

Developer Guide

Amazon Polly



Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon Polly: Developer Guide

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

How it works	What Is Amazon Polly?	1
Are you a first-time user? 3 Getting started 4 Signing up for AWS 4 Sign up for an AWS account 4 Create a user with administrative access 5 Setting up the AWS CLI 7 Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Billingual voices 21 Billingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	How it works	1
Getting started 4 Signing up for AWS 4 Sign up for an AWS account 4 Create a user with administrative access 5 Setting up the AWS CLI 7 Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (arb) 31 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 42 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Benefits	2
Signing up for AWS 4 Sign up for an AWS account 4 Create a user with administrative access 5 Setting up the AWS CLI 7 Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changuages in Amazon Polly 26 Languages in Amazon Polly 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Are you a first-time user?	3
Sign up for an AWS account 4 Create a user with administrative access 5 Setting up the AWS CLI 7 Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Getting started	4
Create a user with administrative access 5 Setting up the AWS CLI 7 Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Signing up for AWS	4
Setting up the AWS CLI 7 Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Bilingual voices 21 Bilingual voices 21 Accented bilingual voices 22 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Sign up for an AWS account	4
Reconfiguring the AWS CLI 8 Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 21 Brand voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Create a user with administrative access	5
Synthesizing speech example 9 Voices in Amazon Polly 13 Available voices 13 Brand voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Setting up the AWS CLI	7
Voices in Amazon Polly 13 Available voices 13 Brand voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Reconfiguring the AWS CLI	8
Available voices	Synthesizing speech example	9
Brand voices 21 Bilingual voices 21 Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Voices in Amazon Polly	. 13
Bilingual voices 21 Accented bilingual voices 22 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Available voices	13
Accented bilingual voices 21 Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Brand voices	. 21
Fully bilingual voices 22 Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Bilingual voices	. 21
Applying the newscaster voice 23 Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Accented bilingual voices	. 21
Listening to voices 25 Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Fully bilingual voices	22
Timing a voice speed 25 Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Applying the newscaster voice	23
Changing a voice speed 26 Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Listening to voices	. 25
Languages in Amazon Polly 28 Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Timing a voice speed	25
Arabic (arb) 31 Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Changing a voice speed	26
Arabic (Gulf) (ar-AE) 36 Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Languages in Amazon Polly	28
Catalan (ca-ES) 42 Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Arabic (arb)	. 31
Chinese (Cantonese) (yue-CN) 46 Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Arabic (Gulf) (ar-AE)	. 36
Chinese (Mandarin) (cmn-CN) 50 Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Catalan (ca-ES)	42
Czech (cs-CZ) 55 Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Chinese (Cantonese) (yue-CN)	. 46
Danish (da-DK) 59 Dutch (Belgian) (nl-BE) 63 Dutch (nl-NL) 67	Chinese (Mandarin) (cmn-CN)	50
Dutch (Belgian) (nl-BE)	Czech (cs-CZ)	. 55
Dutch (nl-NL)	Danish (da-DK)	. 59
	Dutch (Belgian) (nl-BE)	63
English (US) (en-US)	Dutch (nl-NL)	67
	English (US) (en-US)	. 71
English (Australian) (en-AU)	English (Australian) (en-AU)	. 74
English (British) (en-GB)	English (British) (en-GB)	78

	English (Indian) (en-IN)	83
	English (Ireland) (en-IE)	87
	English (New Zealand) (en-NZ)	. 90
	English (South African) (en-ZA)	96
	English (Welsh) (en-GB-WLS)	101
	Finnish (fi-FI)	105
	French (fr-FR)	110
	French (Belgian) (fr-BE)	113
	French (Canadian) (fr-CA)	117
	German (de-DE)	120
	German (Austrian) (de-AT)	124
	German (Swiss standard) (de-CH)	129
	Hindi (hi-IN)	133
	Icelandic (is-IS)	137
	Italian (it-IT)	141
	Japanese (ja-JP)	144
	Korean (ko-KR)	148
	Norwegian (nb-NO)	151
	Polish (pl-PL)	155
	Portuguese (pt-PT)	158
	Portuguese (Brazilian) (pt-BR)	161
	Romanian (ro-RO)	164
	Russian (ru-RU)	167
	Spanish (es-ES)	171
	Spanish (Mexican) (es-MX)	174
	Spanish (US) (es-US)	177
	Swedish (sv-SE)	180
	Turkish (tr-TR)	184
	Welsh (cy-GB)	188
V	pice engines	193
	Generative engine	193
	Available generative voices	194
	Feature and region compatibility	195
	Long-form engine	196
	Available long-form voices	197
	Feature and region compatibility	198

	Neural engine	198
	Available neural voices	199
	Feature and region compatibility	203
	Standard engine	204
	Available Standard voices	204
	Feature and region compatibility	208
	Choosing a voice engine	209
Sp	eech marks	210
	Speech mark types	210
	Visemes and Amazon Polly	211
	Speech mark output	212
	Requesting speech marks	213
	Speech marks without SSML example	215
	Speech marks with SSML example	216
Us	ing SSML	218
	Reserved characters	219
	Using SSML on the console	221
	Using SSML with the Synthesize-Speech command	223
	Synthesizing an SSML-enhanced document	224
	Supported SSML tags	225
	Identifying SSML-enhanced text	227
	Adding a pause	228
	Emphasizing words	229
	Specifying another language for specific words	230
	Placing a custom tag in your text	231
	Adding a pause between paragraphs	231
	Using phonetic pronunciation	232
	Controlling volume, speaking rate, and pitch	234
	Setting a maximum duration for synthesized speech	236
	Adding a pause between sentences	240
	Controlling how special types of words are spoken	240
	Pronouncing acronyms and abbreviations	244
	Improving pronunciation by specifying parts of speech	244
	Adding the sound of breathing	246
	Newscaster speaking style	250
	Adding dynamic range compression	251

	Speaking softly	253
	Controlling timbre	253
	Whispering	255
M	anaging lexicons	257
	Using multiple lexicons	258
	Uploading a lexicon	259
	Applying lexicons (Synthesizing Speech)	265
	Filtering the lexicon list on the console	268
	Downloading lexicons on the console	270
	Deleting a lexicon	271
Lo	ong audio files	273
	Setting up the IAM policy for asynchronous synthesis	274
	Creating long audio files	275
Co	ode and application examples	279
	Java samples	279
	DeleteLexicon	280
	DescribeVoices	280
	GetLexicon	282
	ListLexicons	282
	PutLexicon	283
	StartSpeechSynthesisTask	284
	Speech Marks	286
	SynthesizeSpeech	287
	Python samples	289
	DeleteLexicon	289
	GetLexicon	290
	ListLexicon	291
	PutLexicon	292
	StartSpeechSynthesisTask	
	SynthesizeSpeech	294
	Java example	295
	Python example	
	Python example: index.html	301
	Python example: server.py	306
	iOS example	312
	Android example	314

Ų١	lotas	31 /
	Supported regions	318
	Quotas and throttle rates	318
	Concurrent requests	319
	Best practices to mitigate throttling	319
	Pronunciation lexicons	320
	SynthesizeSpeech API operations	320
	SpeechSynthesisTask API operations	
	Speech Synthesis Markup Language (SSML)	
Se	curity	322
	Data Protection	323
	Encryption at Rest	323
	Encryption in Transit	324
	Internetwork Traffic Privacy	324
	Identity and Access Management	324
	Audience	324
	Authenticating with identities	
	Managing access using policies	
	How Amazon Polly works with IAM	331
	Identity-based policy examples	338
	Amazon Polly API Permissions Reference	345
	Troubleshooting	347
	Logging and Monitoring	348
	Compliance Validation	349
	Resilience	350
	Infrastructure Security	350
	Security Best Practices	350
	Using Interface VPC Endpoints	351
	Availability	
	Creating a VPC endpoint for Amazon Polly	351
	Testing the connection between your VPC and Amazon Polly	351
	Controlling access to your Amazon Polly endpoint	352
	Support for VPC context keys	353
Lo	gging Amazon Polly API calls with AWS CloudTrail	354
	Amazon Polly information in CloudTrail	354
	Example: Amazon Polly Log File Entries	355

CloudWatch integration	357
Getting CloudWatch Metrics (Console)	357
Getting CloudWatch metrics on the AWS CLI	357
Amazon Polly Metrics	358
Dimensions for Amazon Polly Metrics	359
API Reference	361
Actions	361
DeleteLexicon	362
DescribeVoices	364
GetLexicon	368
GetSpeechSynthesisTask	371
ListLexicons	374
ListSpeechSynthesisTasks	377
PutLexicon	380
StartSpeechSynthesisTask	383
SynthesizeSpeech	391
Data Types	397
Lexicon	398
LexiconAttributes	399
LexiconDescription	401
SynthesisTask	402
Voice	407
Common Errors	409
Common Parameters	411
Document History	414

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. With Amazon Polly, you can build speechenabled applications that work in multiple locations and use the ideal voice for your customers. Also, you only pay for the text you synthesize. You can also cache and replay Amazon Polly's generated speech at no additional cost.

Amazon Polly offers many voice options, including generative, long-form, neural, and standard text-to-speech (TTS) options. These voices deliver ground-breaking improvements in speech quality using new machine learning technology to offer the most natural and human-like text-to-speech voices possible. Neural TTS technology also supports a Newscaster speaking style, 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).

How Amazon Polly works

Amazon Polly converts input text into life-like speech. To use an Amazon Polly voice, choose a voice engine, call a speech synthesis method, provide the text that you want to synthesize, then 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 plaintext 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.
- 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

How it works 1

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. 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.



Note

To hear example Amazon Polly voices in your browser, see the Amazon Polly product overview.

Benefits

Some of the benefits of using Amazon Polly include:

- High quality Amazon Polly offers highly-performant generative, long-form, neural, and highquality text-to-speech (TTS) voices. These technologies synthesize natural speech with high pronunciation accuracy (including abbreviations, acronym expansions, date/time interpretations, and homograph disambiguation).
- Low latency Amazon Polly achieves fast responses, which makes it a viable option for lowlatency use cases such as dialogue systems.
- Support for a large portfolio of languages and voices Amazon Polly supports dozens of voices and languages, offering male and female voice options for most languages. 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 that there are no setup costs. 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, smartphones, and so on. In contrast,

Benefits 2

TTS conversion done in the AWS Cloud dramatically reduces local resource requirements. This enables support of all the available languages and voices with outstanding quality. Moreover, speech improvements are instantly available to all end users and don't require additional updates for devices.



Note

To hear example Amazon Polly voices in your browser, see the Amazon Polly product overview.

Are you a first-time user?

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

- 1. How Amazon Polly works This section introduces various Amazon Polly inputs and options that you can work with in order to create a simple experience.
- 2. Getting started with Amazon Polly In this section, you set up your account and test Amazon Polly speech synthesis.
- 3. Code and application examples This section provides additional examples that you can use to explore Amazon Polly.

Are you a first-time user?

Getting started with Amazon Polly

Amazon Polly provides several API operations that you can easily integrate with your existing applications. For a list of supported operations, see Actions.

You can perform almost all of the same operations on the Amazon Polly console and the AWS CLI. However, you can't listen to synthesized speech on the AWS CLI. To work with audio on the AWS CLI, save your text to a file. Then open the file in an audio application of your choice.

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 use Amazon Polly without writing any code.

Before you use Amazon Polly for the first time, you must 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're charged only for the services and resources that you use. If you're a new AWS customer, you can get started with Amazon Polly with no charge. For more information, see AWS Free Usage Tier.

The following sections describe how to get started using Amazon Polly.

Topics

- Signing up for AWS
- Setting up the AWS CLI
- Reconfiguring the AWS CLI

Signing up for AWS

Before you can use any AWS service, including Amazon Polly, you must sign up for AWS.

Sign up for an AWS account

If you do not have an AWS account, complete the following steps to create one.

Signing up for AWS

To sign up for an AWS account

- 1. Open https://portal.aws.amazon.com/billing/signup.
- 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.

When you sign up for an AWS account, an AWS account root user is created. The root user has access to all AWS services and resources in the account. As a security best practice, assign administrative access to a user, and use only the root user to perform tasks that require root user access.

AWS sends you a confirmation email after the sign-up process is complete. At any time, you can view your current account activity and manage your account by going to https://aws.amazon.com/ and choosing **My Account**.

Create a user with administrative access

After you sign up for an AWS account, secure your AWS account root user, enable AWS IAM Identity Center, and create an administrative user so that you don't use the root user for everyday tasks.

Secure your AWS account root user

- 1. Sign in to the <u>AWS Management Console</u> as the account owner by choosing **Root user** and entering your AWS account email address. On the next page, enter your password.
 - For help signing in by using root user, see <u>Signing in as the root user</u> in the *AWS Sign-In User Guide*.
- 2. Turn on multi-factor authentication (MFA) for your root user.

For instructions, see <u>Enable a virtual MFA device for your AWS account root user (console)</u> in the *IAM User Guide*.

Create a user with administrative access

Enable IAM Identity Center.

For instructions, see <u>Enabling AWS IAM Identity Center</u> in the *AWS IAM Identity Center User Guide*.

2. In IAM Identity Center, grant administrative access to a user.

For a tutorial about using the IAM Identity Center directory as your identity source, see Configure user access with the default IAM Identity Center directory in the AWS IAM Identity Center User Guide.

Sign in as the user with administrative access

 To sign in with your IAM Identity Center user, use the sign-in URL that was sent to your email address when you created the IAM Identity Center user.

For help signing in using an IAM Identity Center user, see <u>Signing in to the AWS access portal</u> in the *AWS Sign-In User Guide*.

Assign access to additional users

 In IAM Identity Center, create a permission set that follows the best practice of applying leastprivilege permissions.

For instructions, see Create a permission set in the AWS IAM Identity Center User Guide.

2. Assign users to a group, and then assign single sign-on access to the group.

For instructions, see Add groups in the AWS IAM Identity Center User Guide.

For more information about IAM, see the following:

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



Note your AWS account ID. You will need it in the next steps.

Setting up the AWS CLI

Follow these steps to download and configure the AWS CLI to work with Amazon Polly.

To set up the AWS Command Line Interface

- Download and configure the AWS CLI. For instructions, see the following topics in the AWS 1. Command Line Interface User Guide:
 - Getting Set Up with the AWS Command Line Interface
 - Configuring the AWS Command Line Interface
- Add a named profile for the administrator user in the AWS CLI AWS Config file. You can 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.



If you're using a 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
```

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.

Setting up the AWS CLI

Reconfiguring the AWS CLI

If you've previously downloaded and configured the AWS CLI, Amazon Polly might be unavailable unless you reconfigure the AWS CLI. The following procedure checks to see if this is necessary.

To reactivate Amazon Polly from the AWS CLI

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

```
aws polly help
```

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

- 2. Activate Amazon Polly using one of the two following options:
 - a. Uninstall and reinstall the AWS CLI.

For instructions, see <u>Installing the AWS Command Line Interface</u> 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.

Reconfiguring the AWS CLI 8

Synthesizing speech with Amazon Polly example

This page presents a short speech synthesis example performed in the console, the AWS CLI, and with Python. This example performs speech synthesis from plain text, not SSML.

Console

Synthesize speech on the 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. The text field will load with example text so you can quickly try out Amazon Polly.
- 3. Turn off **SSML**.
- 4. Type or paste this text into the input box.

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

- 5. Under Engine, choose Generative, Long Form, Neural, or Standard.
- 6. Choose a language and AWS Region, then choose a voice. (If you select **Neural** for **Engine**, only the languages and voices that support NTTS are available. All Standard and Long Form voices are disabled.)
- 7. To listen to the speech immediately, choose **Listen**.
- 8. 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**.

AWS CLI

In this exercise, you call the SynthesizeSpeech operation by passing input text. You can save the resulting audio as a file and verify its content.

 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 to be synthesized by a voice of your choice. You must provide a voice ID (explained in the following step) and an 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.

Python

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

The Python code in this example performs the following actions:

- Invokes the AWS SDK for Python (Boto) to send a SynthesizeSpeech request to Amazon Polly (by providing some 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.

```
"""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
    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 more in-depth examples, see the following topics:

- Using SSML on the console
- Applying lexicons (Synthesizing Speech)
- Code and application examples

Voices in Amazon Polly

Amazon Polly provides dozens of lifelike voices and support for a variety of languages. Each voice is created using native language speakers, so there are variations from voice to voice, even within the same language. You can also use the AWS Management Console to test each voice with text of your choice. For most languages, there will be at least one male and one female voice, and often more than one of each. A few languages only have a single voice.

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, provide feedback on this page. Unfortunately, we are not able to comment on plans for specific new languages before they are released.



Note

To hear example Amazon Polly voices in your browser, see the Amazon Polly product overview.

Topics

- Available voices
- Bilingual voices
- Applying the newscaster voice
- Listening to voices
- Timing a voice speed
- Changing a voice speed

Available voices

Amazon Polly provides a variety of lifelike voices in multiple languages for synthesizing speech from text. The following table shows all the voices that Amazon Polly offers.

