.
- Concurrent connections per account: 90
- Throttle rate per operation:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicon</td>
<td></td>
</tr>
<tr>
<td>DeleteLexicon</td>
<td>Any 2 transactions per second (tps) from these operations combined.</td>
</tr>
<tr>
<td>PutLexicon</td>
<td>Maximum allowed burst of 4 tps.</td>
</tr>
<tr>
<td>GetLexicon</td>
<td></td>
</tr>
<tr>
<td>ListLexicons</td>
<td></td>
</tr>
<tr>
<td>Speech</td>
<td></td>
</tr>
<tr>
<td>DescribeVoices</td>
<td>80 tps with a burst limit of 100 tps</td>
</tr>
<tr>
<td>SynthesizeSpeech</td>
<td>Standard voice: 80 tps with a burst limit of 100 tps</td>
</tr>
<tr>
<td></td>
<td>Neural voice: 8 tps with a burst limit of 10 tps</td>
</tr>
<tr>
<td>StartSpeechSynthesisTask</td>
<td>Standard voice: 10 tps with a burst limit of 12 tps</td>
</tr>
<tr>
<td></td>
<td>Neural voice: 1 tps</td>
</tr>
<tr>
<td>GetSynthesizeSpeechTask</td>
<td>Maximum allowed 10 tps combined</td>
</tr>
<tr>
<td>and ListSynthesizeSpeechTask</td>
<td></td>
</tr>
</tbody>
</table>

Pronunciation Lexicons

- You can store up to 100 lexicons per account.
• Lexicon names can be an alphanumeric string up to 20 characters long.
• Each lexicon can be up to 4,000 characters in size. (Note that the size of the lexicon affects the latency of the SynthesizeSpeech operation.)
• You can specify up to 100 characters for each <phoneme> or <alias> replacement in a lexicon.

For information about using lexicons, see Managing Lexicons (p. 133).

**SynthesizeSpeech API Operation**

Note the following limits related to using the SynthesizeSpeech API operation:

• The size of the input text can be up to 3000 billed characters (6000 total characters). SSML tags are not counted as billed characters.
• You can specify up to five lexicons to apply to the input text.
• The output audio stream (synthesis) is limited to 10 minutes. After this is reached, any remaining speech is cut off.

For more information, see SynthesizeSpeech (p. 270).

**Note**
Some limitations of the SynthesizeSpeech API operation can be bypassed using the StartSynthesizeSpeechTask API operation. For more information, see Creating Long Audio Files (p. 144).

**SpeechSynthesisTask API Operations**

Note the following limit relating to using the StartSpeechSynthesisTask, GetSpeechSynthesisTask, and ListSpeechSynthesisTasks API operations:

• The size of the input text can be up to 100,000 billed characters (200,000 total characters). SSML tags are not counted as billed characters.
• You can specify up to five lexicons to apply to the input text.

**Speech Synthesis Markup Language (SSML)**

Note the following limits related to using SSML:

• The <audio>, <lexicon>, <lookup>, and <voice> tags are not supported.
• <break> elements can specify a maximum duration of 10 seconds each.
• The <prosody> tag doesn't support values for the rate attribute lower than -80%.

For more information, see Generating Speech from SSML Documents (p. 105).
Security in Amazon Polly

Cloud security at AWS is the highest priority. As an AWS customer, you benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

Security is a shared responsibility between AWS and you. The shared responsibility model describes this as security of the cloud and security in the cloud:

- **Security of the cloud** – AWS is responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. AWS also provides you with services that you can use securely. Third-party auditors regularly test and verify the effectiveness of our security as part of the AWS Compliance Programs. To learn about the compliance programs that apply to Amazon Polly, see AWS Services in Scope by Compliance Program.
- **Security in the cloud** – Your responsibility is determined by the AWS service that you use. You are also responsible for other factors including the sensitivity of your data, your company’s requirements, and applicable laws and regulations.

This documentation helps you understand how to apply the shared responsibility model when using Amazon Polly. The following topics show you how to configure Amazon Polly to meet your security and compliance objectives. You also learn how to use other AWS services that help you to monitor and secure your Amazon Polly resources.

Topics
- Data Protection in Amazon Polly (p. 227)
- Identity and Access Management in Amazon Polly (p. 228)
- Logging and Monitoring in Amazon Polly (p. 239)
- Compliance Validation for Amazon Polly (p. 239)
- Resilience in Amazon Polly (p. 240)
- Infrastructure Security in Amazon Polly (p. 240)
- Security Best Practices for Amazon Polly (p. 240)

Data Protection in Amazon Polly

Amazon Polly conforms to the AWS shared responsibility model, which includes regulations and guidelines for data protection. AWS is responsible for protecting the global infrastructure that runs all the AWS services. AWS maintains control over data hosted on this infrastructure, including the security configuration controls for handling customer content and personal data. AWS customers and APN partners, acting either as data controllers or data processors, are responsible for any personal data that they put in the AWS Cloud.

For data protection purposes, we recommend that you protect AWS account credentials and set up individual user accounts with AWS Identity and Access Management (IAM), so that each user is given only the permissions necessary to fulfill their job duties. We also recommend that you secure your data in the following ways:

- Use multi-factor authentication (MFA) with each account.
- Use SSL/TLS to communicate with AWS resources.
- Set up API and user activity logging with AWS CloudTrail.
• Use AWS encryption solutions, along with all default security controls within AWS services.

We strongly recommend that you never put sensitive identifying information, such as your customers' account numbers, into free-form fields such as a Name field. This includes when you work with Amazon Polly or other AWS services using the console, API, AWS CLI, or AWS SDKs. Any data that you enter into Amazon Polly or other services might get picked up for inclusion in diagnostic logs. When you provide a URL to an external server, don't include credentials information in the URL to validate your request to that server.

For more information about data protection, see the AWS Shared Responsibility Model and GDPR blog post on the AWS Security Blog.

Encryption at Rest

Output of your Amazon Polly voice synthesis can be saved on your own system. You can also call Amazon Polly, and then encrypt the file with any encryption key of your choice and store it in Amazon Simple Storage Service (Amazon S3) or another secure storage. The Amazon Polly the section called “SynthesizeSpeech” (p. 270) operation is stateless and is not associated with a customer identity. You can't retrieve it from Amazon Polly later.

Encryption in Transit

All text submissions are protected by Secure Sockets Layer (SSL) while in transit. Amazon Polly does not retain the content of text submissions.

Internetwork Traffic Privacy

Access to Amazon Polly is via the AWS console, CLI, or SDKs. Communications utilize Transport Layer Security (TLS) session encryption for confidentiality and digital signatures for authentication and integrity.

Identity and Access Management in Amazon Polly

AWS Identity and Access Management (IAM) is an AWS service that helps an administrator securely control access to AWS resources. IAM administrators control who can be authenticated (signed in) and authorized (have permissions) to use Amazon Polly resources. IAM is an AWS service that you can use with no additional charge.

Topics
• Audience (p. 228)
• Authenticating With Identities (p. 229)
• Managing Access Using Policies (p. 231)
• How Amazon Polly Works with IAM (p. 232)
• Amazon Polly Identity-Based Policy Examples (p. 234)
• Amazon Polly API Permissions: Actions, Permissions, and Resources Reference (p. 238)

Audience

How you use AWS Identity and Access Management (IAM) differs, depending on the work you do in Amazon Polly.
**Service user** – If you use the Amazon Polly service to do your job, your administrator provides you with the credentials and permissions that you need. As you use more Amazon Polly features to do your work, you might need additional permissions. Understanding how access is managed helps you request the right permissions from your administrator.

**Service administrator** – If you're in charge of Amazon Polly resources at your company, you probably have full access to Amazon Polly. It's your job to determine which Amazon Polly features and resources your employees should access. Then submit requests to your IAM administrator to change the permissions of your service users. Review the information on this page to understand the basic concepts of IAM. To learn more about how your company can use IAM with Amazon Polly, see How Amazon Polly Works with IAM (p. 232).

**IAM administrator** – If you're an IAM administrator, you might want to learn details about writing policies to manage access to Amazon Polly. To view example Amazon Polly identity-based policies that you can use in IAM, see Amazon Polly Identity-Based Policy Examples (p. 234).

---

### Authenticating With Identities

Authentication is how you sign in to AWS using your identity credentials. For more information about signing in using the AWS Management Console, see Signing in to the AWS Management Console as an IAM user or root user in the IAM User Guide.

You must be authenticated (signed in to AWS) as the AWS account root user, an IAM user, or by assuming an IAM role. You can also use your company's single sign-on authentication or even sign in using Google or Facebook. In these cases, your administrator previously set up identity federation using IAM roles. When you access AWS using credentials from another company, you are assuming a role indirectly.

To sign in directly to the AWS Management Console, use your password with your root user email address or your IAM user name. You can access AWS programmatically using your root user or IAM users access keys. AWS provides SDK and command line tools to cryptographically sign your request using your credentials. If you don't use AWS tools, you must sign the request yourself. Do this using Signature Version 4, a protocol for authenticating inbound API requests. For more information about authenticating requests, see Signature Version 4 signing process in the AWS General Reference.

Regardless of the authentication method that you use, you might also be required to provide additional security information. For example, AWS recommends that you use multi-factor authentication (MFA) to increase the security of your account. To learn more, see Using multi-factor authentication (MFA) in AWS in the IAM User Guide.

### AWS account root user

When you first create an AWS account, you begin with a single sign-in identity that has complete access to all AWS services and resources in the account. This identity is called the AWS account root user and is accessed by signing in with the email address and password that you used to create the account. We strongly recommend that you do not use the root user for your everyday tasks, even the administrative ones. Instead, adhere to the best practice of using the root user only to create your first IAM user. Then securely lock away the root user credentials and use them to perform only a few account and service management tasks.

### IAM Users and Groups

An IAM user is an identity within your AWS account that has specific permissions for a single person or application. An IAM user can have long-term credentials such as a user name and password or a set of access keys. To learn how to generate access keys, see Managing access keys for IAM users in the IAM User Guide. When you generate access keys for an IAM user, make sure you view and securely save the key pair. You cannot recover the secret access key in the future. Instead, you must generate a new access key pair.
An IAM group is an identity that specifies a collection of IAM users. You can't sign in as a group. You can use groups to specify permissions for multiple users at a time. Groups make permissions easier to manage for large sets of users. For example, you could have a group named IAMAdmins and give that group permissions to administer IAM resources.