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
1	Arabic	arb	Zeina	Female	No	No	No	Yes
2	Arabic	ar-AE	Hala*	Female	No	No	Yes	No
	(Gulf)		Zayd*	Male	No	No	Yes	No
3	Dutch (Belgian)	nl-BE	Lisa	Female	No	No	Yes	No
4	Catalan	ca-ES	Arlet	Female	No	No	Yes	No
5	Czech	cs-CZ	Jitka	Female	No	No	Yes	No
6	Chinese (Cantonese)	yue-CN	Hiujin	Female	No	No	Yes	No
7	Chinese (Mandarii)	cmn-CN	Zhiyu	Female	No	No	Yes	Yes
8	Danish	da-DK	Naja	Female	No	No	No	Yes
			Mads	Male	No	No	No	Yes
			Sofie	Female	No	No	Yes	No
9	Dutch	nl-NL	Laura	Female	No	No	Yes	No
			Lotte	Female	No	No	No	Yes
			Ruben	Male	No	No	No	Yes
10	English (Australi an)	en-AU	Nicole Olivia	Female Female	No Yes	No No	No Yes	Yes No

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
			Russell	Male	No	No	No	Yes
11	English	en-GB	Amy**	Female	Yes	No	Yes	Yes
	(British)		Emma	Female	No	No	Yes	Yes
			Brian	Male	No	No	Yes	Yes
			Arthur	Male	No	No	Yes	No
12	English (Indian)	en-IN	Aditi*	Female	No	No	No	Yes
			Raveena	Female	No	No	No	Yes
			Kajal*	Female	Yes	No	Yes	No
13	English (Ireland)	en-IE	Niamh	Female	No	No	Yes	No
14	English (New Zealand)	en-NZ	Aria	Female	No	No	Yes	No
15	English (South African)	en-ZA	Ayanda	Female	Yes	No	Yes	No

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
16	English	en-US	Danielle	Female	Yes	Yes	Yes	No
	(US)		Gregory	Male	No	Yes	Yes	No
			lvy	Female (c	No	No	Yes	Yes
			Joanna**	Female	Yes	No	Yes	Yes
			Kendra	Female	No	No	Yes	Yes
			Kimberly	Female	No	No	Yes	Yes
			Salli	Female	No	No	Yes	Yes
			Joey	Male	No	No	Yes	Yes
			Justin	Male	No	No	Yes	No
			Kevin	(child)	No	No	Yes	Yes
			Matthew*	Male (child)	Yes	No	Yes	No
			Ruth	Male	Yes	Yes	Yes	No
			Stephen	Female	Yes	No	Yes	No
			Patrick	Male	No	Yes	No	No
				Male				
17	English (Welsh)	en-GB- WLS	Geraint	Male	No	No	No	Yes
18	Finnish	fi-FI	Suvi	Female	No	No	Yes	No

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
19	French	fr-FR	Céline/ Celine	Female	No	No	No	Yes
			Léa	Female	Yes	No	Yes	Yes
			Mathieu	Male	No	No	No	Yes
			Rémi	Male	Yes	No	Yes	No
20	French (Belgian)	fr-BE	Isabelle	Female	No	No	Yes	No
21	French	fr-CA	Chantal	Female	No	No	No	Yes
	(Canadiar)		Gabrielle	Female	No	No	Yes	No
			Liam	Male	No	No	Yes	No
22	German	de-DE	Marlene	Female	No	No	No	Yes
			Vicki	Female	Yes	No	Yes	Yes
			Hans	Male	No	No	No	Yes
			Daniel	Male	Yes	No	Yes	No
23	German (Austrian)	de-AT	Hannah	Female	No	No	Yes	No
24	German (Swiss)	de-CH	Sabrina	Female	No	No	Yes	No
25	Hindi	hi-IN	Aditi*	Female	No	No	No	Yes
			Kajal*	Female	No	No	Yes	No

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
26	Icelandic	is-IS	Dóra/ Dora	Female Male	No No	No No	No No	Yes Yes
			Karl	Mate	140	140	140	103
27	Italian	it-IT	Carla	Female	No	No	No	Yes
			Bianca	Female	Yes	No	Yes	Yes
			Giorgio	Male	No	No	No	Yes
			Adriano	Male	No	No	Yes	No
28	Japanese	ja-JP	Mizuki	Female	No	No	No	Yes
			Takumi	Male	No	No	Yes	Yes
			Kazuha	Female	No	No	Yes	No
			Tomoko	Female	No	No	Yes	No
29	Korean	ko-KR	Seoyeon	Female	No	No	Yes	Yes
30	Norwegia	nb-NO	Liv	Female	No	No	No	Yes
			Ida	Female	No	No	Yes	No
31	Polish	pl-PL	Ewa	Female	No	No	No	Yes
			Maja	Female	No	No	No	Yes
			Jacek	Male	No	No	No	Yes
			Jan	Male	No	No	No	Yes
			Ola	Female	No	No	Yes	No

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
32	Portugue	pt-BR	Camila	Female	No	No	Yes	Yes
	e (Brazilia		Vitória/	Female	No	No	Yes	Yes
	n)		Vitoria	Male	No	No	No	Yes
			Ricardo	Male	No	No	Yes	No
			Thiago					
33	Portugue e	pt-PT	Inês/ Ines	Female	No	No	Yes	Yes
	(Europea		Cristiano	Male	No	No	No	Yes
34	Romania	ro-RO	Carmen	Female	No	No	No	Yes
35	Russian	ru-RU	Tatyana	Female	No	No	No	Yes
			Maxim	Male	No	No	No	Yes
36	Spanish	es-ES	Conchita	Female	No	No	No	Yes
	(Spain)		Lucia	Female	Yes	No	Yes	Yes
			Alba	Female	No	Yes	No	No
			Enrique	Male	No	No	No	Yes
			Sergio	Male	Yes	No	Yes	No
			Raúl	Male	No	Yes	No	No
37	Spanish	es-MX	Mia	Female	Yes	No	Yes	Yes
	(Mexican)		Andrés	Male	Yes	No	Yes	No

	Language and language variants	Language code	Name/ ID	Gender	Generativ e voice	Long Form voice	Neural voice	Standard voice
38	Spanish (US)	es-US	Lupe** Penélope, Penelope Miguel Pedro	Female Female Male Male	Yes No No Yes	No No No	Yes No No Yes	Yes Yes Yes No
39 40	Swedish Turkish	sv-SE tr-TR	Astrid Elin Filiz	Female Female	No No	No No	No Yes No	Yes No Yes
41	Welsh	cy-GB	Burcu Gwyneth	Female Female	No No	No No	Yes No	No Yes

^{*} This voice is bilingual. For more information, see Bilingual voices.

Each Amazon Polly voice engine has unique features. Learn more about features and Region availability for the voice engines offered by Amazon Polly:

- Generative voices
- Long-form voices
- Neural voices
- Standard voices

^{**} These voices can be used with Newscaster speaking styles when used with the Neural format. For more information, see Applying the newscaster voice.

Brand voices

In addition to the available voices listed in the previous table, you can use Amazon Polly to build a custom voice for your brand persona. With a brand voice, you can offer unique and exclusive voices to your customers. To learn more about Amazon Polly brand voices, see Brand Voice.

Bilingual voices

Amazon Polly has two ways of producing bilingual voices:

- Accented bilingual voices
- Fully bilingual voices

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:

```
<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:

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

Brand voices 21

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

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

Fully bilingual voices

A fully bilingual voice like Aditi or Kajal (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, Kajal, Hala, and Zayd are the only fully bilingual voices available.

Using a Bilingual Voice (example: 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 कहाँ, खेल तोह अब शुर 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.

Fully bilingual voices 22

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.

Applying the newscaster voice

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.

Console

To apply the Newscaster style

- 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.

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

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

```
<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**.

AWS CLI

To apply the Newscaster style

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.

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

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

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

Applying the newscaster voice 24

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>
```

For more information about SSML, see Supported SSML tags.

Listening to voices

Once you have set up Amazon Polly, you can test voices using custom text on the console.

To listen to Amazon Polly voices on the 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. For **Engine**, choose **Generative**, **Long Form**, **Neural**, or **Standard**.
- 4. Select 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**.

Timing a voice speed

Because of the natural variation between voices, each available voice speaks at slightly different speeds. For instance, with US English voices, Ivy and Joanna are slightly faster than Matthew, and considerably faster than Joey. Since there is so much variation between voices, there is no standard speed (words per minute) available for Amazon Polly voices. However, you can find how long it takes for your voice to say the selected text using Speech Marks.

Listening to voices 25

</speak>

To time the length of a spoken text passage

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

```
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
```

Open LengthOfText.txt.

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

```
{"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 a 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, if your organization was making an application that reads books to immigrant audiences, you may want to vary the voice speed. Your audience may speak English, but their fluency is limited. Amazon Polly helps you slow down the rate of speech using the SSML prosody> tag.

You can use a percentage:

Changing a voice speed 26

```
</speak>
```

Or a preset speed:

```
<speak>
    In some cases, it might help your audience to rosody 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.

Note

Test your selected voice at various speeds. The speed of each option is approximate and depends on the voice you choose.

For more information on using the prosody tag, see Controlling volume, speaking rate, and pitch.

Changing a voice speed 27

Languages in Amazon Polly

The following languages are supported by Amazon Polly and can be used to synthesize speech. Each language has a unique language code. These language codes are <u>W3C language identification</u> tags (*ISO* 639-3 for the language name and *ISO* 3166 for the country code).

Select a language from the following table for details on the phonemes and visemes that Amazon Polly provides.

Langua	Langu code
<u>Arabic</u>	arb
Arabic (Gulf)	
Catalar	ca- ES
Chineso (Canton e)	
Chineso (Manda)	
Czech	cs- CZ
Danish	da- DK
Dutch (Belgia	
Dutch	nl- NL

Langua	Langu code
English (Austra an)	
English (British	
English (Indian	
English (New Zealan	en- NZ
English (South African	ZA
English (US)	en- US
English (Welsh	
Finnish	
French	
French (Belgia	fr-BE
French (Canad)	fr-CA

Langua Langua code hi-IN Hindi Germa de-DE Germa de-(Austria AT Germa de-(Swiss CH standa Iceland is-IS Italian it-IT Japane ja-JP Korean ko-KR Norwe nb-NO Polish pl-PL Portug pt- BR e (Brazili <u>n)</u> Portug pt-PT e (Europ

Langua	Langu code
Roman	ro- RO
Russiar	ru- RU
Spanisl (Spain)	
Spanisl (Mexica	
Spanisl (US)	
Swedis	sv- SE
Turkish	tr-TR
Welsh	cy- GB

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.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			

IPA	X-SAMPA	Description	Example	Viseme
7	?	glottal stop	أنا	
٢	?\	voiced pharyngeal fricative	ڠُمَر	k
b	b	voiced bilabial plosive	بَلَد	p
d	d	voiced alveolar plosive	داري	t
q _l	d_?\	emphatic voiced alveolar plosive	ضَوء	t
dz	dZ	voiced postalveo lar affricate	جَميل	S
ð	D	voiced dental fricative	ذٰٰلِڬۘ	Т
ðſ	D_?\	emphatic voiced dental fricative	ظَلام	Т
f	f	voiceless labiodent al fricative	فَصل	f
g	g	voiced velar plosive	إنجلترا	k
¥	G	voiced velar fricative	غَرب	k
h	h	voiceless glottal fricative	ەذا	k
j	j	palatal approxima nt	يَ م شي	i

IPA	X-SAMPA	Description	Example	Viseme
k	k	voiceless velar plosive	كَلب	k
l	ι	alveolar lateral approximant	لاقى	t
ļγ	l_G	emphatic alveolar lateral approxima nt	عبدالله	t
m	m	bilabial nasal	ماذا	р
n	n	alveolar nasal	نور	t
p	p	voiceless bilabial plosive	حَبس	р
q	q	voiceless uvular plosive	قَريب	k
r	r	alveolar trill	رَمِل	r
S	S	voiceless alveolar fricative	س ُوْال	S
sr	s_?\	emphatic voiceless alveolar fricative	صاحِب	S
ſ	S	voiceless postalveo شُكر lar fricative		S
t	t	voiceless alveolar plosive	تَمر	t
ţ٢	t_?\	emphatic voiceless alveolar plosive	طالِب	t

IPA	X-SAMPA	Description	Example	Viseme
θ	Т	voiceless dental fricative	ثَلاث	Т
V	V	voiced labiodental fricative	فيتامين	f
W	W	labio-velar approximant	وَلَد	u
X	X	voiceless velar fricative	خَوْف	k
ħ	X\	voiceless pharyngeal fricative	حَوْلَ	k
Z	z	voiced alveolar fricative	زُەور	S
Vowels				
а	a	open front unrounded vowel	بَرد	a
a:	a:	long open front unrounded vowel	دار	a
a_{ℓ}	A_?\	emphatic open Joback unrounded vowel		a
a _c :	A_?\:	emphatic long open back unrounded vowel	ظالِم	a
u	u	close back rounded vowel	شُرب	u

IPA	X-SAMPA	Description	Example	Viseme
u:	u:	long close back rounded vowel		u
u ^ς	u_?\	emphatic close back rounded vowel		u
u ^r :	u_?\:	emphatic long close back rounded vowel	طول	u
i	i	close front unrounded vowel	بِنت	i
i:	i:	long close front unrounded vowel	حَزين	i
j?	i_?\	emphatic close front unrounded vowel	ۻۣڎۜ	i
it:	i_?\:	emphatic long close front unrounded vowel	ماضي	i
е	е	close-mid front unrounded vowel	ماركت	e
e:	e:			e
Э	0	open-mid back rounded vowel	تكنولوجي	0

IPA	X-SAMPA	Description	Example	Viseme
o:	O:	long open-mid back rounded vowel	تليفزيون	0

Arabic (Gulf) (ar-AE)

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 Hala that is supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme	
Consonants						
b	b	voiced bilabial plosive	بلد	/"ba.lad/	b	
d	d	voiced alveolar plosive	رد	/ " r a d d /	d	
q,	d_?\	pharyngea lised voiced alveolar plosive	ضوء	/ " d_?\ a w ? /	D	
f	f	voiceless labiodental fricative	فرن	/"fl.rln/	f	
g	g	voiced velar plosive	قال	/ " g a: l /	k	
j	j	voiced palatal approximant	يمشي	/ " j I m . S i: /	i	

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme
k	k	voiceless velar plosive	كامل	/ " k a: . m i l /	k
l	l	voiced alveolar lateral approximant	ليل	/"le:l/	t
lı	I_G	pharyngea lised voiced alveolar lateral approximant	عبدالله	/ ?\ a b . " d A_?\ l_G . l_G A_?\ /	t
m	m	bilabial nasal stop	مئة	/ " m l j . j a /	p
n	n	alveolar nasal stop	ٺور	/ " n u: r /	t
p	p	voiceless bilabial plosive	أوبرا	/ " ? O . p e . r a: /	p
q	q	voiceless uvular plosive	قصر	/ " q A_?\ s_?\ r /	k
r	r	alveolar trill	رمل	/"ra.mIl/	r
S	S	voiceless alveolar fricative	سمسم	/ " s l m . s l m /	S
sr	s_?\	pharyngea lised voiceless alveolar fricative	صاحب	/ " s_?\ A_?: . X \ I b /	S

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme
t	t	voiceless alveolar plosive	تمر	/ "t a . m a r /	t
ţΥ	t_?\	pharyngea lised voiceless alveolar fricative	طالب	/ " t_?\ A_?: . l I b /	t
V	V	voiced labiodental fricative	فيتامين	/ v i: . t A . " m i: n /	f
W	W	voiced labiovelar approximant	وايد	/ " w a: . j l d /	u
x	х	voiceless velar fricative	خروف	/xa."ru:f/	k
Z	Z	voiceless velar fricative	زەور	/ " z h u: r /	S
ð	D	voiced interdental fricative	ذلك	/ " D a: . l I k /	D
ðΓ	D_?\	pharyngea lised voiced interdental fricative	ظلام	/ D_?\ A_?\ . " l a: m /	D
ħ	X\	voiceless pharyngeal fricative	الجين	/?al."X\i: n/	k

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme
ŋ	N	velar nasal stop	ەونغ كونغ	/ h O N . " k O N g /	k
γ	G	voiced velar fricative	غريبة	/ G I . " r i: . b a /	k
l	S	voiceless postalveolar fricative	شمس	/ " S a m s /	S
3	Z	voiced postalveolar fricative	جاكيت	/Za."ke:t/	S
7	?	glottal stop	مۇسسة	/ m u . " ? a s . s a . s a /	
٢	?\	voiced pharyngeal fricative	عام	/ " ?\ a: m m /	k
dЗ	dZ	voiced postalveolar affricate	جامعة	/ " dZ a: m . ?\ a /	S
θ	Т	voiced interdental fricative	ثلاثة	/ T a . " l a: . T a /	Т
ĥ	h	voiced glottal fricative	ەلال	/"hla:l/	k
Vowels					

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme
æ	a	mid-open front unrounded short vowel	سفر	/"sa.far/	a
a ^r	A_?\	pharyngea lised open back unrounded short vowel	صلب	/ " s_?\ A_?\ l b /	a
æ:	a:	mid-open front unrounded long vowel	باب	/ " b a: b /	a
a ^r :	A_?\:	pharyngea lised open back unrounded long vowel	ناضج	/ " n A_?: . D_? \ i_?\ dZ /	a
a	A	open central unrounded short vowel	wifi	/"wAj.fAj/	a
i	i	tense close front unrounded short vowel (MSA)	إسحاق	/ ? i s . " X\ A_? \: q /	İ
I	I	lax close front unrounded short vowel	بنت	/"bInt/	i

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme
71	i_?\	pharyngea lised close front unrounded short vowel	طفل	/ " t_?\ i_?\ f I l /	i
i:	i:	close front unrounded long vowel	سبيل	/ s a . " b i: l /	i
it:	i_?:	pharyngea lised close front unrounded long vowel	رطيب	/ r A_?\ . " t_?\ i_?: b /	i
u	u	tense close back rounded short vowel (MSA)	مخترع	/"mux.ta. ri?\/	u
ប	U	lax close back rounded short vowel	رسوم	/ r U . " s u: m /	u
u ^r	u_?\	pharyngea lised close back rounded short vowel	عصفور	/ ?\ u_?\ s_?\ . " f u: r /	u
u:	u:	close back rounded long vowel	توت	/ " t u: t /	u

IPA	X-SAMPA	Description	Example	Pronunciation	Viseme
ս ^ւ :	u_?\:	pharyngea lised close back rounded long vowel	صور	/ " s_?\ u_?\: r /	u
е	е	mid front unrounded short vowel	ٳڹۨؾۘڔ۠ڹۣؾ	/ " s e n t /	е
e:	e:	mid front unrounded long vowel	إيش	/ " ? e: S /	е
Э	0	open-mid back rounded short vowel	دولار	/dO."lAr/	0
o:	O:	open-mid back rounded long vowel	لون	/ " l O: n /	0

Catalan (ca-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 Catalan voice of Arlet that is supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme			
Consonants							
р	р	voiceless bilabial plosive	p loure	p			

IPA	X-SAMPA	Description	Example	Viseme
t	t	voiceless alveolar plosive	T arragona	t
k	k	voiceless velar plosive	com	k
b	b	voiced bilabial plosive	b ata	p
d	d	voiced alveolar plosive	en d oll	t
g	g	voiced velar plosive	g ros	k
m	m	voiced bilabial nasal	m anera	p
n	n	voiced alveolar nasal	do n ar	t
Л	J	voiced palatal nasal	a ny	J
ŋ	N	voiced velar nasal	pi n güí	k
†	5	voiced velarized alveolar lateral approximant (dark l)	albercoc	ί
λ	L	voiced palatal lateral approxima nt	llop	J
r	r	voiced alveolar trill	pa rr a	r
ſ	4	voiced alveolar tap	para	t

IPA	X-SAMPA	Description	Example	Viseme
f	f	voiceless labiodent al fricative	èm f asi	f
S	S	voiceless alveolar fricative	s ac	S
Z	Z	voiced alveolar fricative	calzes	S
ſ	S	voiceless postalveo lar fricative	gui x	S
3	Z	voiced postalveo lar fricative	col·le g i	S
<u></u> tſ	tS	voiceless postalveo lar affricate	cotxe	S
d ₃	dZ	voiced postalveo lar affricate	pla tj a	S
β	В	voiced bilabial approximant	o b ert	В
ð	D	voiced dental approximant	be d oll	Т
j	j	voiced palatal approximant	noia	i
γ	G	voiced velar approximant	pe g a	k
V	V	voiced labiodental fricative	a f gà	f
W	W	voiced labiovelar approximant	aig u a	u

IPA	X-SAMPA	Description	Example	Viseme
Х	x	voiceless velar fricative	J iménez	k
j	j\	voiced palatal fricative	y eso	J
l	l	voiced alveolar lateral approxima nt	alondra	t
θ	Т	voiceless dental fricative	Gon z ález	Т
Vowels				
a	a	open back vowel	c a sa	a
е	е	close-mid front unrounded vowel	ll e nya	е
3	Е	open-mid front unrounded vowel	xec	Е
i	i	closed front unrounded vowel	visca	i
0	o	close-mid back rounded vowel	g o s	0
Э	0	open-mid back rounded vowel	j o c	0
u	u	closed back rounded vowel	un	u
ə	@	mid-central vowel	cas a	@
Additiona	al Symbols			

IPA	X-SAMPA	Description	Example	Viseme
1	п	primary stress	Ala ba ma	
	%	secondary stress	Alabama	
		syllable boundary	A.la.ba.ma	

Chinese (Cantonese) (yue-CN)

The following table lists the Jyutping and International Phonetic Alphabet (IPA) phonemes for the Cantonese voice that is supported by Amazon Polly. Jyutping is a romanization system of Cantonese which is commonly used in academia and among Cantonese speakers. 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. Jyutping examples and the corresponding visemes are also shown.

To make Amazon Polly use phonetic pronunciation with Jyutping, use the phoneme alphabet="x-amazon-jyutping" tag.

The following examples show this with each standard.

Jyutping:

```
<speak>
    ## <phoneme alphabet="x-amazon-jyutping" ph="sing2">#</phoneme>#
    ## <phoneme alphabet="x-amazon-jyutping" ph="seng2">#</phoneme>#
</speak>
```

IPA:

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

X-SAMPA:

```
<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 Cantonese input encoded in UTF-8 only.

Phoneme/Viseme Table

Jyutping	IPA	X- SAMPA	Description	Jyutping Example	Viseme				
Consonar	Consonants								
b	p	p	voiceless bilabial plosive	巴, b aa1	p				
С	ts ^h	ts_h	aspirated voiceless alveolar affricate	叉, c aa1	S				
d	t	t	voiceless alveolar plosive	打, d aa2	t				
f	f	f	voiceless labiodental fricative	花, f aa1	f				
g	k	k	voiceless velar plosive	家, g aa1	k				
gw	k ^w	k_w	labialized voiceless velar plosive	瓜, gw aa1	u				
h	h	h	voiceless glottal fricative	哈, h aa1	k				
k	k ^h	k_h	aspirated voiceless velar plosive	卡, k aa1	k				
kw	k ^{wh}	k_wh	labialized aspirated voiceless velar plosive	誇, kw aa1	u				

Jyutping	IPA	X- SAMPA	Description	Jyutping Example	Viseme
l	l	l	alveolar lateral approximant	啦, l aa1	t
m	m	m	bilabial nasal	媽, m aa1	p
m	m	m=	syllabic bilabial nasal	唔, m 4	p
ng	ŋ	N	velar nasal	牙, ng aa4	k
ng	ŋ	N=	syllabic velar nasal	吳, ng 4	k
n	n	n	alveolar nasal	拿, n aa4	t
р	p ^h	p_h	aspirated voiceless bilabial plosive	趴, p aa1	р
S	S	S	voiceless alveolar fricative	沙, s aa1	S
t	t ^h	t_h	aspirated voiceless alveolar plosive	他, t aa1	t
W	w	W	labio-velar approximant	娃, w aa1	u
у	j	j	palatal approximant	也, j aa5	i
Z	ts	ts	voiceless alveolar affricate	渣, z aa1	S
Vowels					
a	В	6	near-open central vowel	吉, g a t1	a
aa	а	Α	open back unrounded vowel	家, g aa 1	a
aai	ai	Ai	dipthong	街, g aai 1	a

Jyutping	IPA	X- SAMPA	Description	Jyutping Example	Viseme
aau	au	Au	dipthong	交, g aau 1	a
ai	ei	6i	dipthong	雞, g ai 1	a
au	вп	6u	dipthong	溝, k au 1	a
е	3	E	open-mid front unrounded vowel	爹, d e 1	E
ei	ei	ei	dipthong	基, g ei 1	е
ео	Θ	8	close-mid central rounded vowel	春, c eo n1	0
eoi	өу	8y	diphthong	居, g eoi 1	0
eu	εu	Eu	diphthong	掉 in 掉垃圾, d eu 6	Е
i	i	i	close front unrounded vowel	斯, si 1	i
i	I	l	near-close near-front unrounded vowel	激, gik1	i
iu	iu	iu	diphthong	驕, g iu 1	i
0	Э	0	open-mid back rounded vowel	哥, g o 1	0
oe	œ	9	open-mid front rounded vowel	鋸, g oe 3	0
oi	oi	Oi	dipthong	該, g oi 1	0
ou	ou	ou	dipthong	高, g ou 1	0

Jyutping	IPA	X- SAMPA	Description	Jyutping Example	Viseme
u	u	u	close back rounded vowel	姑, g u 1	u
u	υ	U	near-close near-back rounded vowel	谷, g u k5	u
ui	ui	ui	dipthong	攰, g ui 6	u
yu	у	у	close front rounded vowel	於, j yu 1	u
Tone mar	ks and Add	ditional Syı	mbols		
1			high level	詩, si 1	
2			medium rising	史, si 2	
3			medium level	試, si 3	
4			very low level	時, si 4	
5			low rising	市, si 5	
6			low level	是, si 6	
-		•	syllable boundary	語音 jyu5- jam1	

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 with Pinyin, use the phoneme alphabet="x-amazon-phonetic standard used" tag.

The following examples show this with each standard.

Pinyin:

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

IPA:

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

X-SAMPA:

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



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

Pinyin	IPA	X- SAMPA	Description	Pinyin Example	Viseme	
Consonants						

Pinyin	IPA	X- SAMPA	Description	Pinyin Example	Viseme
f	f	f	voiceless labiodental fricative	发, f a1	f
h	h	h	voiceless glottal fricative	和, h e2	k
g	k	k	voiceless velar plosive	古, g u3	k
k	k ^h	k_h	aspirated voiceless velar plosive	苦, k u3	k
l	l	l	alveolar lateral approximant	拉, l a1	t
m	m	m	bilabial nasal	骂, m a4	р
n	n	n	alveolar nasal	那, n a4	t
ng	ŋ	N	velar nasal	正, zhe ng 4	k
b	p	p	voiceless bilabial plosive	爸, b a4	р
р	p ^h	p_h	aspirated voiceless bilabial plosive	怕, p a4	р
S	S	S	voiceless alveolar fricative	四, si4	S
X	Ç	s\	voiceless alveolo-palatal fricative	西, x i1	J
sh	ş	s`	voiceless retroflex fricative	是, shi4	S
d	t	t	voiceless alveolar plosive	打, d a3	t

Pinyin	IPA	X- SAMPA	Description	Pinyin Example	Viseme
t	t ^h	t_h	aspirated voiceless alveolar plosive	他, t a1	t
zh	T ţş	t`s`	voiceless retroflex affricate	之, zh i1	S
ch	 t́ફʰ	t`s`_h	aspirated voiceless retroflex affricate	吃, ch i1	S
S	T ts	ts	voiceless alveolar affricate	字, z i4	S
j	t ç	ts\	voiceless alveolo-palatal affricate	鸡, j i1	J
q	T t¢ʰ	ts_h	aspirated voiceless alveolo-palatal affricate	七, q i1	J
С	⊤ts ʰ	ts_h	aspirated voiceless alveolar affricate	次, c i4	S
W	W	W	labio-velar approximant	我, w o3	u
r	z	z`	voiced retroflex fricative	日, ri4	S
"er" and	"r" colored	syllables			
er	∂ •	@`	r-coloured mid central vowel	二, er4	@
-r			r-colored syllable	馅儿, xian r 4	@
Vowels					
е	¥	7	close-mid back unrounded vowel	恶, e 4	е

Pinyin	IPA	X- SAMPA	Description	Pinyin Example	Viseme	
е	ə	@	mid central vowel	恩, e n1	@	
a	a	a	open front unrounded vowel	安, a n1	a	
ai	aı	al	diphthong	爱, ai 4	а	
ao	аʊ	aU	diphthong	奥, ao 4	a	
ei	еі	е	diphthong	诶, ei 4	е	
е	3	E	open-mid front unrounded vowel	姐, ji e 3	E	
i	i	i	close front unrounded vowel	鸡, j i 1	i	
ou	ου	oU	diphthong	欧, ou 1	0	
0	Э	0	open-mid back rounded vowel	哦, o 4	0	
u	u	u	close back rounded vowel	主, zh u 3	u	
yu	у	у	close front rounded vowel	于, yu 2	u	
Tone mai	Tone marks and Additional Symbols					
1			high level tone	淤, yu1		
2			rising tone	鱼, yu2		
3			low (falling-rising) tone	语, yu3		
4			falling tone	育, yu4		

Pinyin	IPA	X- SAMPA	Description	Pinyin Example	Viseme
0			neutral tone	的, de0	
-	•		syllable boundary	语音 yu3-yin1	

Czech (cs-CZ)

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 Czech voice that is supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Consonal	nts			
p	р	voiceless bilabial plosive	p es	p
t	t	voiceless alveolar plosive	tok	t
С	С	voiceless palatal plosive	t 'uk	J
k	k	voiceless velar plosive	kos	k
b	b	voiced bilabial plosive	b ez	р
d	d	voiced alveolar plosive	d ok	t
Ì	٦/	voiced palatal plosive	d 'as	J

IPA	X-SAMPA	Description	Example	Viseme
g	g	voiced velar plosive	g um	k
f	f	voiceless labiodent al fricative	film	f
V	V	voiced labiodental fricative	v es	f
S	S	voiceless alveolar fricative	s en	S
Z	Z	voiced alveolar fricative	zel	S
ſ	S	voiceless postalveo lar fricative	šel	S
3	Z	voiced postalveo lar fricative	žen	S
X	x	voiceless velar fricative	ch at	k
ĥ	h	voiced glottal fricative	h us	k
ts	ts	voiceless alveolar affricate	co	S
ť	tS	voiceless postalveo lar affricate	č in	S
dz	dz	voiced alveolar affricate	špi c berský	S
dʒ	dZ	voiceless alveolar affricate	dž in	S

IPA	X-SAMPA	Description	Example	Viseme
m	m	bilabial nasal	m or	р
n	n	alveolar nasal	nos	t
'n	J	palatal nasal	ň ader	J
ŋ	N	velar nasal	ba n ka	k
r	r	voiced alveolar trill	r ys	r
Ţ	r_r	voiced raised alveolar fricative trill	řez	r
ĵ	r_0_r	voiceless raised alveolar fricative trill	keř	r
l	l	alveolar lateral approximant	les	t
j	j	palatal approxima nt	jen	i
W	w	labiovelar approximant	W atson	u
r	r_=	syllabic voiced alveolar trill	krk	r
Ţ	l_=	syllabic alveolar lateral approxima nt	v l na	t
Vowels				
a	a	open front unrounded vowel	lan	a

IPA	X-SAMPA	Description	Example	Viseme
a:	a:	long open front unrounded vowel	lán	a
3	Е	open-mid front unrounded vowel	let	E
ε:	E:	long open-mid front unrounded vowel	lét	E
I	l	near-close near- front unrounded vowel	bit	i
i:	i:	long close front unrounded vowel	bít	i
0	0	close-mid back rounded vowel	h o l	0
0.	o:	long close-mid back rounded vowel	g ó l	0
u	u	close back rounded vowel	p u l	u
u:	u:	long close back rounded vowel	půl	u
âu	au	diphthong	auto	a
Êu	Eu	diphthong	euro	Е
O u	ou	diphthong	m ou k	0
Additiona	al Symbols			

IPA	X-SAMPA	Description	Example	Viseme
1	и	primary stress		
•		syllable boundary		

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

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			
b	b	voiced bilabial plosive	b at	р
d	d	voiced alveolar plosive	d a	t
ð	D	voiced dental fricative	ma d, th riller	Т
f	f	voiceless labiodent al fricative	fat	f
g	g	voiced velar plosive	g at	k
h	h	voiceless glottal fricative	h at	k
j	j	palatal approxima nt	jo	i

IPA	X-SAMPA	Description	Example	Viseme
k	k	voiceless velar plosive	k at	k
l	l	alveolar lateral approximant	ladt	t
m	m	bilabial nasal	m at	p
n	n	alveolar nasal	n ay	t
ŋ	N	velar nasal	la ng	k
p	p	voiceless bilabial plosive	p ande	p
r	r	alveolar trill	th riller, sto r y	r
R	R	voiced uvular fricative	rat	k
S	S	voiceless alveolar fricative	s at	S
t	t	voiceless alveolar plosive	tal	t
V	V	voiced labiodental fricative	v at	f
W	w	labial-velar approximant	ha v , w eekend	u
Vowels				
Ø	2	close-mid front rounded vowel	øst	0

IPA	X-SAMPA	Description	Example	Viseme
ø:	2:	long close-mid front rounded vowel	øse	0
В	6	near-open central vowel	m or	a
œ	9	open-mid front rounded vowel	skøn, grønt	0
œ:	9:	long open-mid front rounded vowel	høne, gøre	0
ə	@	mid central vowel	an e	@
æ:	{ :	long near-open front unrounded vowel	m a le	a
а	a	open front unrounded vowel	m a n	a
æ	{	near-open front unrounded vowel	adr e sse	a
а	А	open back unrounded vowel	lak, tak	a
a:	A:	long open back unrounded vowel	r a se	a
е	е	close-mid front unrounded vowel	midt	е