Users are different from roles. A user is uniquely associated with one person or application, but a role is intended to be assumable by anyone who needs it. Users have permanent long-term credentials, but roles provide temporary credentials. To learn more, see When to create an IAM user (instead of a role) in the IAM User Guide.

IAM Roles

An IAM role is an identity within your AWS account that has specific permissions. It is similar to an IAM user, but is not associated with a specific person. You can temporarily assume an IAM role in the AWS Management Console by switching roles. You can assume a role by calling an AWS CLI or AWS API operation or by using a custom URL. For more information about methods for using roles, see Using IAM roles in the IAM User Guide.

IAM roles with temporary credentials are useful in the following situations:

- **Temporary IAM user permissions** – An IAM user can assume an IAM role to temporarily take on different permissions for a specific task.
- **Federated user access** – Instead of creating an IAM user, you can use existing identities from AWS Directory Service, your enterprise user directory, or a web identity provider. These are known as federated users. AWS assigns a role to a federated user when access is requested through an identity provider. For more information about federated users, see Federated users and roles in the IAM User Guide.
- **Cross-account access** – You can use an IAM role to allow someone (a trusted principal) in a different account to access resources in your account. Roles are the primary way to grant cross-account access. However, with some AWS services, you can attach a policy directly to a resource (instead of using a role as a proxy). To learn the difference between roles and resource-based policies for cross-account access, see How IAM roles differ from resource-based policies in the IAM User Guide.
- **Cross-service access** – Some AWS services use features in other AWS services. For example, when you make a call in a service, it's common for that service to run applications in Amazon EC2 or store objects in Amazon S3. A service might do this using the calling principal's permissions, using a service role, or using a service-linked role.
- **Principal permissions** – When you use an IAM user or role to perform actions in AWS, you are considered a principal. Policies grant permissions to a principal. When you use some services, you might perform an action that then triggers another action in a different service. In this case, you must have permissions to perform both actions. To see whether an action requires additional dependent actions in a policy, see Actions, Resources, and Condition Keys for Amazon Polly in the Service Authorization Reference.
- **Service role** – A service role is an IAM role that a service assumes to perform actions on your behalf. An IAM administrator can create, modify, and delete a service role from within IAM. For more information, see Creating a role to delegate permissions to an AWS service in the IAM User Guide.
- **Service-linked role** – A service-linked role is a type of service role that is linked to an AWS service. The service can assume the role to perform an action on your behalf. Service-linked roles appear in your IAM account and are owned by the service. An IAM administrator can view, but not edit the permissions for service-linked roles.
- **Applications running on Amazon EC2** – You can use an IAM role to manage temporary credentials for applications that are running on an EC2 instance and making AWS CLI or AWS API requests. This is preferable to storing access keys within the EC2 instance. To assign an AWS role to an EC2 instance and make it available to all of its applications, you create an instance profile that is attached to the instance. An instance profile contains the role and enables programs that are running on the EC2 instance to get temporary credentials. For more information, see Using an IAM role to grant permissions to applications running on Amazon EC2 instances in the IAM User Guide.
To learn whether to use IAM roles or IAM users, see When to create an IAM role (instead of a user) in the IAM User Guide.

Managing Access Using Policies

You control access in AWS by creating policies and attaching them to IAM identities or AWS resources. A policy is an object in AWS that, when associated with an identity or resource, defines its permissions. You can sign in as the root user or an IAM user, or you can assume an IAM role. When you then make a request, AWS evaluates the related identity-based or resource-based policies. Permissions in the policies determine whether the request is allowed or denied. Most policies are stored in AWS as JSON documents. For more information about the structure and contents of JSON policy documents, see Overview of JSON policies in the IAM User Guide.

Administrators can use AWS JSON policies to specify who has access to what. That is, which principal can perform actions on what resources, and under what conditions. Every IAM entity (user or role) starts with no permissions. In other words, by default, users can do nothing, not even change their own password. To give a user permission to do something, an administrator must attach a permissions policy to a user. Or the administrator can add the user to a group that has the intended permissions. When an administrator gives permissions to a group, all users in that group are granted those permissions.

IAM policies define permissions for an action regardless of the method that you use to perform the operation. For example, suppose that you have a policy that allows the iam:GetRole action. A user with that policy can get role information from the AWS Management Console, the AWS CLI, or the AWS API.

Identity-Based Policies

Identity-based policies are JSON permissions policy documents that you can attach to an identity, such as an IAM user, group of users, or role. These policies control what actions users and roles can perform, on which resources, and under what conditions. To learn how to create an identity-based policy, see Creating IAM policies in the IAM User Guide.

Identity-based policies can be further categorized as inline policies or managed policies. Inline policies are embedded directly into a single user, group, or role. Managed policies are standalone policies that you can attach to multiple users, groups, and roles in your AWS account. Managed policies include AWS managed policies and customer managed policies. To learn how to choose between a managed policy or an inline policy, see Choosing between managed policies and inline policies in the IAM User Guide.

Resource-Based Policies

Other services, such as Amazon S3, also support resource-based permissions policies. For example, you can attach a policy to an S3 bucket to manage access permissions to that bucket. Amazon Polly doesn't support resource-based policies.

Other Policy Types

AWS supports additional, less-common policy types. These policy types can set the maximum permissions granted to you by the more common policy types.

- Permissions boundaries – A permissions boundary is an advanced feature in which you set the maximum permissions that an identity-based policy can grant to an IAM entity (IAM user or role). You can set a permissions boundary for an entity. The resulting permissions are the intersection of entity’s identity-based policies and its permissions boundaries. Resource-based policies that specify the user or role in the Principal field are not limited by the permissions boundary. An explicit deny in any of these policies overrides the allow. For more information about permissions boundaries, see Permissions boundaries for IAM entities in the IAM User Guide.
• **Service control policies (SCPs)** – SCPs are JSON policies that specify the maximum permissions for an organization or organizational unit (OU) in AWS Organizations. AWS Organizations is a service for grouping and centrally managing multiple AWS accounts that your business owns. If you enable all features in an organization, then you can apply service control policies (SCPs) to any or all of your accounts. The SCP limits permissions for entities in member accounts, including each AWS account root user. For more information about Organizations and SCPs, see How SCPs work in the AWS Organizations User Guide.

• **Session policies** – Session policies are advanced policies that you pass as a parameter when you programmatically create a temporary session for a role or federated user. The resulting session's permissions are the intersection of the user or role's identity-based policies and the session policies. Permissions can also come from a resource-based policy. An explicit deny in any of these policies overrides the allow. For more information, see Session policies in the IAM User Guide.

### Multiple Policy Types

When multiple types of policies apply to a request, the resulting permissions are more complicated to understand. To learn how AWS determines whether to allow a request when multiple policy types are involved, see Policy evaluation logic in the IAM User Guide.

### How Amazon Polly Works with IAM

Before you use IAM to manage access to Amazon Polly, you should understand what IAM features are available to use with Amazon Polly. To get a high-level view of how Amazon Polly and other AWS services work with IAM, see AWS Services That Work with IAM in the IAM User Guide.

### Topics

- Amazon Polly Identity-Based Policies (p. 232)
- Amazon Polly IAM Roles (p. 233)

### Amazon Polly Identity-Based Policies

With IAM identity-based policies, you can specify allowed or denied actions and resources as well as the conditions under which actions are allowed or denied. Amazon Polly supports specific actions, resources, and condition keys. To learn about all of the elements that you use in a JSON policy, see IAM JSON Policy Elements Reference in the IAM User Guide.

#### Actions

Administrators can use AWS JSON policies to specify who has access to what. That is, which principal can perform actions on what resources, and under what conditions.

The Action element of a JSON policy describes the actions that you can use to allow or deny access in a policy. Policy actions usually have the same name as the associated AWS API operation. There are some exceptions, such as permission-only actions that don't have a matching API operation. There are also some operations that require multiple actions in a policy. These additional actions are called dependent actions.

Include actions in a policy to grant permissions to perform the associated operation.

Policy actions in Amazon Polly use the following prefix before the action: `polly:`. For example, to grant someone permission to delete a Amazon Polly lexicon with the Amazon Polly DeleteLexicon API operation, you include the `polly:DeleteLexicon` action in their policy. Policy statements must include either an Action or NotAction element. Amazon Polly defines its own set of actions that describe tasks that you can perform with this service.
To specify multiple actions in a single statement, separate them with commas as follows:

```
"Action": [
  "polly:action1",
  "polly:action2"
]
```

You can specify multiple actions using wildcards (*). For example, to specify all actions that begin with the word `get`, include the following action:

```
"Action": "polly:Get*"
```

To see a list of Amazon Polly actions, see Actions Defined by Amazon Polly in the IAM User Guide.

**Resources**

Amazon Polly does not support specifying resource ARNs in a policy.

**Condition Keys**

Amazon Polly does not provide any service-specific condition keys, but it does support using some global condition keys. To see all AWS global condition keys, see AWS Global Condition Context Keys in the IAM User Guide.

**Examples**

To view examples of Amazon Polly identity-based policies, see Amazon Polly Identity-Based Policy Examples (p. 234).

**Amazon Polly IAM Roles**

You can attach an identity-based permissions policy to an IAM role to grant cross-account permissions. For example, the administrator in account A can create a role to grant cross-account permissions to another AWS account (for example, account B) or an AWS service as follows:

1. Account A administrator creates an IAM role and attaches a permissions policy to the role that grants permissions on resources in account A.
2. Account A administrator attaches a trust policy to the role identifying account B as the principal who can assume the role.
3. Account B administrator can then delegate permissions to assume the role to any users in account B. Doing this allows users in account B to create or access resources in account A. The principal in the trust policy can also be an AWS service principal if you want to grant an AWS service permissions to assume the role.

For more information about using IAM to delegate permissions, see Access Management in the IAM User Guide.

The following is an example policy that grants permissions to put and get lexicons as well as to list those lexicons currently available.

Amazon Polly supports Identity-based policies for actions at the resource-level. In some cases, the resource can be limited by an ARN. This is true for the `SynthesizeSpeech`, `StartSpeechSynthesisTask`, `PutLexicon`, `GetLexicon`, and `DeleteLexicon` operations. In these cases, the `Resource` value is indicated by the ARN. For example: `arn:aws:polly:us-east-2:account-id:lexicon/*` as the `Resource` value specifies permissions on all owned lexicons within the `us-east-2` Region.

However, not all operations use ARNs. This is the case with the DescribeVoices, ListLexicons, GetSpeechSynthesisTasks, and ListSpeechSynthesisTasks operations.

For more information about users, groups, roles, and permissions, see Identities (Users, Groups, and Roles) in the IAM User Guide.