IPA	X-SAMPA	Description	Example	Viseme
e:	e:	long close-mid front unrounded vowel	m e le	е
3	Е	open-mid front unrounded vowel	mæt	Е
ε:	E:	long open-mid front unrounded vowel	mæle	E
i	i	close front unrounded vowel	mit	i
i:	i:	long close front unrounded vowel	mile	i
0	0	close-mid back rounded vowel	foto	0
0:	o:	long close-mid back rounded vowel	m o le	0
Э	0	open-mid back rounded vowel	m u nd	0
0:	O:	long open-mid back rounded vowel	m å le	0
D:	Q:	long open back rounded vowel	m o rse	0
u	u	close back rounded vowel	l u sk	u

IPA	X-SAMPA	Description	Example	Viseme	
u:	u:	long close back rounded vowel	m u le	u	
٨	V	open-mid back unrounded	kør er	Е	
У	У	close front rounded vowel	y t	u	
y:	y:	long close front rounded vowel	h y le	u	
Additional Symbols					
1	п	primary stress	Ala ba ma		
ı	%	secondary stress	A labama		
		syllable boundary	A.la.ba.ma		

Dutch (Belgian) (nl-BE)

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 Belgian Dutch (Flemish) voices that are supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Consonants				
b	b	voiced bilabial plosive	b ak	р
d	d	voiced alveolar plosive	d ak	t

Dutch (Belgian) (nl-BE) 63

IPA	X-SAMPA	Description	Example	Viseme
d ₃	dZ	voiced postalveo lar affricate	mana g er	S
f	f	voiceless labiodent al fricative	fel	f
g	g	voiced velar plosive	g oal	k
γ	G	voiced velar fricative	h oed	k
ĥ	h\	voiced glottal fricative	h and	k
j	j	palatal approxima nt	ja	i
k	k	voiceless velar plosive	k ap	k
l	l	alveolar lateral approximant	land	t
m	m	bilabial nasal	m et	p
n	n	alveolar nasal	n et	t
ŋ	N	velar nasal	ba ng	k
р	р	voiceless bilabial plosive	p ak	p
r	r	alveolar trill	rand	r
S	S	voiceless alveolar fricative	s ein	S

Dutch (Belgian) (nl-BE) 64

IPA	X-SAMPA	Description	Example	Viseme
ſ	S	voiceless postalveo lar fricative	show	S
t	t	voiceless alveolar plosive	t ak	t
V	V	voiced labiodental fricative	v el	f
υ	v\	labiodental approximant	wit	f
X	x	voiceless velar fricative	to ch	k
Z	Z	voiced alveolar fricative	ziin	S
3	Z	voiced postalveo lar fricative	baga g e	S
Vowels				
Ø:	2:	long close-mid front rounded vowel	n eu s	0
œy	9y	dipthong	b ui t	0
ə	@	mid central vowel	de	@
a:	a:	long open front unrounded vowel	b aa d	a
a:	А	open back unrounded vowel	b a d	a

Dutch (Belgian) (nl-BE) 65

IPA	X-SAMPA	Description	Example	Viseme
e:	e:	long close-mid front unrounded vowel	beet	е
3.	3:	long open-mid central unrounded vowel	barri è re	E
3	E	open-mid front unrounded vowel	bed	Е
εί	Ei	dipthong	beet	Е
i	i	close front unrounded vowel	vier	i
I	I	near-close near- front unrounded vowel	pit	i
0:	0:	long close-mid back rounded vowel	b oo t	0
Э	0	open-mid back rounded vowel	p o t	0
u	u	close back rounded vowel	h oe d	u
ΛU	Vu	dipthong	fout	E
y:	y:	long close front rounded vowel	fuut	u

Dutch (Belgian) (nl-BE) 66

IPA	X-SAMPA	Description	Example	Viseme
Y	Υ	near-close near- front rounded vowel	hut	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	Alabama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b ak	р		
d	d	voiced alveolar plosive	d ak	t		
dz	dZ	voiced postalveo lar affricate	mana g er	S		
f	f	voiceless labiodent al fricative	fel	f		

IPA	X-SAMPA	Description	Example	Viseme
g	g	voiced velar plosive	g oal	k
X	G	voiced velar fricative	h oed	k
h	h\	voiced glottal fricative	h and	k
j	j	palatal approxima nt	ja	i
k	k	voiceless velar plosive	k ap	k
l	l	alveolar lateral approximant	land	t
m	m	bilabial nasal	m et	p
n	n	alveolar nasal	net	t
ŋ	N	velar nasal	ba ng	k
p	р	voiceless bilabial plosive	p ak	р
r	r	alveolar trill	rand	r
S	S	voiceless alveolar fricative	s ein	S
l	S	voiceless postalveo lar fricative	show	S
t	t	voiceless alveolar plosive	t ak	t

IPA	X-SAMPA	Description	Example	Viseme
V	V	voiced labiodental fricative	v el	f
υ	v\	labiodental approximant	wit	f
x	X	voiceless velar fricative	to ch	k
Z	z	voiced alveolar fricative	ziin	S
3	Z	voiced postalveo lar fricative	baga g e	S
Vowels				
ø:	2:	long close-mid front rounded vowel	n eu s	0
œy	9y	dipthong	buit	0
Ә	@	mid central vowel	de	@
a:	a:	long open front unrounded vowel	b aa d	a
a:	А	open back unrounded vowel	b a d	a
e:	e:	long close-mid front unrounded vowel	beet	е
3.	3:	long open-mid central unrounded vowel	barri è re	E

IPA	X-SAMPA	Description	Example	Viseme
3	E	open-mid front unrounded vowel	bed	Е
εί	Ei	dipthong	b ee t	E
i	i	close front unrounded vowel	v ie r	i
I	I	near-close near- front unrounded vowel	pit	i
0:	0:	long close-mid back rounded vowel	b oo t	0
Э	0	open-mid back rounded vowel	p o t	0
u	u	close back rounded vowel	h oe d	u
ΛU	Vu	dipthong	fout	Е
y:	y:	long close front rounded vowel	fuut	u
Y	Υ	near-close near- front rounded vowel	hut	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
•		syllable boundary	A.la.ba.ma	

English (US) (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

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b ed	р		
d	d	voiced alveolar plosive	d ig	t		
dz	dZ	voiced postalveo lar affricate	jump	S		
ð	D	voiced dental fricative	th en	Т		
f	f	voiceless labiodent al fricative	five	f		
g	g	voiced velar plosive	g ame	k		
h	h	voiceless glottal fricative	house	k		
j	j	palatal approxima nt	y es	i		
k	k	voiceless velar plosive	c at	k		

English (US) (en-US) 71

IPA	X-SAMPA	Description	Example	Viseme
l	l	alveolar lateral approximant	lay	l
m	m	bilabial nasal	mouse	p
n	n	alveolar nasal	n ap	t
ŋ	N	velar nasal	thi ng	k
p	р	voiceless bilabial plosive	s p eak	р
J	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	t rap	t
_fl	tS	voiceless postalveo lar affricate	ch art	S
θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	w	labial-velar approximant	west	u
Z	Z	voiced alveolar fricative	z ero	S

English (US) (en-US)

IPA	X-SAMPA	Description	Example	Viseme
3	Z	voiced postalveo lar fricative	vision	S
Vowels				
ə	@	mid-central vowel	a ren a	@
ð	@`	mid-central r- colored vowel	read er	@
æ	{	near open-front unrounded vowel	tr a p	a
aı	al	diphthong	price	a
аυ	aU	diphthong	m ou th	a
а	Α	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е
3,	3`	open mid-centr al unrounded r- colored vowel	n ur se	E
3	Е	open mid-front unrounded vowel	dr e ss	Е
i	i	long close front unrounded vowel	fl ee ce	i
I	I	near-close near- front unrounded vowel	kit	i
ου	oU	diphthong	g oa t	0

English (US) (en-US) 73

IPA	X-SAMPA	Description	Example	Viseme
Э	0	long open mid- back rounded vowel	th ou ght	0
OI	OI	diphthong	ch oi ce	0
u	u	long close-back rounded vowel	g oo se	u
ឋ	U	near-close near- back rounded vowel	f oo t	u
٨	V	open-mid-back unrounded vowel	str u t	Е
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
•		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

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			

English (Australian) (en-AU)

IPA	X-SAMPA	Description	Example	Viseme
b	b	voiced bilabial plosive	b ed	р
d	d	voiced alveolar plosive	d ig	t
d ₃	dZ	voiced postalveo lar affricate	jump	S
ð	D	voiced dental fricative	th en	Т
f	f	voiceless labiodent al fricative	f ive	f
g	g	voiced velar plosive	g ame	k
h	h	voiceless glottal fricative	h ouse	k
j	j	palatal approxima nt	y es	i
k	k	voiceless velar plosive	c at	k
l	l	alveolar lateral approximant	lay	t
ļ	l=	syllabic alveolar lateral approxima nt	batt le	t
m	m	bilabial nasal	m ouse	p
ψ	m=	syllabic bilabial nasal	anth em	p

IPA	X-SAMPA	Description	Example	Viseme
n	n	alveolar nasal	n ap	t
ņ	n=	syllabic alveolar nasal	butto n	t
ŋ	N	velar nasal	thi ng	k
p	p	voiceless bilabial plosive	p in	p
1	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	t ask	t
्री	tS	voiceless postalveo lar affricate	ch art	S
Θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	w	labial-velar approximant	west	u
Z	Z	voiced alveolar fricative	zero	S

IPA	X-SAMPA	Description	Example	Viseme
3	Z	voiced postalveo lar fricative	vision	S
Vowels				
Ә	@	mid central vowel	a ren a	@
әυ	@U	diphthong	g oa t	@
æ	{	near open-front unrounded vowel	tr a p	a
aı	al	diphthong	price	a
аบ	aU	diphthong	m ou th	a
a:	A:	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е
3.	3:	long open mid- central unrounded vowel	n ur se	E
3	Е	open mid-front unrounded vowel	dress	Е
63	E@	diphthong	squ are	Е
i:	i	long close front unrounded vowel	fl ee ce	i
I	I	near-close near- front unrounded vowel	kit	i
ΙƏ	I@	diphthong	near	i

IPA	X-SAMPA	Description	Example	Viseme
0:	OI	long open-mid back rounded vowel	th ou ght	0
OI	OI	Diphthong	ch oi ce	0
D	Q	open back rounded vowel	lot	0
u:	u:	long close-back rounded vowel	g oo se	u
ប	U	near-close near- back rounded vowel	foot	u
υə	U@	diphthong	cure	u
٨	V	Open-mid-back unrounded vowel	str u t	Е
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
ı	%	secondary stress	Alabama	
		syllable boundary	A.la.ba.ma	

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.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b ed	p		
d	d	voiced alveolar plosive	d ig	t		
d ₃	dZ	voiced postalveo lar affricate	jump	S		
ð	D	voiced dental fricative	th en	Т		
f	f	voiceless labiodent al fricative	f ive	f		
g	g	voiced velar plosive	g ame	k		
h	h	voiceless glottal fricative	h ouse	k		
j	j	palatal approxima nt	y es	i		
k	k	voiceless velar plosive	c at	k		
l	l	alveolar lateral approximant	lay	t		
J	[=	syllabic alveolar lateral approxima nt	batt le	t		

IPA	X-SAMPA	Description	Example	Viseme
m	m	bilabial nasal	m ouse	p
ψ	m=	syllabic bilabial nasal	anth em	p
n	n	alveolar nasal	n ap	t
ņ	n=	syllabic alveolar nasal	butto n	t
ŋ	N	velar nasal	thi ng	k
p	p	voiceless bilabial plosive	p in	p
J	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	t ask	t
_fl	tS	voiceless postalveo lar affricate	ch art	S
Θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	W	labial-velar approximant	west	u

IPA	X-SAMPA	Description	Example	Viseme
Z	Z	voiced alveolar fricative	zero	S
3	Z	voiced postalveo lar fricative	vi s ion	S
Vowels				
ə	@	mid central vowel	a ren a	@
әʊ	@U	diphthong	g oa t	@
æ	{	near open-front unrounded vowel	trap	a
aı	al	diphthong	price	a
аυ	aU	diphthong	m ou th	a
a:	A:	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е
3.	3:	long open mid- central unrounded vowel	n ur se	E
3	Е	open mid-front unrounded vowel	dr e ss	Е
63	E@	diphthong	squ are	Е
i:	i	long close front unrounded vowel	fl ee ce	i

IPA	X-SAMPA	Description	Example	Viseme
I	1	near-close near- front unrounded vowel	kit	i
IƏ	I@	diphthong	n ear	i
0.	O:	long open-mid back rounded vowel	th ou ght	0
OI	OI	Diphthong	ch oi ce	0
D	Q	open back rounded vowel	lot	0
u:	u:	long close-back rounded vowel	g oo se	u
ប	U	near-close near- back rounded vowel	foot	u
υə	U@	diphthong	cure	u
٨	V	Open-mid-back unrounded vowel	str u t	E
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	Alabama	
		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).

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme			
Consonar	Consonants						
b	b	voiced bilabial plosive	b ed	p			
d	d	voiced alveolar plosive	d ig	t			
d ₃	dZ	voiced postalveo lar affricate	jump	S			
ð	D	voiced dental fricative	th en	Т			
f	f	voiceless labiodent al fricative	five	f			
g	g	voiced velar plosive	g ame	k			
h	h	voiceless glottal fricative	house	k			
j	j	palatal approxima nt	y es	i			
k	k	voiceless velar plosive	c at	k			

IPA	X-SAMPA	Description	Example	Viseme
l	l	alveolar lateral approximant	lay	t
ļ	l=	syllabic alveolar lateral approxima nt	batt le	t
m	m	bilabial nasal	mouse	p
ψ	m=	syllabic bilabial nasal	anth em	р
n	n	alveolar nasal	n ap	t
ņ	n=	syllabic alveolar nasal	n ap	t
ŋ	N	velar nasal	thi ng	k
p	p	voiceless bilabial plosive	p in	р
J	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	task	t
्री	tS	voiceless postalveo lar affricate	ch art	S

IPA	X-SAMPA	Description	Example	Viseme
Θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	W	labial-velar approximant	west	u
Z	z	voiced alveolar fricative	z ero	S
3	Z	voiced postalveo lar fricative	vision	S
Vowels				
Ә	@	mid central vowel	a ren a	@
әυ	@U	diphthong	g oa t	@
æ	{	near open-front unrounded vowel	trap	a
aı	al	diphthong	price	a
аʊ	aU	diphthong	m ou th	a
a:	A:	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е
3.	3:	long open mid- central unrounded vowel	n ur se	E

IPA	X-SAMPA	Description	Example	Viseme
3	Е	open mid-front unrounded vowel	dr e ss	Е
εә	E@	diphthong	squ are	Е
i:	i	long close front unrounded vowel	fl ee ce	i
I	1	near-close near- front unrounded vowel	kit	i
ΙĐ	I@	diphthong	n ear	i
0.	OI	long open-mid back rounded vowel	th ou ght	0
OI	OI	Diphthong	ch oi ce	0
D	Q	open back rounded vowel	lot	0
u:	u:	long close-back rounded vowel	g oo se	u
ប	U	near-close near- back rounded vowel	f oo t	u
υə	U@	diphthong	cure	u
٨	V	Open-mid-back unrounded vowel	strut	Е
Addition	al Symbols			
ı	п	primary stress	Ala ba ma	

IPA	X-SAMPA	Description	Example	Viseme
	%	secondary stress	A labama	
•		syllable boundary	A.la.ba.ma	

English (Ireland) (en-IE)

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 Irish English voices that are supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			
b	b	voiced bilabial plosive	b ed	p
d	d	voiced alveolar plosive	d ig	t
dz	dZ	voiced postalveo lar affricate	jump	S
ð	D	voiced dental fricative	th en	Т
f	f	voiceless labiodent al fricative	five	f
g	g	voiced velar plosive	g ame	k
h	h	voiceless glottal fricative	house	k

English (Ireland) (en-IE) 87

IPA	X-SAMPA	Description	Example	Viseme
j	j	palatal approxima nt	y es	i
k	k	voiceless velar plosive	c at	k
l	ι	alveolar lateral approximant	lay	t
m	m	bilabial nasal	mouse	p
n	n	alveolar nasal	n ap	t
ŋ	N	velar nasal	thi ng	k
p	р	voiceless bilabial plosive	s p eak	p
J	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	t rap	t
्री	tS	voiceless postalveo lar affricate	ch art	S
θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f

English (Ireland) (en-IE) 88

IPA	X-SAMPA	Description	Example	Viseme
W	W	labial-velar approximant	west	u
Z	Z	voiced alveolar fricative	z ero	S
3	Z	voiced postalveo lar fricative	vi s ion	S
Vowels				
ə	@	mid-central vowel	a ren a	@
э	@`	mid-central r- colored vowel	read er	@
æ	{	near open-front unrounded vowel	trap	а
aı	al	diphthong	price	а
аʊ	aU	diphthong	m ou th	a
а	Α	long open-back unrounded vowel	father	a
еі	el	diphthong	f a ce	е
3,	3`	open mid-centr al unrounded r- colored vowel	n ur se	E
3	Е	open mid-front unrounded vowel	dr e ss	Е
i	i	long close front unrounded vowel	fl ee ce	i

English (Ireland) (en-IE)

IPA	X-SAMPA	Description	Example	Viseme
I	1	near-close near- front unrounded vowel	kit	i
೦೮	oU	diphthong	g oa t	0
Э	0	long open mid- back rounded vowel	th ou ght	0
OI	OI	diphthong	ch oi ce	0
u	u	long close-back rounded vowel	g oo se	u
ប	U	near-close near- back rounded vowel	foot	u
٨	V	open-mid-back unrounded vowel	str u t	Е
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	Alabama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			
b	b	voiced bilabial plosive	b ed	p
d	d	voiced alveolar plosive	d ig	t
dz	dZ	voiced postalveo lar affricate	jump	S
ð	D	voiced dental fricative	th en	Т
f	f	voiceless labiodent al fricative	five	f
g	g	voiced velar plosive	g ame	k
h	h	voiceless glottal fricative	house	k
j	j	palatal approxima nt	y es	i
k	k	voiceless velar plosive	c at	k
l	ι	alveolar lateral approximant	lay	t
Ţ	[=	syllabic alveolar lateral approxima nt	batt le	t

IPA	X-SAMPA	Description	Example	Viseme
m	m	bilabial nasal	m ouse	p
ψ	m=	syllabic bilabial nasal	anth em	p
n	n	alveolar nasal	n ap	t
ņ	n=	syllabic alveolar nasal	butto n	t
ŋ	N	velar nasal	thi ng	k
p	p	voiceless bilabial plosive	p in	p
J	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	t ask	t
_fl	tS	voiceless postalveo lar affricate	ch art	S
Θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	W	labial-velar approximant	west	u

IPA	X-SAMPA	Description	Example	Viseme
Z	Z	voiced alveolar fricative	zero	S
3	Z	voiced postalveo lar fricative	vision	S
Vowels				
Э	@	mid central vowel	a ren a	@
эυ	@U	diphthong	g oa t	@
æ	{	near open-front unrounded vowel	tr a p	a
aı	al	diphthong	price	a
аυ	aU	diphthong	m ou th	a
a:	A:	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е
3.	3:	long open mid- central unrounded vowel	n ur se	E
3	Е	open mid-front unrounded vowel	dr e ss	Е
63	E@	diphthong	squ are	Е
i:	i	long close front unrounded vowel	fl ee ce	i

IPA	X-SAMPA	Description	Example	Viseme
I	I	near-close near- front unrounded vowel	kit	i
ΙƏ	I@	diphthong	near	i
0.	O:	long open-mid back rounded vowel	th ou ght	0
OI	OI	Diphthong	ch oi ce	0
D	Q	open back rounded vowel	lot	0
u:	u:	long close-back rounded vowel	g oo se	u
ប	U	near-close near- back rounded vowel	foot	u
υə	U@	diphthong	cure	u
٨	V	Open-mid-back unrounded vowel	str u t	Е
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
ı	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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.

English	Maori
Hello/cheers	Kia ora
Welcome (to)	Nau mai (ki)
Hello (one person)/thank you	Tēnā koe
Hello (three or more people)/thank you	Tēnā koutou
Good morning	Ata mārie
Good morning	Mōrena
Thank you	Ngā mihi
Take care	Ngā manaakitanga
See you	Ka kite
See you later	Mā te wā
Have a good day	Kia pai tō rā
Merry Christmas	Meri Kirihimete
Maori	Māori
Maori language	te reo Māori
Maori language week	Te wiki o te reo Māori
New Zealand	Aotearoa
Maori New Year	Mātariki
Town in New Zealand / Waitangi Day is the national day of New Zealand	Waitangi
One	tahi
Two	rua

English	Maori
Three	toru
Four	whā
Five	rima
Six	ono
Seven	whitu
Eight	waru
Nine	iwa
Ten	tekau
Twenty	rua tekau
Thirty	Toru tekau

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

IPA	X-SAMPA	Description	Example	Viseme
Consona	nts			
b	b	voiced bilabial plosive	b ed	р
d	d	voiced alveolar plosive	d ig	t

IPA	X-SAMPA	Description	Example	Viseme
dz	dZ	voiced postalveo lar affricate	jump	S
ð	D	voiced dental fricative	th en	Т
f	f	voiceless labiodent al fricative	five	f
g	g	voiced velar plosive	g ame	k
h	h	voiceless glottal fricative	h ouse	k
j	j	palatal approxima nt	y es	i
k	k	voiceless velar plosive	c at	k
l	l	alveolar lateral approximant	lay	t
ļ	l=	syllabic alveolar lateral approxima nt	batt le	t
†	K	voiceless lateral fricative	um hl anga	t
m	m	bilabial nasal	mouse	p
ψ	m=	syllabic bilabial nasal	anth em	p
n	n	alveolar nasal	n ap	t

IPA	X-SAMPA	Description	Example	Viseme
ņ	n=	syllabic alveolar nasal	butt on	t
ŋ	N	velar nasal	thi ng	k
p	р	voiceless bilabial plosive	p in	p
1	r\	alveolar approxima nt	red	r
r	r	alveolar trill	pareis	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S
t	t	voiceless alveolar plosive	t ask	t
t f	tS	voiceless postalveo lar affricate	ch art	S
Θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	w	labial-velar approximant	west	u
X	x	voiceless velar fricative	g auteng	k

IPA	X-SAMPA	Description	Example	Viseme
Z	z	voiced alveolar fricative	zero	S
!	!\	post-alveolar click	gq eberha	k
I	Ν	dental click	n c ube	t
II	\	lateral click	xh osa	t
Vowels				
ə	@	mid central vowel	a rena	@
əi	@i	diphthong	nelspr ui t	i
әυ	@U	diphthong	g oa t	@
æ	{	near open-front unrounded vowel	trap	a
аі	al	diphthong	price	a
аυ	aU	diphthong	m ou th	a
a:	A:	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е
3.	3:	long open mid- central unrounded vowel	n ur se	E
3	Е	open mid-front unrounded vowel	dr e ss	E
63	E@	diphthong	squ are	Е

IPA	X-SAMPA	Description	Example	Viseme
i:	i	long close front unrounded vowel	fl ee ce	i
iə	I@	diphthong	du pr eez	i
I	1	near-close near- front unrounded vowel	kit	i
IÐ	I@	diphthong	near	i
э:	O:	long open-mid back rounded vowel	th ou ght	0
OI	OI	Diphthong	ch oi ce	0
D	Q	open back rounded vowel	lot	0
u:	u:	long close-back rounded vowel	g oo se	u
ប	U	near-close near- back rounded vowel	f oo t	u
υә	U@	diphthong	cure	u
٨	V	Open-mid-back unrounded vowel	strut	Е
У	У	close front rounded vowel	van v uu ren	u
Additional Symbols				
1	II	primary stress	Ala ba ma	

IPA	X-SAMPA	Description	Example	Viseme
1	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

English (Welsh) (en-GB-WLS)

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

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b ed	р		
d	d	voiced alveolar plosive	d ig	t		
dz	dZ	voiced postalveo lar affricate	jump	S		
ð	D	voiced dental fricative	th en	Т		
f	f	voiceless labiodent al fricative	five	f		
g	g	voiced velar plosive	g ame	k		
h	h	voiceless glottal fricative	h ouse	k		

101

English (Welsh) (en-GB-WLS)

IPA	X-SAMPA	Description	Example	Viseme
j	j	palatal approxima nt	y es	i
k	k	voiceless velar plosive	c at	k
l	l	alveolar lateral approximant	lay	t
Ţ	l =	syllabic alveolar lateral approxima nt	batt le	t
m	m	bilabial nasal	mouse	р
ψ	m=	syllabic bilabial nasal	anth em	p
n	n	alveolar nasal	n ap	t
ù	n=	syllabic alveolar nasal	n ap	t
ŋ	N	velar nasal	thi ng	k
p	p	voiceless bilabial plosive	p in	p
J	r\	alveolar approxima nt	red	r
S	S	voiceless alveolar fricative	s eem	S
ſ	S	voiceless postalveo lar fricative	sh ip	S

IPA	X-SAMPA	Description	Example	Viseme
t	t	voiceless alveolar plosive	t ask	t
⁻tſ	tS	voiceless postalveo lar affricate	ch art	S
Θ	Т	voiceless dental fricative	th in	Т
V	V	voiced labiodental fricative	v est	f
W	W	labial-velar approximant	west	u
Z	z	voiced alveolar fricative	z ero	S
3	Z	voiced postalveo lar fricative	vision	S
Vowels				
ə	@	mid central vowel	a ren a	@
əυ	@U	diphthong	g oa t	@
æ	{	near open-front unrounded vowel	trap	a
aı	al	diphthong	price	a
аυ	aU	diphthong	m ou th	a
a:	A:	long open-back unrounded vowel	f a ther	a
еі	el	diphthong	f a ce	е

IPA	X-SAMPA	Description	Example	Viseme
3.	3:	long open mid- central unrounded vowel	n ur se	E
3	E	open mid-front unrounded vowel	dr e ss	Е
εә	E@	diphthong	squ are	E
i:	i	long close front unrounded vowel	fl ee ce	i
I	1	near-close near- front unrounded vowel	kit	i
ΙĐ	I@	diphthong	n ear	i
9.	OI	long open-mid back rounded vowel	th ou ght	0
OI	OI	Diphthong	ch oi ce	0
D	Q	open back rounded vowel	lot	0
u:	u:	long close-back rounded vowel	g oo se	u
ឋ	U	near-close near- back rounded vowel	foot	u
υə	U@	diphthong	cure	u
٨	V	Open-mid-back unrounded vowel	strut	E

IPA	X-SAMPA	Description	Example	Viseme	
Additional Symbols					
1	п	primary stress	Ala ba ma		
ı	%	secondary stress	A labama		
		syllable boundary	A.la.ba.ma		

Finnish (fi-FI)

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 Finnish voice that is supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Finnish c	onsonants			
p	р	voiceless bilabial plosive	[p]ankki	p
t	t	voiceless alveolar plosive	[t]alo	t
k	k	voiceless velar plosive	[k]aali	k
d	d	voiced alveolar plosive	[d]ata	t
S	S	voiceless alveolar fricative	[s]ali	S
h	h	voiceless glottal fricative	[h]attu	k

IPA	X-SAMPA	Description	Example	Viseme
υ	v\	voiced labiodental approximant	[v]aiva	V
j	j	palatal approxima nt	[j]oki	i
l	l	alveolar lateral approximant	[l]oma	t
r	r	voiced alveolar trill	[r]iita	r
m	m	bilabial nasal	[m]ato	p
n	n	alveolar nasal	[n]enäa	t
ŋ	N	velar nasal	he[n]ki	k
Consonar	nts found in loanword	ls		
b	b	voiced bilabial plosive	[b]ussi	р
f	f	voiceless labiodent al fricative	[f]irma	V
W	w	labial-velar approximant	[w]iki	u
Z	Z	voiced alveolar fricative	[z]ulu	S
g	g	voiced velar plosive	[g]aala	k
ſ	S	voiceless postalveo lar fricative	[sh]akki	S

IPA	X-SAMPA	Description	Example	Viseme
3	Z	voiced postalveo lar fricative	[g]enre	S
θ	Т	voiceless dental fricative	ear[th]	Т
ð	D	voiced dental fricative	ei[th]er	Т
Short vov	wels			
i	i	close front unrounded vowel	k[i]lo	i
3	Е	open mid-front unrounded vowel	k[e]sä	E
æ	{	near open-front unrounded vowel	k[ä]ly	А
у	У	close front rounded vowel	k[y]lä	u
Ø	2	close mid-front rounded vowel	p[ö]ly	0
u	u	close back rounded vowel	k[u]lo	u
Э	0	open mid-back rounded vowel	k[o]lo	0
а	А	open back unrounded vowel	k[a]la	А
Long vow	vels			

IPA	X-SAMPA	Description	Example	Viseme
i:	i:	long close front unrounded vowel	s[ii]li	i
ε:	E:	long open mid- front unrounded vowel	[ee]tu	E
æ:	{ :	long near open- front unrounded vowel	t[ää]llä	A
y:	y:	long close front unrounded vowel	t[yy]li	u
ø:	2:	long close mid- front rounded vowel	t[öö]lö	0
u:	u:	long close back rounded vowel	t[uu]li	u
0.	O:	long open mid- back rounded vowel	r[oo]li	0
a:	A:	long open back unrounded vowel	k[aa]su	Α
Dipthong	S			
εί	Ei	dipthong	l[ei]pä	Е
æi	{i	dipthong	[äi]ti	Α
ui	ui	dipthong	k[ui]n	u
ai	Ai	dipthong	k[ai]kki	Α

IPA	X-SAMPA	Description	Example	Viseme
ic	Oi	dipthong	p[oi]ka	0
øi	2i	dipthong	s[öi]n	0
yi	yi	dipthong	l[yi]jy	u
au	Au	dipthong	s[au]na	А
ou	Ou	dipthong	k[ou]lu	0
εu	Eu	dipthong	r[eu]na	Е
iu	iu	dipthong	v[iu]lu	i
æy	{y	dipthong	t[äy]nnä	A
øy	2y	dipthong	k[öy]hä	0
εγ	Еу	dipthong	pes[ey]tyä	Е
iy	iy	dipthong	käär[iy]tyä	i
iε	iE	dipthong	t[ie]	i
yø	y2	dipthong	[yö]	u
uɔ	uO	dipthong	t[uo]	u
Vowels fo	ound in English loanw	ords		
I	I	near-close near- front unrounded vowel	b[i]t	i
υ	U	near-close near- back rounded vowel	b[oo]k	u
ə	@	mid-central vowel	[a]bout	@

IPA	X-SAMPA	Description	Example	Viseme
٨	V	open-mid-back unrounded vowel	c[u]t	Е

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

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b oire	р		
d	d	voiced alveolar plosive	ma d ame	t		
f	f	voiceless labiodent al fricative	f emme	f		
g	g	voiced velar plosive	g rand	k		
Ч	Н	labial-palatal approximant	br u it	u		
j	j	palatal approxima nt	mei ll eur	i		
k	k	voiceless velar plosive	q uatre	k		

French (fr-FR) 110

IPA	X-SAMPA	Description	Example	Viseme
l	l	alveolar lateral approximant	malade	t
m	m	bilabial nasal	m aison	p
n	n	alveolar nasal	astro n ome	t
'n	J	palatal nasal	bai gn er	J
ŋ	N	velar nasal	parki ng	k
p	p	voiceless bilabial plosive	p omme	p
R	R	voiced uvular fricative	amou r eux	k
S	S	voiceless alveolar fricative	s anté	S
l	S	voiceless postalveo lar fricative	ch at	S
t	t	voiceless alveolar plosive	t éléphone	t
V	V	voiced labiodental fricative	v rai	f
W	w	labial-velar approximant	s o ir	u
Z	Z	voiced alveolar fricative	rai s on	S
3	Z	voiced postalveo lar fricative	auber g ine	S

French (fr-FR)

IPA	X-SAMPA	Description	Example	Viseme
Vowels				
Ø	2	close-mid front rounded vowel	d eu x	0
œ	9	open-mid front rounded vowel	n eu f	0
œ̃	9~	nasal open-mid front rounded vowel	br un	0
ə	@	mid central vowel	je	@
a	a	open front unrounded vowel	t a ble	a
ã	A~	nasal open back unrounded vowel	cam em bert	a
е	е	close-mid front unrounded vowel	march é	е
3	E	open-mid front unrounded vowel	n ei ge	Е
ε̃	E~	nasal open-mid front unrounded vowel	sap in	E
i	i	close front unrounded vowel	mille	i
0	o	close-mid back rounded vowel	h ô pital	0
Э	0	open-mid back rounded vowel	h o mme	0

French (fr-FR)

IPA	X-SAMPA	Description	Example	Viseme
õ	O~	nasal open-mid back rounded vowel	b on	O
u	u	close back rounded vowel	s ou s	u
у	У	close front rounded vowel	dur	u
Addition	al Symbols			
1	n .	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

French (Belgian) (fr-BE)

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 Belgian French voices that are supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Consonai	nts			
b	b	voiced bilabial plosive	b oire	р
d	d	voiced alveolar plosive	ma d ame	t

IPA	X-SAMPA	Description	Example	Viseme
f	f	voiceless labiodent al fricative	femme	f
g	g	voiced velar plosive	g rand	k
Ч	Н	labial-palatal approximant	br u it	u
j	j	palatal approxima nt	mei ll eur	i
k	k	voiceless velar plosive	q uatre	k
l	l	alveolar lateral approximant	malade	t
m	m	bilabial nasal	m aison	р
n	n	alveolar nasal	astro n ome	t
'n	J	palatal nasal	bai gn er	J
ŋ	N	velar nasal	parki ng	k
р	р	voiceless bilabial plosive	p omme	р
R	R	voiced uvular fricative	amou r eux	k
S	S	voiceless alveolar fricative	s anté	S
ſ	S	voiceless postalveo lar fricative	ch at	S