Amazon Polly Identity-Based Policy Examples

By default, IAM users and roles don't have permission to create or modify Amazon Polly resources. They also can't perform tasks using the AWS Management Console, AWS CLI, or AWS API. An IAM administrator must create IAM policies that grant users and roles permission to perform specific API operations on the specified resources they need. The administrator must then attach those policies to the IAM users or groups that require those permissions.

To learn how to create an IAM identity-based policy using these example JSON policy documents, see Creating Policies on the JSON Tab in the IAM User Guide.

Topics

• Policy Best Practices (p. 234)
• Using the Amazon Polly Console (p. 235)
• Allow Users to View Their Own Permissions (p. 235)
• AWS Managed (Predefined) Policies for Amazon Polly (p. 236)
• Customer Managed Policy Examples (p. 236)

Policy Best Practices

Identity-based policies are very powerful. They determine whether someone can create, access, or delete Amazon Polly resources in your account. These actions can incur costs for your AWS account. When you create or edit identity-based policies, follow these guidelines and recommendations:

• Get started using AWS managed policies – To start using Amazon Polly quickly, use AWS managed policies to give your employees the permissions they need. These policies are already available in your account and are maintained and updated by AWS. For more information, see Get started using permissions with AWS managed policies in the IAM User Guide.

• Grant least privilege – When you create custom policies, grant only the permissions required to perform a task. Start with a minimum set of permissions and grant additional permissions as necessary. Doing so is more secure than starting with permissions that are too lenient and then trying to tighten them later. For more information, see Grant least privilege in the IAM User Guide.

• Enable MFA for sensitive operations – For extra security, require IAM users to use multi-factor authentication (MFA) to access sensitive resources or API operations. For more information, see Using multi-factor authentication (MFA) in AWS in the IAM User Guide.
• **Use policy conditions for extra security** – To the extent that it's practical, define the conditions under which your identity-based policies allow access to a resource. For example, you can write conditions to specify a range of allowable IP addresses that a request must come from. You can also write conditions to allow requests only within a specified date or time range, or to require the use of SSL or MFA. For more information, see IAM JSON policy elements: Condition in the IAM User Guide.

**Using the Amazon Polly Console**

For a user to work with the Amazon Polly console, that user must have a minimum set of permissions to describe the Amazon Polly resources in their AWS account.

If you create an IAM policy that is more restrictive than the minimum required permissions, the console doesn't function as intended for users with that IAM policy.

You don't need to allow minimum console permissions for users that are making calls only to the AWS CLI or the Amazon Polly API.

To use the Amazon Polly console, grant permissions to all the Amazon Polly APIs. There are no additional permissions needed. To get full console functionality you can use following policy:

```json
```

**Allow Users to View Their Own Permissions**

This example shows how you might create a policy that allows IAM users to view the inline and managed policies that are attached to their user identity. This policy includes permissions to complete this action on the console or programatically using the AWS CLI or AWS API.

```json
```
AWS Managed (Predefined) Policies for Amazon Polly

AWS addresses many common use cases by providing standalone IAM policies that are created and administered by AWS. These AWS managed policies grant necessary permissions for common use cases so that you can avoid having to investigate what permissions are needed. For more information, see AWS Managed Policies in the IAM User Guide.

The following AWS managed policies, which you can attach to users in your account, are specific to Amazon Polly:

- **AmazonPollyReadOnlyAccess** – Grants read-only access to resources, allows listing lexicons, fetching lexicons, listing available voices and synthesizing speech (including, applying lexicons to the synthesized speech).
- **AmazonPollyFullAccess** – Grants full access to resources and all the supported operations.

**Note**
You can review these permissions policies by signing in to the IAM console and searching for specific policies there.

You can also create your own custom IAM policies to allow permissions for Amazon Polly actions and resources. You can attach these custom policies to the IAM users or groups that require those permissions.

Customer Managed Policy Examples

In this section, you can find example user policies that grant permissions for various Amazon Polly actions. These policies work when you are using AWS SDKs or the AWS CLI. When you are using the console, grant permissions to all the Amazon Polly APIs.

**Note**
All examples use the us-east-2 Region and contain fictitious account IDs.

**Examples**

- Example 1: Allow All Amazon Polly Actions (p. 236)
- Example 2: Allow All Amazon Polly Actions Except DeleteLexicon (p. 237)
- Example 3: Allow DeleteLexicon (p. 237)
- Example 4: Allow Delete Lexicon in a Specified Region (p. 238)
- Example 5: Allow DeleteLexicon for Specified Lexicon (p. 238)

**Example 1: Allow All Amazon Polly Actions**

After you sign up (see Step 1.1: Sign up for AWS (p. 4)) create an administrator user to manage your account, including creating users and managing their permissions.
You might create a user who has permissions for all Amazon Polly actions. Think of this user as a service-specific administrator for working with Amazon Polly. You can attach the following permissions policy to this user.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "AllowAllPollyActions",
        "Effect": "Allow",
        "Action": [
            "polly:*"],
        "Resource": "*"
    }]
}
```

**Example 2: Allow All Amazon Polly Actions Except DeleteLexicon**

The following permissions policy grants the user permissions to perform all actions except `DeleteLexicon`, with the permissions for delete explicitly denied in all Regions.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "AllowAllActions-DenyDelete",
        "Effect": "Allow",
        "Action": [
            "polly:DescribeVoices",
            "polly:GetLexicon",
            "polly:PutLexicon",
            "polly:SynthesizeSpeech",
            "polly:ListLexicons"],
        "Resource": "*"
    }
    {
        "Sid": "DenyDeleteLexicon",
        "Effect": "Deny",
        "Action": [
            "polly:DeleteLexicon"],
        "Resource": "*"
    }]
}
```

**Example 3: Allow DeleteLexicon**

The following permissions policy grants the user permissions to delete any lexicon that you own regardless of the project or Region in which it is located.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "AllowDeleteLexicon",
        "Effect": "Allow",
        "Action": [
            "polly:DeleteLexicon"],
        "Resource": "*"
    }]
}
```
Example 4: Allow Delete Lexicon in a Specified Region

The following permissions policy grants the user permissions to delete any lexicon in any project that you own that is located in a single Region (in this case, us-east-2).

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "AllowDeleteSpecifiedRegion",
        "Effect": "Allow",
        "Action": [
            "polly:DeleteLexicon"
        ],
    }
}
```

Example 5: Allow DeleteLexicon for Specified Lexicon

The following permissions policy grants the user permissions to delete a specific lexicon that you own (in this case, myLexicon) in a specific Region (in this case, us-east-2).