IPA	X-SAMPA	Description	Example	Viseme
t	t	voiceless alveolar plosive	t éléphone	t
V	V	voiced labiodental fricative	v rai	f
W	W	labial-velar approximant	s o ir	u
Z	z	voiced alveolar fricative	rai s on	S
3	Z	voiced postalveo lar fricative	auber g ine	S
Vowels				
Ø	2	close-mid front rounded vowel	d eu x	0
œ	9	open-mid front rounded vowel	n eu f	0
œ̃	9~	nasal open-mid front rounded vowel	br un	0
Ә	@	mid central vowel	je	@
a	a	open front unrounded vowel	t a ble	a
ã	A~	nasal open back unrounded vowel	cam em bert	a
е	е	close-mid front unrounded vowel	march é	е

IPA	X-SAMPA	Description	Example	Viseme
3	Е	open-mid front unrounded vowel	n ei ge	Е
Ē	E~	nasal open-mid front unrounded vowel	sap in	E
i	i	close front unrounded vowel	mille	i
0	o	close-mid back rounded vowel	h ô pital	0
Э	0	open-mid back rounded vowel	h o mme	0
õ	0~	nasal open-mid back rounded vowel	b on	0
u	u	close back rounded vowel	s ou s	u
У	У	close front rounded vowel	dur	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
•		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b oire	p		
d	d	voiced alveolar plosive	ma d ame	t		
f	f	voiceless labiodent al fricative	f emme	f		
g	g	voiced velar plosive	g rand	k		
Ч	Н	labial-palatal approximant	br u it	u		
j	j	palatal approxima nt	mei ll eur	i		
k	k	voiceless velar plosive	q uatre	k		
l	l	alveolar lateral approximant	malade	t		
m	m	bilabial nasal	m aison	p		
n	n	alveolar nasal	astro n ome	t		
'n	J	palatal nasal	bai gn er	J		

French (Canadian) (fr-CA)

IPA	X-SAMPA	Description	Example	Viseme
ŋ	N	velar nasal	parki ng	k
p	p	voiceless bilabial plosive	p omme	p
R	R	voiced uvular fricative	amou r eux	k
S	S	voiceless alveolar fricative	s anté	S
ſ	S	voiceless postalveo lar fricative	ch at	S
t	t	voiceless alveolar plosive	t éléphone	t
V	V	voiced labiodental fricative	v rai	f
W	w	labial-velar approximant	s o ir	u
Z	Z	voiced alveolar fricative	rai s on	S
3	Z	voiced postalveo lar fricative	auber g ine	S
Vowels				
Ø	2	close-mid front rounded vowel	d eu x	o
œ	9	open-mid front rounded vowel	n eu f	0

French (Canadian) (fr-CA) 118

IPA	X-SAMPA	Description	Example	Viseme
œ̃	9~	nasal open-mid front rounded vowel	br un	0
ə	@	mid central vowel	j e	@
a	a	open front unrounded vowel	t a ble	a
ã	A~	nasal open back unrounded vowel	cam em bert	a
е	е	close-mid front unrounded vowel	march é	е
3	Е	open-mid front unrounded vowel	n ei ge	Е
Ē	E~	nasal open-mid front unrounded vowel	sap in	E
i	i	close front unrounded vowel	mille	i
0	o	close-mid back rounded vowel	h ô pital	o
Э	0	open-mid back rounded vowel	h o mme	0
õ	0~	nasal open-mid back rounded vowel	b on	0
u	u	close back rounded vowel	s ou s	u

French (Canadian) (fr-CA) 119

IPA	X-SAMPA	Description	Example	Viseme
у	У	close front rounded vowel	dur	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
ı	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme		
Consonai	Consonants					
7	?	glottal stop				
b	b	voiced bilabial plosive	B ier	p		
d	d	voiced alveolar plosive	D ach	t		
Ç	С	voiceless palatal fricative	ich	k		
dz	dZ	voiced postalveo lar affricate	Dsch ungel	S		

IPA	X-SAMPA	Description	Example	Viseme
f	f	Voiceless labiodent al fricative	V ogel	f
g	g	Voiced velar plosive	G abel	k
h	h	Voiceless glottal fricative	Haus	k
j	j	Voiceless glottal fricative	jemand	i
k	k	Voiceless velar plosive	K leid	k
l	l	Alveolar lateral approximant	Loch	t
m	m	Bilabial nasal	Milch	p
n	n	Alveolar nasal	N atur	t
ŋ	N	Velar nasal	kli ng en	k
р	р	Voiceless bilabial plosive	P ark	р
pf	pf	Voiceless labiodent al affricate	A pf el	
R	R	Uvular trill	Regen	
S	S	voiceless alveolar fricative	Me ss er	S
ſ	S	Voiceless postalveolar fricative	Fi sch er	S

IPA	X-SAMPA	Description	Example	Viseme
t	t	Voiceless alveolar plosive	Topf	Т
t s	Ts	Voiceless alveolar affricate	Z ahl	
_fl	tS	Voiceless postalveolar affricate	deu tsch	S
V	V	Voiced labiodental fricative	W asser	f
X	х	Voiceless velar fricative	ko ch en	k
Z	z	Voiced alveolar fricative	See	S
3	Z	Voiced postalveo lar fricative	Oran g e	S
Vowels				
Ø:	2:	long close-mid front rounded vowel	b ö se	0
В	6	near-open central vowel	bess er	a
Ř	6_^	non-syllabic near- open central vowel	Klar	a
œ	9	open-mid front rounded vowel	k ö nnen	0
Э	@	mid central vowel	Red e	@

IPA	X-SAMPA	Description	Example	Viseme
a	a	open front unrounded vowel	Salz	a
a:	a:	long open front unrounded vowel	S ah ne	a
aı	al	diphthong	nein	a
аυ	aU	diphthong	Au gen	a
ã	A~	nasal open back unrounded vowel	Restaur ant	a
e:	e:	long close-mid front unrounded vowel	R e de	е
3	Е	open-mid front unrounded vowel	K e ller	Е
ε̃	E~	nasal open-mid front unrounded vowel	Terr ain	E
i:	i:	long close front unrounded vowel	Lied	i
I	1	near-close near- front unrounded vowel	bitte	i
0:	o:	long close-mid back rounded vowel	K oh l	0
Э	0	open-mid back rounded vowel	K o ffer	0

IPA	X-SAMPA	Description	Example	Viseme
õ	0~	nasal open-mid back rounded vowel	Ann on ce	0
ΟΥ	OY	diphthong	neu	0
u:	u:	long close back rounded vowel	Br u der	u
ឋ	U	near-close near- back rounded vowel	W u nder	u
y:	y:	long close front rounded vowel	k üh l	u
Y	Υ	near-close near- front rounded vowel	K ü che	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
I	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

German (Austrian) (de-AT)

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 Austrian German voices that are supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
7	?	glottal stop				
b	b	voiced bilabial plosive	B ier	р		
d	d	voiced alveolar plosive	D ach	t		
Ç	С	voiceless palatal fricative	ich	k		
dz	dZ	voiced postalveo lar affricate	Dsch ungel	S		
f	f	Voiceless labiodent al fricative	V ogel	f		
g	g	Voiced velar plosive	G abel	k		
h	h	Voiceless glottal fricative	Haus	k		
j	j	Voiceless glottal fricative	jemand	i		
k	k	Voiceless velar plosive	K leid	k		
l	T	Alveolar lateral approximant	Loch	t		
m	m	Bilabial nasal	Milch	р		
n	n	Alveolar nasal	Natur	t		

IPA	X-SAMPA	Description	Example	Viseme
ŋ	N	Velar nasal	kli ng en	k
p	p	Voiceless bilabial plosive	P ark	p
pf	pf	Voiceless labiodent al affricate	A pf el	
R	R	Uvular trill	Regen	
S	S	voiceless alveolar fricative	Me ss er	S
l	S	Voiceless postalveolar fricative	Fi sch er	S
t	t	Voiceless alveolar plosive	Topf	Т
t s	Ts	Voiceless alveolar affricate	Zahl	
─tſ	tS	Voiceless postalveolar affricate	deu tsch	S
V	V	Voiced labiodental fricative	W asser	f
X	x	Voiceless velar fricative	ko ch en	k
Z	z	Voiced alveolar fricative	S ee	S
3	Z	Voiced postalveo lar fricative	Oran g e	S

IPA	X-SAMPA	Description	Example	Viseme
Vowels				
Ø:	2:	long close-mid front rounded vowel	b ö se	0
В	6	near-open central vowel	bess er	a
Ĕ	6_^	non-syllabic near- open central vowel	Klar	a
œ	9	open-mid front rounded vowel	k ö nnen	0
ə	@	mid central vowel	Red e	@
a	a	open front unrounded vowel	Salz	a
a:	a:	long open front unrounded vowel	S ah ne	a
aı	al	diphthong	nein	a
аυ	aU	diphthong	Augen	a
ã	A~	nasal open back unrounded vowel	Restaur ant	a
e:	e:	long close-mid front unrounded vowel	R e de	е
3	Е	open-mid front unrounded vowel	K e ller	Е

IPA	X-SAMPA	Description	Example	Viseme
ε̃	E~	nasal open-mid front unrounded vowel	Terr ain	Е
i:	i:	long close front unrounded vowel	Lied	i
I	1	near-close near- front unrounded vowel	bitte	i
0:	o:	long close-mid back rounded vowel	K oh l	o
Э	0	open-mid back rounded vowel	K o ffer	0
õ	O~	nasal open-mid back rounded vowel	Ann on ce	0
ΟΥ	OY	diphthong	n eu	0
u:	u:	long close back rounded vowel	Br u der	u
ប	U	near-close near- back rounded vowel	W u nder	u
y:	y:	long close front rounded vowel	k üh l	ü
Y	Υ	near-close near- front rounded vowel	K ü che	u

IPA	X-SAMPA	Description	Example	Viseme
Addition	al Symbols			
1	11	primary stress	Ala ba ma	
ı	%	secondary stress	Alabama	
		syllable boundary	A.la.ba.ma	

German (Swiss standard) (de-CH)

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 (Swiss standard) voice that is supported by Amazon Polly.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme		
Consonai	Consonants					
7	?	glottal stop				
b	b	voiced bilabial plosive	B ier	р		
d	d	voiced alveolar plosive	D ach	t		
Ç	С	voiceless palatal fricative	ich	k		
dz	dZ	voiced postalveo lar affricate	Dsch ungel	S		
f	f	Voiceless labiodent al fricative	V ogel	f		

IPA	X-SAMPA	Description	Example	Viseme
g	g	Voiced velar plosive	G abel	k
h	h	Voiceless glottal fricative	Haus	k
j	j	Voiceless glottal fricative	j emand	i
k	k	Voiceless velar plosive	K leid	k
l	l	Alveolar lateral approximant	Loch	t
m	m	Bilabial nasal	Milch	p
n	n	Alveolar nasal	Natur	t
ŋ	N	Velar nasal	kli ng en	k
p	p	Voiceless bilabial plosive	Park	p
pf	pf	Voiceless labiodent al affricate	A pf el	
R	R	Uvular trill	Regen	
S	S	voiceless alveolar fricative	Me ss er	S
l	S	Voiceless postalveolar fricative	Fi sch er	S
t	t	Voiceless alveolar plosive	Topf	Т

IPA	X-SAMPA	Description	Example	Viseme
ີts	Ts	Voiceless alveolar affricate	Z ahl	
्री	tS	Voiceless postalveolar affricate	deu tsch	S
V	V	Voiced labiodental fricative	W asser	f
Х	x	Voiceless velar fricative	ko ch en	k
Z	Z	Voiced alveolar fricative	See	S
3	Z	Voiced postalveo lar fricative	Oran g e	S
Vowels				
Ø:	2:	long close-mid front rounded vowel	b ö se	0
В	6	near-open central vowel	bess er	a
Ğ	6_^	non-syllabic near- open central vowel	Klar	a
œ	9	open-mid front rounded vowel	k ö nnen	0
Э	@	mid central vowel	Red e	@
a	a	open front unrounded vowel	Salz	a

IPA	X-SAMPA	Description	Example	Viseme
a:	a:	long open front unrounded vowel	S ah ne	a
aı	al	diphthong	n ei n	a
аบ	aU	diphthong	Au gen	a
ã	A~	nasal open back unrounded vowel	Restaur ant	a
e:	e:	long close-mid front unrounded vowel	R e de	е
3	E	open-mid front unrounded vowel	K e ller	Е
ε̃	E~	nasal open-mid front unrounded vowel	Terr ain	E
i:	i:	long close front unrounded vowel	Lied	i
I	1	near-close near- front unrounded vowel	bitte	i
o:	o:	long close-mid back rounded vowel	Kohl	0
Э	0	open-mid back rounded vowel	K o ffer	0

IPA	X-SAMPA	Description	Example	Viseme
õ	0~	nasal open-mid back rounded vowel	Ann on ce	0
ΟΥ	OY	diphthong	neu	0
u:	u:	long close back rounded vowel	Br u der	u
ឋ	U	near-close near- back rounded vowel	W u nder	u
y:	y:	long close front rounded vowel	k üh l	u
Y	Υ	near-close near- front rounded vowel	K ü che	u
Additional Symbols				
1	и	primary stress	Ala ba ma	
I	%	secondary stress	A labama	
•		syllable boundary	A.la.ba.ma	

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).

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example
Consonants			
p ^h	p_h	voiceless aspirated bilabial plosive	फूल (phool)
b ^{fi}	b_h	voiced aspirated bilabial plosive	भारी (bhaari)
<u>t</u>	t_d	voiceless dental plosive	तापमान (taapmaan)
ţ ^h	t_d_h	voiceless aspirated dental plosive	थोड़ा (thoda)
ď	d_d	voiced dental plosive	दल्लि (dilli)
Ğμ	d_d_h	voiced aspirated dental plosive	धोबी (dhobi)
t	t`	voiceless retroflex plosive	कटोरा (katora)
t h	t`_h	voiceless aspirated retroflex plosive	ਠਂਫ (thand)
d	d`	voiced retroflex plosive	डर (darr)
d ⁿ	d`_h	voiced aspirated retroflex plosive	ढाल (dhal)
tʃʰ	tS_h	voiceless aspirated palatal affricate	छाल (chaal)
d3 ^h	dZ_h	voiced aspirated palatal affricate	झाल (jhaal)
k ^h	k_h	voiceless aspirated velar plosive	खान (khan)

IPA	X-SAMPA	Description	Example
g ^h	g_h	voiced aspirated velar plosive	घान (ghaan)
η	n`	retroflex nasal	क्षण (kshan)
ſ	4	alveolar flap	राम (ram)
r	r`	plain retroflex flap	बड़ा (bada)
(^h	r`_h	voiced aspirated retroflex flap	बढ़ी (barhi)
υ	v\	bilabial approximant	वसूल (wasool)
Vowels			
ə	@_o	mid central vowel	अच्छा (achhaa)
ð	@~	nasalised mid central vowel	हँसना (hansnaa)
a	A_o	open front unrounded vowel	आग (aag)
a~	A~	nasalised open front unrounded vowel	घडियाँ (ghariyaan)
I	l_o	near-close near-front unrounded vowel	इक्कीस (ikkees)
Ĩ	 ~	nasalised near-close near front unrounded vowel	संचािई (sinchai)
i	i_o	close front unrounded vowel	बल्ली (billee)
ĩ	i~	nasalised close front unrounded vowel	नही (nahin)

IPA	X-SAMPA	Description	Example
ប	U_o	near-close near-back rounded vowel	उलूल (ullu)
ũ	U~	nasalised near-close near-back rounded vowel	मुँह (munh)
u	u_o	close back rounded vowel	फूल (phool)
u~	u~	nasalised close back rounded vowel	ऊँट (oont)
Э	O_o	open-mid back rounded vowel	कौन (kaun)
õ	O~	nasalised open-mid back rounded vowel	भौ (bhaun)
0	0	close-mid back rounded vowel	सोना (sona)
o~	0~	nasalised close-mid back rounded vowel	क्यो (kyon)
3	E_o	open-mid front unrounded vowel	पैसा (paisa)
ε̃	E~	nasalised open-mid front unrounded vowel	में (main)
е	е	close-mid front unrounded vowel	एक (ek)
e~	e~	nasalised close-mid front unrounded vowel	कतिाबें (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

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			
b	b	voiced bilabial plosive	gras b akkanum	0
С	С	voiceless palatal plosive	pak k in	k
C _h	c_h	aspirated voiceless palatal plosive	anar k istai	k
Ç	С	voiceless palatal fricative	h éðan	k
d	d	voiced alveolar plosive	bón d i	t
ð	D	voiced dental fricative	bor ð	Т
f	f	voiceless labiodent al fricative	du f t	f
g	g	voiced velar plosive	hol g óma	k
γ	G	voiced velar fricative	hu g ur	k

IPA	X-SAMPA	Description	Example	Viseme
h	h	voiceless glottal fricative	h eili	k
j	j	palatal approxima nt	jökull	i
k ^h	k_h	aspirated voiceless velar plosive	ós k öpunum	k
l	l	alveolar lateral approximant	gó l f	t
Ĵ	l_0	voiceless alveolar lateral approxima nt	fólk	t
m	m	bilabial nasal	septe m ber	р
ŵ	m_0	voiceless bilabial nasal	ko m pa	р
n	n	alveolar nasal	n úmer	t
ů	n_0	voiceless alveolar nasal	pö n tun	t
n	J	palatal nasal	pæli n gar	J
ŋ	N	velar nasal	sö n gvarann	k
ŋ˚	N_0	voiceless velar nasal	fræ n ka	k
p ^h	p_h	aspirated voiceless bilabial plosive	af p lánun	p
r	r	alveolar trill	afsk r ifta	r

IPA	X-SAMPA	Description	Example	Viseme
î.	r_0	voiceless alveolar trill	andvö r pum	r
S	S	voiceless alveolar fricative	baðhú s	S
t h	t_h	aspirated voiceless alveolar plosive	t anki	t
θ	Т	voiceless dental fricative	þ eldökki	Т
V	V	voiced labiodental fricative	sil f ur	f
W	w	labial-velar approximant		u
X	X	voiceless velar fricative	samféla g s	k
Vowels				
œ	9	open-mid front rounded vowel	þr ö skuldinum	0
œ:	9:	long open-mid front rounded vowel	tv ö	0
a	a	open front unrounded vowel	nefn a	a
a:	a:	long open front unrounded vowel	f a ra	a
au	au	diphthong	á tta	a

IPA	X-SAMPA	Description	Example	Viseme
au:	au:	diphthong	á tján	a
3	Е	open-mid front unrounded vowel	k e nnari	Е
ε:	E:	long open-mid front unrounded vowel	dr e ka	E
i	i	close front unrounded vowel	Gúl í ver	i
i:	i:	long close front unrounded vowel	þr í r	i
I	1	near-close near- front unrounded vowel	samsp i l	i
I:	l:	long near-clos e near-front unrounded vowel	stig	i
Э	0	open-mid back rounded vowel	regndr o par	0
0:	O:	long open-mid back rounded vowel	ullarb o lur	0
ou	Ou	diphthong	t ó lf	0
ou:	Ou:	diphthong	fj ó rir	0
u	u	close back rounded vowel	st ú lkan	u

IPA	X-SAMPA	Description	Example	Viseme
u:	u:	long close back rounded vowel	fr ú	u
Y	Υ	near-close near- front rounded vowel	tíu	u
Y:	Υ	long near-clos e near-front rounded vowel	gr u ninn	u
Addition	al Symbols			
Г	п	primary stress	Ala ba ma	
ı	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme	
Consonants					
b	b	voiced bilabial plosive	b acca	р	
d	d	voiced alveolar plosive	d ama	t	

Italian (it-IT)

IPA	X-SAMPA	Description	Example	Viseme
dz	dz	voiced alveolar affricate	zero	S
dz	dZ	voiced postalveo lar affricate	g iro	S
f	f	voiceless labiodent al fricative	f amiglia	f
g	g	voiced velar plosive	g atto	k
h	h	voiceless glottal fricative	h orror	k
j	j	palatal approxima nt	d i eci	i
k	k	voiceless velar plosive	c ampo	k
l	ι	alveolar lateral approximant	lido	t
٨	L	palatal lateral approximant	a gli o	J
m	m	bilabial nasal	m ille	р
n	n	alveolar nasal	nove	t
'n	J	palatal nasal	lasa gn e	J
p	p	voiceless bilabial plosive	p izza	p
r	r	alveolar trill	risata	r

Italian (it-IT)

IPA	X-SAMPA	Description	Example	Viseme
S	S	voiceless alveolar fricative	sei	S
ſ	S	voiceless postalveo lar fricative	sci enza	S
t	t	voiceless alveolar plosive	t avola	t
t s	ts	voiceless alveolar affricate	for z a	S
t f	tS	voiceless postalveo lar affricate	ci elo	S
V	V	voiced labiodental fricative	v enti	f
W	W	labial-velar approximant	quattro	u
Z	z	voiced alveolar fricative	bi s ogno	S
3	Z	voiced postalveo lar fricative	bi j ou	S
Vowels				
a	a	open front unrounded vowel	a rco	a
е	е	close-mid front unrounded vowel	tre	е
3	Е	open-mid front unrounded vowel	ettaro	Е

Italian (it-IT)

IPA	X-SAMPA	Description	Example	Viseme
i	i	close front unrounded vowel	impero	i
0	0	close-mid back rounded vowel	cent o	0
Э	0	open-mid back rounded vowel	o tto	0
u	u	close back rounded vowel	uno	u
Addition	al Symbols			
1	"	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
•		syllable boundary	A.la.ba.ma	

Japanese (ja-JP)

Amazon Polly supports the Pronunciation Kana and Yomigana alphabets for Japanese. To make Amazon Polly use phonetic pronunciation with these alphabets, use the phoneme alphabet="x-amazon-phonetic standard used" attribute.

- x-amazon-pron-kana indicates that Pronunciation Kana is used. Pronunciation Kana are special Katakana characters used for phonetic transcription and can encode pitch accent.
- x-amazon-yomigana indicates that Yomigana is used. Yomigana can be conventional Katakana, Hiragana, and Latin alphabets interpreted as hepburn romanization.

The following examples show how these are used:

Pronunciation Kana

<spea< th=""><th>ak></th><th></th><th></th><th></th></spea<>	ak>			

```
###<phoneme alphabet="x-amazon-pron-kana" ph="###'#">##</phoneme>###
</speak>
```

Yomigana

```
<speak>
    ###<phoneme alphabet="x-amazon-yomigana" ph="####">##</phoneme>###
    ###<phoneme alphabet="x-amazon-yomigana" ph="###">##</phoneme>###
    ###<phoneme alphabet="x-amazon-yomigana" ph="Hirokazu">##</phoneme>###
</speak>
```

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.

IPA	X-SAMPA	Description	Example	Viseme		
Consonai	Consonants					
١	4	alveolar flap	練習, renshuu	t		
7	?	glottal stop	あつっ, atsu'			
b	b	voiced bilabial plosive	舞踊, b uyou	р		
β	В	voiced bilabial fricative	ヴィンテージ, v inteeji	В		
С	С	voiceless palatal plosive	ききょう, kiky ou	k		
Ç	С	voiceless palatal fricative	人, h ito	k		
d	d	voiced alveolar plosive	濁点, d akuten	t		
dz.	dz\	voiced alveolo-p alatal affricate	純, j un	J		

IPA	X-SAMPA	Description	Example	Viseme
g	g	voiced velar plosive	ご飯, g ohan	k
h	h	voiceless glottal fricative	本, h on	k
j	j	palatal approxima nt	屋根, y ane	i
Ĵ	٦/	voiced palatal plosive	行儀, gy ou g i	J
k	k	voiceless velar plosive	漢字, k anji	k
J	I\	alveolar lateral flap	釣り, tsuri	r
Jj	l∖j	alveolar lateral flap, palatal approximant	流行, ry uukou	r
m	m	bilabial nasal	飯, m eshi	p
n	n	alveolar nasal	猫, n eko	t
'n	J	palatal nasal	日本, n ippon	J
N	N\	uvular nasal	缶, ka n	k
p	p	voiceless bilabial plosive	パン, p an	p
ф	p\	voiceless bilabial fricative	福, h uku	f
S	S	voiceless alveolar fricative	層, s ou	S

IPA	X-SAMPA	Description	Example	Viseme
G	s\	voiceless alveolo-p alatal fricative	書簡, sh okan	J
t	t	voiceless alveolar plosive	手紙, t egami	t
t s	ts	voiceless alveolar affricate	釣り, ts uri	S
⁻tç	ts\	voiceless alveolo-p alatal affricate	吉, ki ch i	J
W	W	labial-velar approximant	電話, den w a	u
Z	z	voiced alveolar fricative	座敷, z ashiki	S
Vowels				
ä:	a:_"	long open central unrounded vowel	羽蟻, h aa ri	a
ä	a_"	open central unrounded vowel	仮名, k a n a	a
e:	e:_o	long mid front unrounded vowel	学生, gakus ei	@
е	e_o	mid front unrounded vowel	歴, r e ki	@
i	i	close front unrounded vowel	気, ki	i
i:	i:	long close front unrounded vowel	詩歌, shiika	i

IPA	X-SAMPA	Description	Example	Viseme
ш	М	close back unrounded vowel	運, u n	i
w:	M:	long close back unrounded vowel	宗教, sh uu kyou	i
0:	0:_0	long mid back rounded vowel	購読, k oo doku	0
0	0_0	mid back rounded vowel	読者, d o kusha	0

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.

IPA	X-SAMPA	Description	Example	Viseme		
Consona	Consonants					
k	k	voiceless velar plosive	강, [g]ang	k		
k#	k_t	strong voiceless velar plosive	깨, [kk]e	k		
n	n	alveolar nasal	남, [n]am	t		
t	t	voiceless alveolar plosive	도, [d]o	t		
t#	t_t	strong voiceless alveolar plosive	때, [tt]e	t		

Korean (ko-KR) 148

IPA	X-SAMPA	Description	Example	Viseme
١	4	alveolar flap	사랑, sa[r]ang	t
l	l	alveolar lateral approximant	돌, do[l]	t
m	m	bilabial nasal	무, [m]u	р
p	p	voiceless bilabial plosive	봄, [b]om	р
p#	p_t	strong voiceless bilabial plosive	뻘, [pp]eol	p
S	S	voiceless alveolar fricative	새, [s]e	S
s#	s_t	strong voiceless alveolar fricative	씨, [ss]i	S
ŋ	N	velar nasal	방, ba[ng]	k
Tç	ts\	voiceless alveolo-p alatal affricate	조, [j]o	J
T#¢	ts_t	strong voiceless alveolo-palatal affricate	찌 , [jj]i	J
Tt ¢ʰ	ts_h	aspirated voiceless alveolo-palatal affricate	차, [ch]a	J
k ^h	k_h	aspirated voiceless velar plosive	코, [k]o	k
t ^h	t_h	aspirated voiceless alveolar plosive	통, [t]ong	t

Korean (ko-KR) 149

IPA	X-SAMPA	Description	Example	Viseme
p ^h	p_h	aspirated voiceless bilabial plosive	패, [p]e	р
h	h	voiceless glottal fricative	힘, [h]im	k
j	j	palatal approxima nt	양, [y]ang	i
W	W	labial-velar approximant	왕, [w]ang	u
щ	M\	velar approxima nt>	의, [wj]i	i
Vowels				
а	a	open front unrounded vowel	밥, b[a]b	a
٨	V	open-mid back unrounded vowel	정, j[eo]ng	Е
3	E	open-mid front unrounded vowel	배, b[e]	E
0	0	close-mid back rounded vowel	노, n[o]	0
u	u	close back rounded vowel	둘, d[u]l	u
ш	М	close back unrounded vowel	은, [eu]n	i
i	i	close front unrounded vowel	김, k[i]m	i

Korean (ko-KR)

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		
Consonar	Consonants					
١	4	alveolar flap	prøv	t		
b	b	voiced bilabial plosive	la bb	p		
Ç	С	voiceless palatal fricative	k ino	k		
d	d	voiced alveolar plosive	la dd	t		
đ	d`	voiced retroflex plosive	ve rd i	t		
f	f	voiceless labiodent al fricative	fot	f		
g	g	voiced velar plosive	ta gg	k		
h	h	voiceless glottal fricative	ha	k		
j	j	palatal approxima nt	gi	i		
k	k	voiceless velar plosive	ta kk	k		

IPA	X-SAMPA	Description	Example	Viseme
l	ι	alveolar lateral approximant	fall, ball	t
l	ľ	retroflex lateral approximant	æ rl ig	t
m	m	bilabial nasal	la m	p
n	n	alveolar nasal	va nn	t
η	n`	retroflex nasal	ga rn	t
ŋ	N	velar nasal	sa ng	k
p	p	voiceless bilabial plosive	ho pp	р
S	S	voiceless alveolar fricative	la ss	S
Ş	s`	voiceless retroflex fricative	å rs	S
ſ	S	voiceless postalveo lar fricative	s kyt	S
t	t	voiceless alveolar plosive	la t	t
t	t`	voiceless retroflex plosive	ha rdt	t
υ	v\	labiodental approximant	v in	f
W	W	labial-velar approximant	will	х

IPA	X-SAMPA	Description	Example	Viseme
Vowels				
Ø:	2:	long close-mid front rounded vowel	søt	0
œ	9	open-mid front rounded vowel	søtt	0
ə	@	mid central vowel	ap e	@
æ:	{ :	long near-open front unrounded vowel	vær	a
u	}	close central rounded vowel	l u nd	u
u ː	} :	long close central rounded vowel	lun	u
æ	{	near-open front unrounded vowel	vært	a
а	Α	open back unrounded vowel	hatt	a
a:	A:	long open back unrounded vowel	hat	a
e:	e:	long close-mid front unrounded vowel	sen	е
3	E	open-mid front unrounded vowel	send	Е

IPA	X-SAMPA	Description	Example	Viseme
i:	i:	long close front unrounded vowel	vin	i
I	I	near-close near- front unrounded vowel	vind	i
0.	o:	long close-mid back rounded vowel	våt	0
Э	0	open-mid back rounded vowel	vått	0
u:	u:	long close back rounded vowel	b o k	u
υ	U	near-close near- back rounded vowel	b u kk	u
y:	y:	long close front rounded vowel	l y n	u
Y	Υ	near-close near- front rounded vowel	l y nne	u
Addition	al Symbols			
ı	П	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b obas, b elka	р		
d	d	voiced alveolar plosive	dar, do	t		
dz	dz	voiced alveolar affricate	dz won, wi dz owie	S		
dz	dz\	voiced alveolo-p alatal affricate	dź więk	J		
dz	dz`	voiced retroflex affricate	dż em, dż ungla	S		
f	f	voiceless labiodent al fricative	furtka, film	f		
g	g	voiced velar plosive	g azeta, wa g a	k		
h	h	voiceless glottal fricative	ch leb, h andel	k		
j	j	palatal approxima nt	jak, maja	i		

Polish (pl-PL)

IPA	X-SAMPA	Description	Example	Viseme
k	k	voiceless velar plosive	k ura, mare k	k
l	l	alveolar lateral approximant	lipa, alicja	t
m	m	bilabial nasal	m atka, m olo	p
n	n	alveolar nasal	n orka	t
'n	J	palatal nasal	ko ń , toru ń	J
p	p	voiceless bilabial plosive	p ora, sto p	p
r	r	alveolar trill	rok, park	r
S	S	voiceless alveolar fricative	sum, pas	S
G	s\	voiceless alveolo-p alatal fricative	ś ruba, ś nieg	J
Ş	s`	voiceless retroflex fricative	szum, masz	S
t	t	voiceless alveolar plosive	tok, stół	t
t s	ts	voiceless alveolar affricate	car, co	S
⁻ tç	ts\	voiceless alveolo-p alatal affricate	ć ma, mie ć	J
t ş	ts`	voiceless retroflex affricate	czas, raczej	S

Polish (pl-PL)

IPA	X-SAMPA	Description	Example	Viseme
V	V	voiced labiodental fricative	worek, mewa	f
W	W	labial-velar approximant	łaska, mało	u
Z	Z	voiced alveolar fricative	z ero	S
Z	z\	voiced alveolo-p alatal fricative	źrebię, bieliźnie	J
Z,	z`	voiced retroflex fricative	żar, żona	S
Vowels				
а	a	open front unrounded vowel	ja	a
3	Е	open-mid front unrounded vowel	echo	Е
ε̃	E~	nasal open-mid front unrounded vowel	w ę że	E
i	i	close front unrounded vowel	ile	i
Э	0	open-mid back rounded vowel	o czy	0
õ	O~	nasal open-mid back rounded vowel	w ą ż	0

Polish (pl-PL)

IPA	X-SAMPA	Description	Example	Viseme
u	u	close back rounded vowel	uczta	u
i	1	close central unrounded vowel	b y k	i
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
ı	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme		
Consonai	Consonants					
ſ	4	alveolar flap	pira	t		
b	b	voiced bilabial plosive	d ato	р		
d	d	voiced alveolar plosive	d ato	t		
f	f	voiceless labiodent al fricative	facto	f		

Portuguese (pt-PT) 158

IPA	X-SAMPA	Description	Example	Viseme
g	g	voiced velar plosive	g ato	k
j	j	palatal approxima nt	paragua y	i
k	k	voiceless velar plosive	c acto	k
l	l	alveolar lateral approximant	galo	t
λ	L	palatal lateral approximant	ga lh o	J
m	m	bilabial nasal	m ato	p
n	n	alveolar nasal	n ato	t
'n	J	palatal nasal	pi nh a	J
p	p	voiceless bilabial plosive	p ato	p
R	R\	uvular trill	barroso	k
S	S	voiceless alveolar fricative	s aca	S
ſ	S	voiceless postalveo lar fricative	ch ato	S
t	t	voiceless alveolar plosive	tacto	t
V	V	voiced labiodental fricative	v aca	f