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "AllowDeleteForSpecifiedLexicon",
        "Effect": "Allow",
        "Action": [
            "polly:DeleteLexicon"
        ],
    }
}
```

Amazon Polly API Permissions: Actions, Permissions, and Resources Reference

When you are setting up a permissions policy that you can attach to an IAM identity (identity-based policies), you can use the following list as a reference. The list includes each Amazon Polly API operation, the corresponding actions for which you can grant permissions to perform the action, and the AWS resource for which you can grant the permissions. You specify the actions in the policy's `Action` field, and you specify the resource value in the policy's `Resource` field.

You can use AWS-wide condition keys in your Amazon Polly policies to express conditions. For a complete list of AWS-wide keys, see Available Keys in the IAM User Guide.

**Note**

To specify an action, use the `polly` prefix followed by the API operation name (for example, `polly:GetLexicon`).

Amazon Polly supports Identity-based policies for actions at the resource-level. Therefore, the `Resource` value is indicated by the ARN. For example: `arn:aws:polly:us-east-2:account-id:lexicon/*` as the `Resource` value specifies permissions on all owned lexicons within the us-east-2 Region.

Because Amazon Polly doesn't support permissions for actions at the resource-level, most policies specify a wildcard character (*) as the `Resource` value. However, if it is necessary to limit permissions to a specific Region this wildcard character is replaced with the appropriate ARN: `arn:aws:polly:region:account-id:lexicon/*`. 

238
Amazon Polly API and Required Permissions for Actions

API Operation: DeleteLexicon (p. 249)
Required Permissions (API Action): polly:DeleteLexicon
Resources: arn:aws:polly:region:account-id:lexicon/LexiconName

API Operation: DescribeVoices (p. 251)
Required Permissions (API Action): polly:DescribeVoices
Resources: arn:aws:polly:region:account-id:lexicon/voice-name

API Operation: GetLexicon (p. 254)
Required Permissions (API Action): polly:GetLexicon
Resources: arn:aws:polly:region:account-id:lexicon/voice-name

API Operation: ListLexicons (p. 258)
Required Permissions (API Action): polly:ListLexicons
Resources: arn:aws:polly:region:account-id:lexicon/*

API Operation: PutLexicon (p. 262)
Required Permissions (API Action): polly:ListLexicons
Resources: *

API Operation: SynthesizeSpeech (p. 270)
Required Permissions (API Action): polly:SynthesizeSpeech
Resources: *

Logging and Monitoring in Amazon Polly

Monitoring is an important part of maintaining the reliability, availability, and performance of your Amazon Polly applications. To monitor Amazon Polly API calls, you can use AWS CloudTrail. To monitor the status of your jobs, use Amazon CloudWatch Logs.

- **Amazon CloudWatch Alarms** – Using CloudWatch alarms, you watch a single metric over a time period that you specify. If the metric exceeds a given threshold, a notification is sent to an Amazon Simple Notification Service topic or AWS Auto Scaling policy. CloudWatch alarms do not invoke actions when a metric is in a particular state. Rather the state must have changed and been maintained for a specified number of periods. For more information, see Integrating CloudWatch with Amazon Polly (p. 245).

- **CloudTrail logs** – CloudTrail provides a record of actions taken by a user, role, or an AWS service in Amazon Polly. Using the information collected by CloudTrail, you can determine the request that was made to Amazon Polly. You can also determine the IP address from which the request was made, who made the request, when it was made, and additional details. For more information, see Logging Amazon Polly API Calls with AWS CloudTrail (p. 242).

Compliance Validation for Amazon Polly

Third-party auditors assess the security and compliance of Amazon Polly as part of multiple AWS compliance programs. These include SOC, PCI, FedRAMP, HIPAA, and others.
Resilience in Amazon Polly

The AWS global infrastructure is built around AWS Regions and Availability Zones. AWS Regions provide multiple physically separated and isolated Availability Zones, which are connected with low-latency, high-throughput, and highly redundant networking. With Availability Zones, you can design and operate applications and databases that automatically fail over between zones without interruption. Availability Zones are more highly available, fault tolerant, and scalable than traditional single or multiple data center infrastructures.

For more information about AWS Regions and Availability Zones, see AWS Global Infrastructure.

Infrastructure Security in Amazon Polly

As a managed service, Amazon Polly is protected by the AWS global network security procedures that are described in the Amazon Web Services: Overview of Security Processes whitepaper.

You use AWS published API calls to access Amazon Polly through the network. Clients must support Transport Layer Security (TLS) 1.0 or later. We recommend TLS 1.2 or later. Clients must also support cipher suites with perfect forward secrecy (PFS) such as Ephemeral Diffie-Hellman (DHE) or Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). Most modern systems such as Java 7 and later support these modes.

Additionally, requests must be signed by using an access key ID and a secret access key that is associated with an IAM principal. Or you can use the AWS Security Token Service (AWS STS) to generate temporary security credentials to sign requests.

Security Best Practices for Amazon Polly

Your trust, privacy, and the security of your content are our highest priorities. We implement responsible and sophisticated technical and physical controls designed to prevent unauthorized access to, or
disclosure of, your content and ensure that our use complies with our commitments to you. For more information, see AWS Data Privacy FAQ.

Amazon Polly does not retain the content of text submissions.

For a broad view of AWS security, including compliance, penetration testing, bulletins, and resources, visit the AWS Cloud Security website.
Logging Amazon Polly API Calls with AWS CloudTrail

Amazon Polly is integrated with AWS CloudTrail, a service that provides a record of actions taken by a user, role, or an AWS service in Amazon Polly. CloudTrail captures all API calls for Amazon Polly as events. The calls captured include calls from the Amazon Polly console and code calls to the Amazon Polly API operations. If you create a trail, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon Polly. If you don't configure a trail, you can still view the most recent events in the CloudTrail console in Event history. Using the information collected by CloudTrail, you can determine the request that was made to Amazon Polly, the IP address from which the request was made, who made the request, when it was made, and additional details.

To learn more about CloudTrail, including how to configure and enable it, see the AWS CloudTrail User Guide.

Amazon Polly Information in CloudTrail

CloudTrail is enabled on your AWS account when you create the account. When supported event activity occurs in Amazon Polly, that activity is recorded in a CloudTrail event along with other AWS service events in Event history. You can view, search, and download recent events in your AWS account. For more information, see Viewing Events with CloudTrail Event History.

For an ongoing record of events in your AWS account, including events for Amazon Polly, create a trail. A trail enables CloudTrail to deliver log files to an Amazon S3 bucket. By default, when you create a trail in the console, the trail applies to all AWS Regions. The trail logs events from all Regions in the AWS partition and delivers the log files to the Amazon S3 bucket that you specify. Additionally, you can configure other AWS services to further analyze and act upon the event data collected in CloudTrail logs. For more information, see the following:

- Overview for Creating a Trail
- CloudTrail Supported Services and Integrations
- Configuring Amazon SNS Notifications for CloudTrail
- Receiving CloudTrail Log Files from Multiple Regions and Receiving CloudTrail Log Files from Multiple Accounts

Amazon Polly supports logging the following actions as events in CloudTrail log files:

- DeleteLexicon (p. 249)
- DescribeVoices (p. 251)
- GetLexicon (p. 254)
- GetSpeechSynthesisTask (p. 256)
- ListLexicons (p. 258)
- ListSpeechSynthesisTasks (p. 260)
- PutLexicon (p. 262)
- StartSpeechSynthesisTask (p. 264)
- SynthesizeSpeech (p. 270)
Every event or log entry contains information about who generated the request. The identity information helps you determine the following:

- Whether the request was made with root or AWS Identity and Access Management (IAM) user credentials.
- Whether the request was made with temporary security credentials for a role or federated user.
- Whether the request was made by another AWS service.

For more information, see the CloudTrail userIdentity Element.

Example: Amazon Polly Log File Entries

A trail is a configuration that enables delivery of events as log files to an Amazon S3 bucket that you specify. CloudTrail log files contain one or more log entries. An event represents a single request from any source and includes information about the requested action, the date and time of the action, request parameters, and so on. CloudTrail log files aren't an ordered stack trace of the public API calls, so they don't appear in any specific order.

The following example shows a CloudTrail log entry that demonstrates the SynthesizeSpeech.

```json
{
  "Records": [
    {
      "awsRegion": "us-east-2",
      "eventID": "19bd70f7-5e60-4cdc-9825-936c552278ae",
      "eventName": "SynthesizeSpeech",
      "eventSource": "polly.amazonaws.com",
      "eventTime": "2016-11-02T03:49:39Z",
      "eventType": "AwsApiCall",
      "eventVersion": "1.05",
      "recipientAccountId": "123456789012",
      "requestID": "414288c2-a1af-11e6-b17f-d7cfc06cb461",
      "requestParameters": {
        "lexiconNames": [
          "SampleLexicon"
        ],
        "outputFormat": "mp3",
        "sampleRate": "22050",
        "text": "**********",
        "textType": "text",
        "voiceId": "Kendra"
      },
      "responseElements": {
        "contentType": "audio/mpeg",
        "requestCharacters": 25
      },
      "sourceIPAddress": "1.2.3.4",
      "userAgent": "Amazon CLI/Polly 1.10 API 2016-06-10",
      "userIdentity": {
        "accessKeyId": "EXAMPLE_KEY_ID",
        "accountId": "123456789012",
        "arn": "arn:aws:iam::123456789012:user/Alice",
        "principalId": "EX_PRINCIPAL_ID",
        "type": "IAMUser",
        "userName": "Alice"
      }
    }
  ]
}
```

243
Integrating CloudWatch with Amazon Polly

When you interact with Amazon Polly, it sends the following metrics and dimensions to CloudWatch every minute. You can use the following procedures to view the metrics for Amazon Polly.

You can monitor Amazon Polly using CloudWatch, which collects and processes raw data from Amazon Polly into readable, near real-time metrics. These statistics are recorded for a period of two weeks, so that you can access historical information and gain a better perspective on how your web application or service is performing. By default, Amazon Polly metric data is sent to CloudWatch in 1 minute intervals. For more information, see What Is Amazon CloudWatch in the Amazon CloudWatch User Guide.

Getting CloudWatch Metrics (Console)

2. In the navigation pane, choose Metrics.
3. In the CloudWatch Metrics by Category pane, under the metrics category for Amazon Polly, select a metrics category, and then in the upper pane, scroll down to view the full list of metrics.

Getting CloudWatch Metrics (CLI)

The following code display available metrics for Amazon Polly.

```
aws cloudwatch list-metrics --namespace "AWS/Polly"
```

The preceding command returns a list of Amazon Polly metrics similar to the following. The MetricName element identifies what the metric is.

```json
{
  "Metrics": [
    {
      "Namespace": "AWS/Polly",
      "Dimensions": [
        {
          "Name": "Operation",
          "Value": "SynthesizeSpeech"
        }
      ],
      "MetricName": "ResponseLatency"
    },
    {
      "Namespace": "AWS/Polly",
      "Dimensions": [
        {
          "Name": "Operation",
          "Value": "SynthesizeSpeech"
        }
      ],
      "MetricName": "SynthesizeSpeech"
    }
  ]
}
```
For more information, see `GetMetricStatistics` in the *Amazon CloudWatch API Reference*.

## Amazon Polly Metrics

Amazon Polly produces the following metrics for each request. These metrics are aggregated and in one minute intervals sent to CloudWatch where they are available.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RequestCharacters</td>
<td>The number of characters in the request. This is billable characters only and does not include SSML tags.</td>
</tr>
<tr>
<td></td>
<td>Valid Dimension: Operation</td>
</tr>
<tr>
<td></td>
<td>Valid Statistics: Minimum, Maximum, Average, SampleCount, Sum</td>
</tr>
<tr>
<td></td>
<td>Unit: Count</td>
</tr>
<tr>
<td>ResponseLatency</td>
<td>The latency between when the request was made and the start of the streaming response.</td>
</tr>
<tr>
<td></td>
<td>Valid Dimensions: Operation</td>
</tr>
<tr>
<td></td>
<td>Valid Statistics: Minimum, Maximum, Average, SampleCount</td>
</tr>
<tr>
<td></td>
<td>Unit: milliseconds</td>
</tr>
<tr>
<td>2XXCount</td>
<td>HTTP 200 level code returned upon a successful response.</td>
</tr>
<tr>
<td></td>
<td>Valid Dimensions: Operation</td>
</tr>
<tr>
<td></td>
<td>Valid Statistics: Average, SampleCount, Sum</td>
</tr>
<tr>
<td></td>
<td>Unit: Count</td>
</tr>
<tr>
<td>4XXCount</td>
<td>HTTP 400 level error code returned upon an error. For each successful response, a zero (0) is emitted.</td>
</tr>
<tr>
<td></td>
<td>Valid Dimensions: Operation</td>
</tr>
<tr>
<td></td>
<td>Valid Statistics: Average, SampleCount, Sum</td>
</tr>
<tr>
<td></td>
<td>Unit: Count</td>
</tr>
<tr>
<td>5XXCount</td>
<td>HTTP 500 level error code returned upon an error. For each successful response, a zero (0) is emitted.</td>
</tr>
<tr>
<td></td>
<td>Valid Dimensions: Operation</td>
</tr>
<tr>
<td></td>
<td>Valid Statistics: Average, SampleCount, Sum</td>
</tr>
<tr>
<td></td>
<td>Unit: Count</td>
</tr>
</tbody>
</table>

246
## Dimensions for Amazon Polly Metrics

Amazon Polly metrics use the AWS/Polly namespace and provide metrics for the following dimension:

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation</td>
<td>Metrics are grouped by the API method they refer to. Possible values are SynthesizeSpeech, PutLexicon, DescribeVoices, etc.</td>
</tr>
</tbody>
</table>
Amazon Polly API Reference

This section contains the Amazon Polly API reference.

**Note**
Authenticated API calls must be signed using the Signature Version 4 Signing Process. For more information, see Signing AWS API Requests in the Amazon Web Services General Reference.

**Topics**
- Actions (p. 248)
- Data Types (p. 274)

**Actions**

The following actions are supported:

- **DeleteLexicon** (p. 249)
- **DescribeVoices** (p. 251)
- **GetLexicon** (p. 254)
- **GetSpeechSynthesisTask** (p. 256)
- **ListLexicons** (p. 258)
- **ListSpeechSynthesisTasks** (p. 260)
- **PutLexicon** (p. 262)
- **StartSpeechSynthesisTask** (p. 264)
- **SynthesizeSpeech** (p. 270)
DeleteLexicon

Deletes the specified pronunciation lexicon stored in an AWS Region. A lexicon which has been deleted is not available for speech synthesis, nor is it possible to retrieve it using either the GetLexicon or ListLexicon APIs.

For more information, see Managing Lexicons.

Request Syntax

```
DELETE /v1/lexicons/LexiconName HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

**LexiconName (p. 249)**

The name of the lexicon to delete. Must be an existing lexicon in the region.

Pattern: [0-9A-Za-z]{1,20}

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

**LexiconNotFoundException**

Amazon Polly can't find the specified lexicon. This could be caused by a lexicon that is missing, its name is misspelled or specifying a lexicon that is in a different region.

Verify that the lexicon exists, is in the region (see ListLexicons (p. 258)) and that you spelled its name is spelled correctly. Then try again.

HTTP Status Code: 404

**ServiceFailureException**

An unknown condition has caused a service failure.

HTTP Status Code: 500
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeVoices

Returns the list of voices that are available for use when requesting speech synthesis. Each voice speaks a specified language, is either male or female, and is identified by an ID, which is the ASCII version of the voice name.

When synthesizing speech (SynthesizeSpeech), you provide the voice ID for the voice you want from the list of voices returned by DescribeVoices.

For example, you want your news reader application to read news in a specific language, but giving a user the option to choose the voice. Using the DescribeVoices operation you can provide the user with a list of available voices to select from.

You can optionally specify a language code to filter the available voices. For example, if you specify en-US, the operation returns a list of all available US English voices.

This operation requires permissions to perform the polly:DescribeVoices action.

Request Syntax

GET /v1/voices?
Engine=Engine&IncludeAdditionalLanguageCodes=IncludeAdditionalLanguageCodes&LanguageCode=LanguageCode&NextToken=NextToken
HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

**Engine** (p. 251)

Specifies the engine (standard or neural) used by Amazon Polly when processing input text for speech synthesis.

Valid Values: standard | neural

**IncludeAdditionalLanguageCodes** (p. 251)

Boolean value indicating whether to return any bilingual voices that use the specified language as an additional language. For instance, if you request all languages that use US English (es-US), and there is an Italian voice that speaks both Italian (it-IT) and US English, that voice will be included if you specify yes but not if you specify no.

**LanguageCode** (p. 251)

The language identification tag (ISO 639 code for the language name-ISO 3166 country code) for filtering the list of voices returned. If you don't specify this optional parameter, all available voices are returned.


**NextToken** (p. 251)

An opaque pagination token returned from the previous DescribeVoices operation. If present, this indicates where to continue the listing.

Length Constraints: Minimum length of 0. Maximum length of 4096.
Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json
{
  "NextToken": "string",
  "Voices": [
    {
      "AdditionalLanguageCodes": [ "string" ],
      "Gender": "string",
      "Id": "string",
      "LanguageCode": "string",
      "LanguageName": "string",
      "Name": "string",
      "SupportedEngines": [ "string" ]
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 252)**

The pagination token to use in the next request to continue the listing of voices. NextToken is returned only if the response is truncated.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 4096.

**Voices (p. 252)**

A list of voices with their properties.

Type: Array of Voice (p. 282) objects

Errors

**InvalidNextTokenException**

The NextToken is invalid. Verify that it's spelled correctly, and then try again.

HTTP Status Code: 400

**ServiceFailureException**

An unknown condition has caused a service failure.

HTTP Status Code: 500
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetLexicon

Returns the content of the specified pronunciation lexicon stored in an AWS Region. For more information, see Managing Lexicons.

Request Syntax

```
GET /v1/lexicons/LexiconName HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

**LexiconName (p. 254)**

Name of the lexicon.

Pattern: [0-9A-Za-z]{1,20}

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json
{
    "Lexicon": {
        "Content": "string",
        "Name": "string"
    },
    "LexiconAttributes": {
        "Alphabet": "string",
        "LanguageCode": "string",
        "LastModified": number,
        "LexemesCount": number,
        "LexiconArn": "string",
        "Size": number
    }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Lexicon (p. 254)**

Lexicon object that provides name and the string content of the lexicon.

Type: Lexicon (p. 275) object
LexiconAttributes (p. 254)

Metadata of the lexicon, including phonetic alphabetic used, language code, lexicon ARN, number of lexemes defined in the lexicon, and size of lexicon in bytes.

Type: LexiconAttributes (p. 276) object

Errors

LexiconNotFoundException

Amazon Polly can't find the specified lexicon. This could be caused by a lexicon that is missing, its name is misspelled or specifying a lexicon that is in a different region.

Verify that the lexicon exists, is in the region (see ListLexicons (p. 258)) and that you spelled its name is spelled correctly. Then try again.

HTTP Status Code: 404

ServiceFailureException

An unknown condition has caused a service failure.

HTTP Status Code: 500

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetSpeechSynthesisTask

Retrieves a specific SpeechSynthesisTask object based on its TaskId. This object contains information about the given speech synthesis task, including the status of the task, and a link to the S3 bucket containing the output of the task.

Request Syntax

```
GET /v1/synthesisTasks/TaskId HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

**TaskId (p. 256)**

The Amazon Polly generated identifier for a speech synthesis task.

Pattern: ^[a-zA-Z0-9-]{1,100}#

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "SynthesisTask": {
      "CreationTime": number,
      "Engine": "string",
      "LanguageCode": "string",
      "LexiconNames": [ "string" ],
      "OutputFormat": "string",
      "OutputUri": "string",
      "RequestCharacters": number,
      "SampleRate": "string",
      "SnsTopicArn": "string",
      "SpeechMarkTypes": [ "string" ],
      "TaskId": "string",
      "TaskStatus": "string",
      "TaskStatusReason": "string",
      "TextType": "string",
      "VoiceId": "string"
   }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**SynthesisTask** (p. 256)

SynthesisTask object that provides information from the requested task, including output format, creation time, task status, and so on.

Type: SynthesisTask (p. 279) object

**Errors**

**InvalidTaskIdException**

The provided Task ID is not valid. Please provide a valid Task ID and try again.

HTTP Status Code: 400

**ServiceFailureException**

An unknown condition has caused a service failure.

HTTP Status Code: 500

**SynthesisTaskNotFoundException**

The Speech Synthesis task with requested Task ID cannot be found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListLexicons

Returns a list of pronunciation lexicons stored in an AWS Region. For more information, see Managing Lexicons.

Request Syntax

GET /v1/lexicons?NextToken=NextToken HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

NextToken (p. 258)

An opaque pagination token returned from previous ListLexicons operation. If present, indicates where to continue the list of lexicons.

Length Constraints: Minimum length of 0. Maximum length of 4096.

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200
Content-type: application/json

{  
  "Lexicons": [  
    {  
      "Attributes": {  
        "Alphabet": "string",
        "LanguageCode": "string",
        "LastModified": number,
        "LexemesCount": number,
        "LexiconArn": "string",
        "Size": number
      },  
      "Name": "string"
    }  
  ],  
  "NextToken": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Lexicons (p. 258)

A list of lexicon names and attributes.
Type: Array of LexiconDescription (p. 278) objects

NextToken (p. 258)

The pagination token to use in the next request to continue the listing of lexicons. NextToken is returned only if the response is truncated.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 4096.

Errors

InvalidNextTokenException

The NextToken is invalid. Verify that it's spelled correctly, and then try again.

HTTP Status Code: 400

ServiceFailureException

An unknown condition has caused a service failure.

HTTP Status Code: 500

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListSpeechSynthesisTasks

Returns a list of SpeechSynthesisTask objects ordered by their creation date. This operation can filter the tasks by their status, for example, allowing users to list only tasks that are completed.

Request Syntax

GET /v1/synthesisTasks?MaxResults=MaxResults&NextToken=NextToken&Status=Status HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

MaxResults (p. 260)

Maximum number of speech synthesis tasks returned in a List operation.

Valid Range: Minimum value of 1. Maximum value of 100.

NextToken (p. 260)

The pagination token to use in the next request to continue the listing of speech synthesis tasks.

Length Constraints: Minimum length of 0. Maximum length of 4096.

Status (p. 260)

Status of the speech synthesis tasks returned in a List operation

Valid Values: scheduled | inProgress | completed | failed

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200
Content-type: application/json

{  "NextToken": "string",
  "SynthesisTasks": [
  {"CreationTime": number,
   "Engine": "string",
   "LanguageCode": "string",
   "LexiconNames": ["string"],
   "OutputFormat": "string",
   "OutputUri": "string",
   "RequestCharacters": number,
   "SampleRate": "string",
   "SnsTopicArn": "string",
   "SpeechMarkTypes": ["string"],
   "TaskId": "string",
   "TaskStatus": "string",
   "TaskStatusReason": "string",
   "TextType": "string",
   "VoiceId": "string"}  ]}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

NextToken (p. 260)

An opaque pagination token returned from the previous List operation in this request. If present, this indicates where to continue the listing.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 4096.

SynthesisTasks (p. 260)

List of SynthesisTask objects that provides information from the specified task in the list request, including output format, creation time, task status, and so on.

Type: Array of SynthesisTask (p. 279) objects

Errors

InvalidNextTokenException

The NextToken is invalid. Verify that it's spelled correctly, and then try again.

HTTP Status Code: 400

ServiceFailureException

An unknown condition has caused a service failure.

HTTP Status Code: 500

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
PutLexicon

Stores a pronunciation lexicon in an AWS Region. If a lexicon with the same name already exists in the region, it is overwritten by the new lexicon. Lexicon operations have eventual consistency, therefore, it might take some time before the lexicon is available to the SynthesizeSpeech operation.

For more information, see Managing Lexicons.

Request Syntax

```
PUT /v1/lexicons/LexiconName HTTP/1.1
Content-type: application/json
{
    "Content": "string"
}
```

URI Request Parameters

The request uses the following URI parameters.

LexiconName (p. 262)

Name of the lexicon. The name must follow the regular express format [0-9A-Za-z][1,20]. That is, the name is a case-sensitive alphanumeric string up to 20 characters long.

Pattern: [0-9A-Za-z]{1,20}

Required: Yes

Request Body

The request accepts the following data in JSON format.

Content (p. 262)

Content of the PLS lexicon as string data.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

InvalidLexiconException

Amazon Polly can't find the specified lexicon. Verify that the lexicon's name is spelled correctly, and then try again.
HTTP Status Code: 400
**LexiconSizeExceeded**Exception

The maximum size of the specified lexicon would be exceeded by this operation.

HTTP Status Code: 400
**MaxLexemeLengthExceeded**Exception

The maximum size of the lexeme would be exceeded by this operation.

HTTP Status Code: 400
**MaxLexiconsNumberExceeded**Exception

The maximum number of lexicons would be exceeded by this operation.

HTTP Status Code: 400
**ServiceFailure**Exception

An unknown condition has caused a service failure.

HTTP Status Code: 500
**UnsupportedPlsAlphabet**Exception

The alphabet specified by the lexicon is not a supported alphabet. Valid values are x–sampa and ipa.

HTTP Status Code: 400
**UnsupportedPlsLanguage**Exception

The language specified in the lexicon is unsupported. For a list of supported languages, see Lexicon Attributes.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StartSpeechSynthesisTask

Allows the creation of an asynchronous synthesis task, by starting a new SpeechSynthesisTask. This operation requires all the standard information needed for speech synthesis, plus the name of an Amazon S3 bucket for the service to store the output of the synthesis task and two optional parameters (OutputS3KeyPrefix and SnsTopicArn). Once the synthesis task is created, this operation will return a SpeechSynthesisTask object, which will include an identifier of this task as well as the current status. The SpeechSynthesisTask object is available for 72 hours after starting the asynchronous synthesis task.

Request Syntax

```
POST /v1/synthesisTasks HTTP/1.1
Content-type: application/json

{
    "Engine": "string",
    "LanguageCode": "string",
    "LexiconNames": [ "string" ],
    "OutputFormat": "string",
    "OutputS3BucketName": "string",
    "OutputS3KeyPrefix": "string",
    "SampleRate": "string",
    "SnsTopicArn": "string",
    "SpeechMarkTypes": [ "string" ],
    "Text": "string",
    "TextType": "string",
    "VoiceId": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**Engine (p. 264)**

Specifies the engine (standard or neural) for Amazon Polly to use when processing input text for speech synthesis. Using a voice that is not supported for the engine selected will result in an error.

Type: String

Valid Values: standard | neural

Required: No

**LanguageCode (p. 264)**

Optional language code for the Speech Synthesis request. This is only necessary if using a bilingual voice, such as Aditi, which can be used for either Indian English (en-IN) or Hindi (hi-IN).

If a bilingual voice is used and no language code is specified, Amazon Polly uses the default language of the bilingual voice. The default language for any voice is the one returned by the DescribeVoices operation for the LanguageCode parameter. For example, if no language code is specified, Aditi will use Indian English rather than Hindi.
StartSpeechSynthesisTask

**Type:** String

**Valid Values:**
- arb
- cmn-CN
- cy-GB
- da-DK
- de-DE
- en-AU
- en-GB
- en-GB-WLS
- en-IN
- en-US
- es-ES
- es-MX
- es-US
- fr-CA
- fr-FR
- is-IS
- it-IT
- ja-JP
- hi-IN
- ko-KR
- nb-NO
- nl-NL
- pl-PL
- pt-BR
- pt-PT
- ro-RO
- ru-RU
- sv-SE
- tr-TR
- en-NZ
- en-ZA

**Required:** No

**LexiconNames (p. 264)**

List of one or more pronunciation lexicon names you want the service to apply during synthesis. Lexicons are applied only if the language of the lexicon is the same as the language of the voice.

**Type:** Array of strings

**Array Members:** Maximum number of 5 items.

**Pattern:** `[0-9A-Za-z]{1,20}`

**Required:** No

**OutputFormat (p. 264)**

The format in which the returned output will be encoded. For audio stream, this will be mp3, ogg_vorbis, or pcm. For speech marks, this will be json.

**Type:** String

**Valid Values:**
- json
- mp3
- ogg_vorbis
- pcm

**Required:** Yes

**OutputS3BucketName (p. 264)**

Amazon S3 bucket name to which the output file will be saved.

**Type:** String

**Pattern:** `^[a-z0-9][\./-a-z0-9]{1,61}[a-z0-9]$`

**Required:** Yes

**OutputS3KeyPrefix (p. 264)**

The Amazon S3 key prefix for the output speech file.

**Type:** String

**Pattern:** `^[0-9a-zA-Z/\%!\_-\./\*\:\;\@=\+$\?\&]{0,800}$`

**Required:** No

**SampleRate (p. 264)**

The audio frequency specified in Hz.

The valid values for mp3 and ogg_vorbis are "8000", "16000", "22050", and "24000". The default value for standard voices is "22050". The default value for neural voices is "24000".

Valid values for pcm are "8000" and "16000" The default value is "16000".

**Type:** String

**Required:** No
SnsTopicArn (p. 264)

ARN for the SNS topic optionally used for providing status notification for a speech synthesis task.

Type: String

Pattern: ^arn:aws(-\[(cn|iso(-b)?|us-gov)\]?\:\[a-z0-9-]{1,50}::\d{12}::[a-zA-Z0-9-_]{1,256}$

Required: No

SpeechMarkTypes (p. 264)

The type of speech marks returned for the input text.

Type: Array of strings

Array Members: Maximum number of 4 items.

Valid Values: sentence | ssml | viseme | word

Required: No

Text (p. 264)

The input text to synthesize. If you specify ssml as the TextType, follow the SSML format for the input text.

Type: String

Required: Yes

TextType (p. 264)

Specifies whether the input text is plain text or SSML. The default value is plain text.

Type: String

Valid Values: ssml | text

Required: No

VoiceId (p. 264)

Voice ID to use for the synthesis.

Type: String

Valid Values: Aditi | Amy | Astrid | Bianca | Brian | Camila | Carla | Carmen | Celine | Chantal | Conchita | Cristiano | Dora | Emma | Enrique | Ewa | Filiz | Gabrielle | Geraint | Giorgio | Gwyneth | Hans | Ines | Ivy | Jacek | Jan | Joanna | Joey | Justin | Karl | Kendra | Kevin | Kimberly | Lea | Liv | Lotte | Lucia | Lupe | Mads | Maja | Marlene | Mathieu | Matthew | Maxim | Mia | Miguel | Mizuki | Naja | Nicole | Olivia | Penelope | Raveena | Ricardo | Ruben | Russell | Salli | Seoyeon | Takumi | Tatyana | Vicki | Vitoria | Zeina | Zhiyu | Aria | Ayanda

Required: Yes

Response Syntax

HTTP/1.1 200
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**SynthesisTask (p. 266)**

SynthesisTask object that provides information and attributes about a newly submitted speech synthesis task.

Type: SynthesisTask (p. 279) object

Errors

**EngineNotSupportedException**

This engine is not compatible with the voice that you have designated. Choose a new voice that is compatible with the engine or change the engine and restart the operation.

HTTP Status Code: 400

**InvalidS3BucketException**

The provided Amazon S3 bucket name is invalid. Please check your input with S3 bucket naming requirements and try again.

HTTP Status Code: 400

**InvalidS3KeyException**

The provided Amazon S3 key prefix is invalid. Please provide a valid S3 object key name.

HTTP Status Code: 400

**InvalidSampleRateException**

The specified sample rate is not valid.

HTTP Status Code: 400
InvalidSnsTopicArnException

The provided SNS topic ARN is invalid. Please provide a valid SNS topic ARN and try again.

HTTP Status Code: 400

InvalidSsmlException

The SSML you provided is invalid. Verify the SSML syntax, spelling of tags and values, and then try again.

HTTP Status Code: 400

LanguageNotSupportedException

The language specified is not currently supported by Amazon Polly in this capacity.

HTTP Status Code: 400

LexiconNotFoundException

Amazon Polly can't find the specified lexicon. This could be caused by a lexicon that is missing, its name is misspelled or specifying a lexicon that is in a different region.

Verify that the lexicon exists, is in the region (see ListLexicons (p. 258)) and that you spelled its name is spelled correctly. Then try again.

HTTP Status Code: 404

MarksNotSupportedException

Speech marks are not supported for the OutputFormat selected. Speech marks are only available for content in json format.

HTTP Status Code: 400

ServiceFailureException

An unknown condition has caused a service failure.

HTTP Status Code: 500

SsmlMarksNotSupportedException

SSML speech marks are not supported for plain text-type input.

HTTP Status Code: 400

TextLengthExceededException

The value of the "Text" parameter is longer than the accepted limits. For the SynthesizeSpeech API, the limit for input text is a maximum of 6000 characters total, of which no more than 3000 can be billed characters. For the StartSpeechSynthesisTask API, the maximum is 200,000 characters, of which no more than 100,000 can be billed characters. SSML tags are not counted as billed characters.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
SynthesizeSpeech

Synthesizes UTF-8 input, plain text or SSML, to a stream of bytes. SSML input must be valid, well-formed SSML. Some alphabets might not be available with all the voices (for example, Cyrillic might not be read at all by English voices) unless phoneme mapping is used. For more information, see How it Works.

Request Syntax

```json
POST /v1/speech HTTP/1.1
Content-type: application/json

{
  "Engine": "string",
  "LanguageCode": "string",
  "LexiconNames": [ "string" ],
  "OutputFormat": "string",
  "SampleRate": "string",
  "SpeechMarkTypes": [ "string" ],
  "Text": "string",
  "TextType": "string",
  "VoiceId": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

Engine  (p. 270)

Specifies the engine (standard or neural) for Amazon Polly to use when processing input text for speech synthesis. For information on Amazon Polly voices and which voices are available in standard-only, NTTS-only, and both standard and NTTS formats, see Available Voices.

NTTS-only voices

When using NTTS-only voices such as Kevin (en-US), this parameter is required and must be set to neural. If the engine is not specified, or is set to standard, this will result in an error.

Type: String

Valid Values: standard | neural

Required: Yes

Standard voices

For standard voices, this is not required; the engine parameter defaults to standard. If the engine is not specified, or is set to standard and an NTTS-only voice is selected, this will result in an error.

Type: String

Valid Values: standard | neural

Required: No
LanguageCode (p. 270)

Optional language code for the Synthesize Speech request. This is only necessary if using a bilingual voice, such as Aditi, which can be used for either Indian English (en-IN) or Hindi (hi-IN).

If a bilingual voice is used and no language code is specified, Amazon Polly uses the default language of the bilingual voice. The default language for any voice is the one returned by the DescribeVoices operation for the LanguageCode parameter. For example, if no language code is specified, Aditi will use Indian English rather than Hindi.

Type: String

Valid Values:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>arb</td>
<td>cmn-CN</td>
</tr>
<tr>
<td>cy-GB</td>
<td>da-DK</td>
</tr>
<tr>
<td>de-DE</td>
<td>en-AU</td>
</tr>
<tr>
<td>en-GB</td>
<td>en-GB-WLS</td>
</tr>
<tr>
<td>ja-JP</td>
<td>hi-IN</td>
</tr>
<tr>
<td>ko-KR</td>
<td>nb-NO</td>
</tr>
<tr>
<td>nl-NL</td>
<td>pl-PL</td>
</tr>
<tr>
<td>pt-BR</td>
<td>pt-PT</td>
</tr>
<tr>
<td>ro-RO</td>
<td>ru-RU</td>
</tr>
<tr>
<td>sv-SE</td>
<td>tr-TR</td>
</tr>
<tr>
<td>en-NZ</td>
<td>en-ZA</td>
</tr>
</tbody>
</table>

Required: No

LexiconNames (p. 270)

List of one or more pronunciation lexicon names you want the service to apply during synthesis. Lexicons are applied only if the language of the lexicon is the same as the language of the voice. For information about storing lexicons, see PutLexicon.

Type: Array of strings

Array Members: Maximum number of 5 items.

Pattern: [0-9A-Za-z]{1,20}

Required: No

OutputFormat (p. 270)

The format in which the returned output will be encoded. For audio stream, this will be mp3, ogg_vorbis, or pcm. For speech marks, this will be json.

When pcm is used, the content returned is audio/pcm in a signed 16-bit, 1 channel (mono), little-endian format.

Type: String

Valid Values: json | mp3 | ogg_vorbis | pcm

Required: Yes

SampleRate (p. 270)

The audio frequency specified in Hz.

The valid values for mp3 and ogg_vorbis are "8000", "16000", "22050", and "24000". The default value for standard voices is "22050". The default value for neural voices is "24000".

Valid values for pcm are "8000" and "16000" The default value is "16000".

Type: String

Required: No

SpeechMarkTypes (p. 270)

The type of speech marks returned for the input text.

Type: Array of strings
Array Members: Maximum number of 4 items.

Valid Values: sentence | ssml | viseme | word

Required: No

**Text (p. 270)**

Input text to synthesize. If you specify ssml as the TextType, follow the SSML format for the input text.

Type: String

Required: Yes

**TextType (p. 270)**

Specifies whether the input text is plain text or SSML. The default value is plain text. For more information, see Using SSML.

Type: String

Valid Values: ssml | text

Required: No

**VoiceId (p. 270)**

Voice ID to use for the synthesis. You can get a list of available voice IDs by calling the DescribeVoices operation.

Type: String

Valid Values: Aditi | Amy | Astrid | Bianca | Brian | Camila | Carla | Carmen | Celine | Chantal | Conchita | Cristiano | Dora | Emma | Enrique | Ewa | Filiz | Gabrielle | Geraint | Giorgio | Gwyneth | Hans | Ines | Ivy | Jacek | Jan | Joanna | Joey | Justin | Karl | Kendra | Kevin | Kimberly | Lea | Liv | Lotte | Lucia | Lupe | Mads | Maja | Marlene | Mathieu | Matthew | Maxim | Mia | Miguel | Mizuki | Naja | Nicole | Olivia | Penelope | Raveena | Ricardo | Ruben | Russell | Salli | Seoyeon | Takumi | Tatyana | Vicki | Vitoria | Zeina | Zhiyu | Aria | Ayanda

Required: Yes

**Response Syntax**

HTTP/1.1 200
Content-Type: Content-Type
x-amzn-RequestCharacters: RequestCharacters

AudioStream

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.

**ContentType (p. 272)**

Specifies the type audio stream. This should reflect the OutputFormat parameter in your request.
• If you request mp3 as the OutputFormat, the ContentType returned is audio/mpeg.
• If you request ogg_vorbis as the OutputFormat, the ContentType returned is audio/ogg.
• If you request pcm as the OutputFormat, the ContentType returned is audio/pcm in a signed 16-bit, 1 channel (mono), little-endian format.
• If you request json as the OutputFormat, the ContentType returned is audio/json.

RequestCharacters (p. 272)
Number of characters synthesized.

The response returns the following as the HTTP body.

AudioStream (p. 272)
Stream containing the synthesized speech.

Errors

EngineNotSupportedException
This engine is not compatible with the voice that you have designated. Choose a new voice that is compatible with the engine or change the engine and restart the operation.

HTTP Status Code: 400

InvalidSampleRateException
The specified sample rate is not valid.

HTTP Status Code: 400

InvalidSsmlException
The SSML you provided is invalid. Verify the SSML syntax, spelling of tags and values, and then try again.

HTTP Status Code: 400

LanguageNotSupportedException
The language specified is not currently supported by Amazon Polly in this capacity.

HTTP Status Code: 400

LexiconNotFoundException
Amazon Polly can't find the specified lexicon. This could be caused by a lexicon that is missing, its name is misspelled or specifying a lexicon that is in a different region.

Verify that the lexicon exists, is in the region (see ListLexicons (p. 258)) and that you spelled its name is spelled correctly. Then try again.

HTTP Status Code: 404

MarksNotSupportedForFormatException
Speech marks are not supported for the OutputFormat selected. Speech marks are only available for content in json format.

HTTP Status Code: 400
ServiceFailureException
An unknown condition has caused a service failure.
HTTP Status Code: 500

SsmlMarksNotSupportedForTextTypeException
SSML speech marks are not supported for plain text-type input.
HTTP Status Code: 400

TextLengthExceededException
The value of the "Text" parameter is longer than the accepted limits. For the SynthesizeSpeech API, the limit for input text is a maximum of 6000 characters total, of which no more than 3000 can be billed characters. For the StartSpeechSynthesisTask API, the maximum is 200,000 characters, of which no more than 100,000 can be billed characters. SSML tags are not counted as billed characters.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

Data Types
The following data types are supported:

- Lexicon (p. 275)
- LexiconAttributes (p. 276)
- LexiconDescription (p. 278)
- SynthesisTask (p. 279)
- Voice (p. 282)
Lexicon

Provides lexicon name and lexicon content in string format. For more information, see Pronunciation Lexicon Specification (PLS) Version 1.0.

Contents

Content

Lexicon content in string format. The content of a lexicon must be in PLS format.

Type: String

Required: No

Name

Name of the lexicon.

Type: String

Pattern: [0–9A–Za–z]{1,20}

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LexiconAttributes

Contains metadata describing the lexicon such as the number of lexemes, language code, and so on. For more information, see Managing Lexicons.

Contents

Alphabet

Phonetic alphabet used in the lexicon. Valid values are ipa and x-sampa.

Type: String
Required: No

LanguageCode

Language code that the lexicon applies to. A lexicon with a language code such as "en" would be applied to all English languages (en-GB, en-US, en-AUS, en-WLS, and so on.

Type: String
Valid Values:
- arb
- cmn-CN
- cy-GB
- da-DK
- de-DE
- en-AU
- en-GB
- en-GB-WLS
- en-IN
- en-US
- es-ES
- es-MX
- es-US
- fr-CA
- fr-FR
- is-IS
- it-IT
- ja-JP
- hi-IN
- ko-KR
- nb-NO
- nl-NL
- pl-PL
- pt-BR
- pt-PT
- ro-RO
- ru-RU
- sv-SE
- tr-TR
- en-NZ
- en-ZA
Required: No

LastModified

Date lexicon was last modified (a timestamp value).

Type: Timestamp
Required: No

LexemesCount

Number of lexemes in the lexicon.

Type: Integer
Required: No

LexiconArn

Amazon Resource Name (ARN) of the lexicon.

Type: String
Required: No

Size

Total size of the lexicon, in characters.

Type: Integer
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LexiconDescription

Describes the content of the lexicon.

Contents

Attributes

Provides lexicon metadata.

Type: LexiconAttributes (p. 276) object

Required: No

Name

Name of the lexicon.

Type: String

Pattern: [0-9A-Za-z]{1,20}

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SynthesisTask

SynthesisTask object that provides information about a speech synthesis task.

Contents

CreationTime

Timestamp for the time the synthesis task was started.

Type: Timestamp
Required: No

Engine

Specifies the engine (standard or neural) for Amazon Polly to use when processing input text for speech synthesis. Using a voice that is not supported for the engine selected will result in an error.

Type: String
Valid Values: standard | neural
Required: No

LanguageCode

Optional language code for a synthesis task. This is only necessary if using a bilingual voice, such as Aditi, which can be used for either Indian English (en-IN) or Hindi (hi-IN).

If a bilingual voice is used and no language code is specified, Amazon Polly uses the default language of the bilingual voice. The default language for any voice is the one returned by the DescribeVoices operation for the LanguageCode parameter. For example, if no language code is specified, Aditi will use Indian English rather than Hindi.

Type: String
Required: No

LexiconNames

List of one or more pronunciation lexicon names you want the service to apply during synthesis. Lexicons are applied only if the language of the lexicon is the same as the language of the voice.

Type: Array of strings
Array Members: Maximum number of 5 items.
Pattern: [0-9A-Za-z]{1,20}
Required: No

OutputFormat

The format in which the returned output will be encoded. For audio stream, this will be mp3, ogg_vorbis, or pcm. For speech marks, this will be json.

Type: String
Valid Values: json | mp3 | ogg_vorbis | pcm

Required: No

OutputUri

Pathway for the output speech file.

Type: String

Required: No

RequestCharacters

Number of billable characters synthesized.

Type: Integer

Required: No

SampleRate

The audio frequency specified in Hz.

The valid values for mp3 and ogg_vorbis are "8000", "16000", "22050", and "24000". The default value for standard voices is "22050". The default value for neural voices is "24000".

Valid values for pcm are "8000" and "16000" The default value is "16000".

Type: String

Required: No

SnsTopicArn

ARN for the SNS topic optionally used for providing status notification for a speech synthesis task.

Type: String

Pattern: ^arn:aws(-(cn|iso(-b)?|us-gov))?:sns:[a-z0-9_-]{1,50}:\d{12}:[a-zA-Z0-9_-]{1,256}$

Required: No

SpeechMarkTypes

The type of speech marks returned for the input text.

Type: Array of strings

Array Members: Maximum number of 4 items.

Valid Values: sentence | ssml | viseme | word

Required: No

TaskId

The Amazon Polly generated identifier for a speech synthesis task.

Type: String

Pattern: ^(a-zA-Z0-9_-){1,100}$

Required: No
**TaskStatus**

Current status of the individual speech synthesis task.

Type: String

Valid Values: scheduled | inProgress | completed | failed

Required: No

**TaskStatusReason**

Reason for the current status of a specific speech synthesis task, including errors if the task has failed.

Type: String

Required: No

**TextType**

Specifies whether the input text is plain text or SSML. The default value is plain text.

Type: String

Valid Values: ssml | text

Required: No

**VoiceId**

Voice ID to use for the synthesis.

Type: String

Valid Values: Aditi | Amy | Astrid | Bianca | Brian | Camila | Carla | Carmen | Celine | Chantal | Conchita | Cristiano | Dora | Emma | Enrique | Ewa | Filiz | Gabrielle | Geraint | Giorgio | Gwyneth | Hans | Ines | Ivy | Jacek | Jan | Joanna | Joey | Justin | Karl | Kendra | Kevin | Kimberly | Lea | Liv | Lotte | Lucia | Lupe | Mads | Maja | Marlene | Mathieu | Matthew | Maxim | Mia | Miguel | Mizuki | Naja | Nicole | Olivia | Penelope | Raveena | Ricardo | Ruben | Russell | Salli | Seoyeon | Takumi | Tatyana | Vicki | Vitoria | Zeina | Zhiyu | Aria | Ayanda

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Voice

Description of the voice.

Contents

AdditionalLanguageCodes

Additional codes for languages available for the specified voice in addition to its default language.

For example, the default language for Aditi is Indian English (en-IN) because it was first used for that language. Since Aditi is bilingual and fluent in both Indian English and Hindi, this parameter would show the code hi-IN.

Type: Array of strings


Required: No

Gender

Gender of the voice.

Type: String

Valid Values: Female | Male

Required: No

Id

Amazon Polly assigned voice ID. This is the ID that you specify when calling the SynthesizeSpeech operation.

Type: String

Valid Values: Aditi | Amy | Astrid | Bianca | Brian | Camila | Carla | Carmen | Celine | Chantal | Conchita | Cristiano | Dora | Emma | Enrique | Ewa | Filiz | Gabrielle | Geraint | Giorgio | Gwyneth | Hans | Ines | Ivy | Jacek | Jan | Joanna | Joey | Justin | Karl | Kendra | Kevin | Kimberly | Lea | Liv | Lotte | Lucia | Lupe | Mads | Maja | Marlene | Mathieu | Matthew | Maxim | Mia | Miguel | Mizuki | Naja | Nicole | Olivia | Penelope | Raveena | Ricardo | Ruben | Russell | Salli | Seoyeon | Takumi | Tatyana | Vicki | Vitoria | Zeina | Zhiyu | Aria | Ayanda

Required: No

LanguageCode

Language code of the voice.

Type: String

Voice

**LanguageName**

Human readable name of the language in English.

Type: String

Required: No

**Name**

Name of the voice (for example, Salli, Kendra, etc.). This provides a human readable voice name that you might display in your application.

Type: String

Required: No

**SupportedEngines**

Specifies which engines (standard or neural) that are supported by a given voice.

Type: Array of strings

Valid Values: standard | neural

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
## Document History for Amazon Polly

The following table describes important changes in each release of the Amazon Polly Developer Guide. For notification about updates to this documentation, you can subscribe to an RSS feed.

- **Latest documentation update:** September 1, 2021

<table>
<thead>
<tr>
<th>update-history-change</th>
<th>update-history-description</th>
<th>update-history-date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New voice added for NTTS (p. 284)</strong></td>
<td>Amazon Polly now provides a new South African English voice, Ayanda. The voice is available as an NTTS voice only. See Neural voices for a list of NTTS voices.</td>
<td>September 1, 2021</td>
</tr>
<tr>
<td><strong>New Region added (p. 284)</strong></td>
<td>Amazon Polly is now available in the Africa (Cape Town) AWS Region. This Region supports neural TTS (NTTS). For more information, see Feature and Region Compatibility for a list of regions that support NTTS.</td>
<td>September 1, 2021</td>
</tr>
<tr>
<td><strong>New language and voice added (p. 284)</strong></td>
<td>Amazon Polly now supports New Zealand English (en-NZ). A new NTTS voice, Aria, speaks New Zealand English and a selection of Maori words.</td>
<td>August 24, 2021</td>
</tr>
<tr>
<td><strong>New feature (p. 284)</strong></td>
<td>Amazon Polly makes the conversational speaking style the default version for the neural Matthew and Joanna voices. We removed references to the conversational speaking style.</td>
<td>June 28, 2021</td>
</tr>
<tr>
<td><strong>New voice added for NTTS (p. 284)</strong></td>
<td>Amazon Polly now provides the German voice Vicki as an NTTS voice.</td>
<td>June 15, 2021</td>
</tr>
<tr>
<td><strong>New voice added (p. 284)</strong></td>
<td>A new female voice, Gabrielle, has been added to the French (Canadian) (fr-CA) locale. The voice is high quality and only available as an NTTS voice. Like all neural voices, it is only available in certain regions. For a list of regions, see Feature and region compatibility.</td>
<td>June 1, 2021</td>
</tr>
<tr>
<td><strong>New voice added for NTTS (p. 284)</strong></td>
<td>Amazon Polly now provides the Korean voice Seoyeon as an NTTS voice.</td>
<td>May 11, 2021</td>
</tr>
<tr>
<td>Date</td>
<td>New Region added for NTTS (p. 284)</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>March 17, 2021</td>
<td>Amazon Polly now supports neural TTS (NTTS) in the Canada (Central) AWS Region. For more information, see Feature and Region Compatibility for NTTS.</td>
<td></td>
</tr>
<tr>
<td>November 10, 2020</td>
<td>In addition to the Matthew, Joanna, and Lupe voices for the Newscaster speaking style, Amazon Polly now provides an additional option for this speaking style. Using the neural engine, you can use the Amy voice in British English for the Newscaster style. For more information, see NTTS Speaking Styles.</td>
<td></td>
</tr>
<tr>
<td>September 3, 2020</td>
<td>In addition to the existing Regions for NTTS (us-east-1, us-west-2, eu-west-1, and ap-southeast-2), neural voices are now supported in four additional Regions: (ap-northeast-1 (Tokyo), ap-southeast-1 (Singapore), eu-central-1 (Frankfurt), and eu-west-2 (London). For more information, see Feature and Region Compatibility for NTTS.</td>
<td></td>
</tr>
<tr>
<td>June 16, 2020</td>
<td>In addition to child voices Ivy and Justin, a new male child voice, Kevin, has been added to American English (en-US). This new voice is very high quality and is only available as an NTTS voice. Like all neural voices, it is only supported in four Regions: us-east-1 (N. Virginia), us-west-2 (Oregon), eu-west-1 (Ireland), and ap-southeast-2 (Sydney). For more information, see NTTS Voices.</td>
<td></td>
</tr>
<tr>
<td>April 16, 2020</td>
<td>In addition to the Matthew and Joanna voices for the Newscaster speaking style, Amazon Polly now provides an additional option for this speaking style. Using the neural engine, you can use the Lupe voice in Spanish (American) for the Newscaster style. For more information, see NTTS Speaking Styles.</td>
<td></td>
</tr>
<tr>
<td>Feature Type</td>
<td>Description</td>
<td>Date</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>New feature (p. 284)</td>
<td>In addition to the Newscaster speaking style, Amazon Polly now provides a second NTTS speaking style to help you synthesize even better text to speech passages. The Conversational style uses the neural system to generate speech in a more friendly and expressive conversational style that can be used in many use cases. For more information, see NTTS Speaking Styles.</td>
<td>November 25, 2019</td>
</tr>
<tr>
<td>New voices added (p. 284)</td>
<td>Two new voices added: Camila (female, Portuguese-Brazil) and Lupe (female, Spanish-US).</td>
<td>October 23, 2019</td>
</tr>
<tr>
<td>New feature added (p. 284)</td>
<td>Addition of Amazon Polly for Windows plugin to incorporate the full range of Amazon Polly voices into Windows SAPI-compliant applications.</td>
<td>September 26, 2019</td>
</tr>
<tr>
<td>Major new feature (p. 284)</td>
<td>In addition to the standard text-to-speech (TTS) voices supported by Amazon Polly since its launch, Amazon Polly now provides an improved Neural TTS (NTTS) system that can provide even higher quality voices, thereby providing you with the most natural and human-like text-to-speech voices possible. For more information, see Neural Text-to-Speech.</td>
<td>July 30, 2019</td>
</tr>
<tr>
<td>New voices added (p. 284)</td>
<td>New voices added: Lucia (female, Spanish), and Bianca (female, Italian).</td>
<td>August 2, 2018</td>
</tr>
<tr>
<td>New language added (p. 284)</td>
<td>New language added: Mexican Spanish (es-MX). This language uses the female voice of Mia.</td>
<td>August 2, 2018</td>
</tr>
<tr>
<td>New language added (p. 284)</td>
<td>New language added: Hindi (hi-IN). This voice uses the female voice of Aditi, which is also used for Indian English, making Aditi Amazon Polly’s first bilingual voice.</td>
<td>August 2, 2018</td>
</tr>
<tr>
<td>New feature added (p. 284)</td>
<td>Addition of Speech synthesis of long text passages (up to 100,000 billed characters).</td>
<td>July 17, 2018</td>
</tr>
<tr>
<td>Event Description</td>
<td>Description</td>
<td>Date</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>New SSML feature added (p. 284)</td>
<td>Addition of Maximum Duration for Synthesized Speech.</td>
<td>July 17, 2018</td>
</tr>
<tr>
<td>New voice added (p. 284)</td>
<td>New voice added: Léa (female, French).</td>
<td>June 5, 2018</td>
</tr>
<tr>
<td>Region expansion (p. 284)</td>
<td>Expansion of Amazon Polly service to all commercial regions.</td>
<td>June 4, 2018</td>
</tr>
<tr>
<td>Expanded feature (p. 284)</td>
<td>Expansion of Amazon Polly WordPress Plugin feature, including addition of Amazon Translate capabilities.</td>
<td>June 4, 2018</td>
</tr>
<tr>
<td>New voices added (p. 284)</td>
<td>Two new voices added: Aditi (female, Indian English) and Seoyeon (female, Korean).</td>
<td>November 15, 2017</td>
</tr>
<tr>
<td>New feature (p. 284)</td>
<td>Addition of new Speech Marks feature, as well as an expansion of SSML capabilities.</td>
<td>April 19, 2017</td>
</tr>
<tr>
<td>New guide (p. 284)</td>
<td>This is the first release of the Amazon Polly Developer Guide.</td>
<td>November 30, 2016</td>
</tr>
</tbody>
</table>
AWS glossary

For the latest AWS terminology, see the AWS glossary in the AWS General Reference.