Portuguese (pt-PT) 159

IPA	X-SAMPA	Description	Example	Viseme
W	w	labial-velar approximant	ma u	u
Z	Z	voiced alveolar fricative	z aca	S
3	Z	voiced postalveo lar fricative	jacto	S
Vowels				
а	a	open front unrounded vowel	parto	a
a~	a~	nasal open front unrounded vowel	p e ga	a
е	е	close-mid front unrounded vowel	p e ga	е
e~	e~	nasal close-mid front unrounded vowel	movem	е
3	Е	open-mid front unrounded vowel	caf é	Е
i	i	close front unrounded vowel	lingueta	i
ĩ	i~	nasal close front unrounded vowel	cinto	i
0	0	close-mid back rounded vowel	p o der	0

Portuguese (pt-PT) 160

IPA	X-SAMPA	Description	Example	Viseme
o~	0~	nasal close-mid back rounded vowel	c o mpra	0
Э	0	open-mid back rounded vowel	cot ó	0
u	u	close back rounded vowel	fui	u
u~	u~	nasal close back rounded vowel	sunto	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	Alabama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme
Consonai	nts			
ſ	4	alveolar flap	pira	t
b	b	voiced bilabial plosive	b ato	р

IPA	X-SAMPA	Description	Example	Viseme
d	d	voiced alveolar plosive	d ato	t
dz	dZ	voiced postalveo lar affricate	ida de	S
f	f	voiceless labiodent al fricative	f acto	f
g	g	voiced velar plosive	g ato	k
j	j	palatal approxima nt	paragua y	i
k	k	voiceless velar plosive	c acto	k
l	l	alveolar lateral approximant	galo	t
λ	L	palatal lateral approximant	ga lh o	J
m	m	bilabial nasal	m ato	p
n	n	alveolar nasal	n ato	t
'n	J	palatal nasal	pi nh a	J
p	p	voiceless bilabial plosive	p ato	p
S	S	voiceless alveolar fricative	s aca	S
ſ	S	voiceless postalveo lar fricative	ch ato	S

IPA	X-SAMPA	Description	Example	Viseme
t	t	voiceless alveolar plosive	t acto	t
्री	tS	voiceless postalveo lar affricate	noi te	S
V	V	voiced labiodental fricative	v aca	f
W	W	labial-velar approximant	ma u	u
Χ	X	voiceless uvular fricative	carro	k
Z	z	voiced alveolar fricative	z aca	S
3	Z	voiced postalveo lar fricative	jacto	S
Vowels				
а	a	open front unrounded vowel	p a rto	a
a~	a~	nasal open front unrounded vowel	pens a mos	a
е	е	close-mid front unrounded vowel	p e ga	е
e [~]	e~	nasal close-mid front unrounded vowel	movem	е
3	Е	open-mid front unrounded vowel	caf é	Е

IPA	X-SAMPA	Description	Example	Viseme
i	i	close front unrounded vowel	lingueta	i
ĩ	j~	nasal close front unrounded vowel	cinto	i
0	O	close-mid back rounded vowel	p o der	0
o~	0~	nasal close-mid back rounded vowel	c o mpra	0
Э	0	open-mid back rounded vowel	cot ó	0
u	u	close back rounded vowel	fui	u
u~	u~	nasal close back rounded vowel	sunto	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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.

Romanian (ro-RO) 164

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
b	b	voiced bilabial plosive	b ubă	p		
d	d	voiced alveolar plosive	d upă	t		
d̃ʒ	dZ	voiced postalveo lar affricate	ge orge	S		
f	f	voiceless labiodent al fricative	a f acere	f		
g	g	voiced velar plosive	a g riș	k		
h	h	voiceless glottal fricative	h arpă	k		
j	j	palatal approxima nt	baie	i		
k	k	voiceless velar plosive	c oș	k		
l	l	alveolar lateral approximant	lampa	t		
m	m	bilabial nasal	m ama	p		
n	n	alveolar nasal	n or	t		
p	p	voiceless bilabial plosive	pilă	p		
r	r	alveolar trill	rampă	r		

Romanian (ro-RO) 165

IPA	X-SAMPA	Description	Example	Viseme
S	S	voiceless alveolar fricative	soare	S
ſ	S	voiceless postalveo lar fricative	ma ș ină	S
t	t	voiceless alveolar plosive	t ata	t
t s	ts	voiceless alveolar affricate	ț ară	S
्री	tS	voiceless postalveo lar affricate	ce ai	S
V	V	voiced labiodental fricative	v iață	f
W	W	labial-velar approximant	bea u	u
Z	z	voiced alveolar fricative	mo z ol	S
3	Z	voiced postalveo lar fricative	j oacă	S
Vowels				
ə	@	mid central vowel	bab ă	@
a	a	open front unrounded vowel	c a sa	a
е	е	close-mid front unrounded vowel	e lan	е

Romanian (ro-RO) 166

IPA	X-SAMPA	Description	Example	Viseme
ě	e_^	non-syllabic close-mid front unrounded vowel	beau	е
i	i	close front unrounded vowel	m i e	i
0	0	close-mid back rounded vo	o ră	0
oa	o_^a	diphthong	oa re	0
u	u	close back rounded vowel	u nde	u
÷	1	close central unrounded vowel	Rom â nia	i
Addition	al Symbols			
1	11	primary stress	Ala ba ma	
г	%	secondary stress	A labama	
		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

IPA	X-SAMPA	Description	Example	Viseme
Consonants				

IPA	X-SAMPA	Description	Example	Viseme
b	b	voiced bilabial plosive	б орт	р
b ⁱ	b'	palatalized voiced bilabial plosive	б юро	р
d	d	voiced alveolar plosive	дом	t
d ^j	d'	palatalized voiced alveolar plosive	д я д я	t
f	f	voiceless labiodent al fricative	флаг	f
fi	f'	palatalized voiceless labiodent al fricative	ф евраль	f
g	g	voiced velar plosive	нога	k
g ⁱ	g'	palatalized voiced velar plosive	герой	k
j	j	palatal approxima nt	диза й н, я щик	i
k	k	voiceless velar plosive	кот	k
k ^j	k'	palatalized voiceless velar plosive	к ино	k
l	l	alveolar lateral approximant	лампа	t

IPA	X-SAMPA	Description	Example	Viseme
Įi	l'	palatalized alveolar lateral approximant	лес	t
m	m	bilabial nasal	м ама	p
m ^j	m'	palatalized bilabial nasal	мяч	p
n	n	alveolar nasal	нос	t
n ^j	n'	palatalized alveolar nasal	няня	t
p	р	voiceless bilabial plosive	п апа	p
p ^j	p'	palatalized voiceless bilabial plosive	перо	p
r	r	alveolar trill	р оза	r
r ⁱ	r'	palatalized alveolar trill	р юмка	r
S	S	voiceless alveolar fricative	с ыр	S
S ^j	s'	palatalized voiceless alveolar fricative	с ердце, ру сь	S
G:	s\:	long voiceless alveolo-palatal fricative	щ ека	J

IPA	X-SAMPA	Description	Example	Viseme
Ş	s`	voiceless retroflex fricative	шум	S
t	t	voiceless alveolar plosive	т очка	t
t ⁱ	t'	palatalized voiceless alveolar plosive	тётя	t
t s	ts	voiceless alveolar affricate	ц арь	S
tç	ts\	voiceless alveolo-p alatal affricate	час	J
V	V	voiced labiodental fricative	вор	f
V ^j	v'	palatalized voiced labiodental fricative	верфь	f
X	x	voiceless velar fricative	хор	k
X ^j	x'	palatalized voiceless velar fricative	х ими х	k
Z	Z	voiced alveolar fricative	зуб	S
Z ^j	z'	palatalized voiced alveolar fricative	з има	S

IPA	X-SAMPA	Description	Example	Viseme
Z :	z\:	long voiced alveolo-palatal fricative	уе зж ать	J
ζ	z`	voiced retroflex fricative	жена	S
Vowels				
Ә	@	mid central vowel	канарейк а	@
a	a	open front unrounded vowel	дв а, я блоко	a
е	е	close-mid front unrounded vowel	печь	е
3	Е	open-mid front unrounded vowel	Э ТО	Е
i	i	close front unrounded vowel	од и н, ч е тыре	i
0	0	close-mid back rounded vowel	к о т	0
u	u	close back rounded vowel	м у ж, вь ю га	u
ŧ	1	close central unrounded vowel	М ы ШЬ	i

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.

Spanish (es-ES) 171

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme	
Consonants					
ſ	4	alveolar flap	pero, bravo, amor, eterno	t	
b	b	voiced bilabial plosive	b e stia	p	
β	В	voiced bilabial fricative	be b é	В	
d	d	voiced alveolar plosive	cuan d o	t	
ð	D	voiced dental fricative	ar d er	Т	
f	f	voiceless labiodent al fricative	fase, café	f	
g	g	voiced velar plosive	gato, lengua, guerra	k	
Y	G	voiced velar fricative	tri g o, Ar g os	k	
j	j	palatal approxima nt	hacia, tierra, radio, viuda	i	
j	j\	voiced palatal fricative	enhielar, sa y o, in y ectado, des y erba	J	
k	k	voiceless velar plosive	caña, laca, q uisimos	k	

Spanish (es-ES) 172

IPA	X-SAMPA	Description	Example	Viseme
l	l	alveolar lateral approximant	lino, calor, principal	t
λ	L	palatal lateral approximant	llave, pollo	J
m	m	bilabial nasal	m adre, co m er, a n fibio	p
n	n	alveolar nasal	nido, anillo, sin	t
'n	J	palatal nasal	caba ñ a, ñ oquis	J
ŋ	N	velar nasal	ci n co, ve n ga	k
p	р	voiceless bilabial plosive	pozo, topo	p
r	r	alveolar trill	perro, enrachado	r
S	S	voiceless alveolar fricative	saco, casa, puertas	S
t	t	voiceless alveolar plosive	t amiz, á t omo	t
्री	tS	voiceless postalveo lar affricate	ch ubasco	S
θ	Т	voiceless dental fricative	cereza, zorro, la c ero, paz	Т
W	w	labial-velar approximant	fuego, fuimos, cuota, cuadro	u
X	x	voiceless velar fricative	jamón, g eneral, su j e, reloj	k

Spanish (es-ES) 173

IPA	X-SAMPA	Description	Example	Viseme		
Z	Z	voiced alveolar fricative	ra s go, mi s mo	S		
Vowels	Vowels					
a	a	open front unrounded vowel	tanque	a		
е	е	close-mid front unrounded vowel	peso	е		
i	i	close front unrounded vowel	cinco	i		
0	0	close-mid back rounded vowel	b o sque	0		
u	u	close-mid front unrounded vowel	p u blicar	u		
Addition	al Symbols					
1	п	primary stress	Ala ba ma			
	%	secondary stress	A labama			
		syllable boundary	A.la.ba.ma			

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.

Spanish (Mexican) (es-MX) 174

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme		
Consonar	Consonants					
ſ	4	alveolar flap	pero, bravo, amor, eterno	t		
b	b	voiced bilabial plosive	b estia	p		
β	В	voiced bilabial fricative	be b é	В		
d	d	voiced alveolar plosive	cuan d o	t		
ð	D	voiced dental fricative	ar d er	Т		
f	f	voiceless labiodent al fricative	fase, café	f		
g	g	voiced velar plosive	g ato, len g ua, g uerra	k		
Y	G	voiced velar fricative	tri g o, Ar g os	k		
j	j	palatal approxima nt	hacia, tierra, radio, viuda	i		
Ţ	j\	voiced palatal fricative	enh i elar, sa y o, in y ectado, des y erba	J		
k	k	voiceless velar plosive	c aña, la c a, q uisimos	k		

Spanish (Mexican) (es-MX) 175

IPA	X-SAMPA	Description	Example	Viseme
l	l	lateral alveolar approximant	lino, calor, principal	t
m	m	bilabial nasal	m adre, co m er, a n fibio	p
n	n	alveolar nasal	nido, anillo, sin	t
'n	J	palatal nasal	caba ñ a, ñ oquis	J
ŋ	N	velar nasal	angosto, increíble	k
p	p	voiceless bilabial plosive	pozo, topo	р
r	r	alveolar trill	perro, enrachado	r
S	S	voiceless alveolar fricative	saco, casa, puertas	S
ſ	S	voiceless postalveo lar fricative	show, flash	S
t	t	voiceless alveolar plosive	tamiz, átomo	t
्री	tS	voiceless postalveo lar affricate	ch ubasco	S
W	w	labial-velar approximant	fuego, fuimos, cuota, cuadro	u
X	x	voiceless velar fricative	jamón, g eneral, peaje, reloj	k
Z	Z	voiced alveolar fricative	rasgo, mismo	S

Spanish (Mexican) (es-MX) 176

IPA	X-SAMPA	Description	Example	Viseme		
Vowels	Vowels					
а	a	central open unrounded vowel	t a nque	a		
е	е	close-mid front unrounded vowel	p e so	е		
i	i	close front unrounded vowel	cinco	i		
0	o	close-mid back rounded vowel	b o sque	0		
u	u	close back rounded vowel	p u blicar	u		
Addition	al Symbols					
ı	П	primary stress	Ala ba ma			
ı	%	secondary stress	A labama			
		syllable boundary	A.la.ba.ma			

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

IPA	X-SAMPA	Description	Example	Viseme
Consonar	nts			

Spanish (US) (es-US) 177

IPA	X-SAMPA	Description	Example	Viseme
ſ	4	alveolar flap	pero, bravo, amor, eterno	t
b	b	voiced bilabial plosive	b estia	р
β	В	voiced bilabial fricative	be b é	В
d	d	voiced alveolar plosive	cuan d o	t
ð	D	voiced dental fricative	ar d er	Т
f	f	voiceless labiodent al fricative	fase, café	f
g	g	voiced velar plosive	g ato, len g ua, g uerra	k
X	G	voiced velar fricative	tri g o, Ar g os	k
j	j	palatal approxima nt	hacia, tierra, radio, viuda	i
Ţ	j\	voiced palatal fricative	enh i elar, sa y o, in y ectado, des y erba	J
k	k	voiceless velar plosive	c aña, la c a, q uisimos	k
l	ι	lateral alveolar approximant	lino, calor, principal	t

Spanish (US) (es-US) 178

IPA	X-SAMPA	Description	Example	Viseme
m	m	bilabial nasal	madre, comer, anfibio	p
n	n	alveolar nasal	nido, anillo, sin	t
л	J	palatal nasal	caba ñ a, ñ oquis	J
ŋ	N	velar nasal	angosto, increíble	k
p	р	voiceless bilabial plosive	pozo, topo	p
r	r	alveolar trill	perro, enrachado	r
S	S	voiceless alveolar fricative	saco, casa, puertas	S
ſ	S	voiceless postalveo lar fricative	show, flash	S
t	t	voiceless alveolar plosive	tamiz, átomo	t
्री	tS	voiceless postalveo lar affricate	ch ubasco	S
W	w	labial-velar approximant	fuego, fuimos, cuota, cuadro	u
X	x	voiceless velar fricative	jamón, g eneral, peaje, reloj	k
Z	Z	voiced alveolar fricative	rasgo, mismo	S
Vowels				

Spanish (US) (es-US) 179

IPA	X-SAMPA	Description	Example	Viseme
a	a	central open unrounded vowel	tanque	a
е	е	close-mid front unrounded vowel	p e so	е
i	i	close front unrounded vowel	cinco	i
0	0	close-mid back rounded vowel	b o sque	0
u	u	close back rounded vowel	p u blicar	u
Addition	al Symbols			
I	II	primary stress	Ala ba ma	
ı	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme
Consonai	nts			
b	b	voiced bilabial plosive	bil	p

IPA	X-SAMPA	Description	Example	Viseme
d	d	voiced alveolar plosive	d al	t
đ	d`	voiced retroflex plosive	bo rd	t
f	f	voiceless labiodent al fricative	fil	f
g	g	voiced velar plosive	g ås	k
h	h	voiceless glottal fricative	h al	k
j	j	palatal approxima nt	j ag	i
k	k	voiceless velar plosive	k al	k
l	l	alveolar lateral approximant	lös	t
l	ľ	retroflex lateral approximant	hä rl ig	t
m	m	bilabial nasal	mil	p
n	n	alveolar nasal	n ålar	t
η	n`	retroflex nasal	ba rn	t
ŋ	N	velar nasal	ri ng	k
p	p	voiceless bilabial plosive	pil	p

IPA	X-SAMPA	Description	Example	Viseme
r	r	alveolar trill	ris	r
S	S	voiceless alveolar fricative	sil	S
Ç	s\	voiceless alveolo-p alatal fricative	tj ock	J
\$	s`	voiceless retroflex fricative	fo rs, sch lager	S
t	t	voiceless alveolar plosive	t al	t
t	t`	voiceless retroflex plosive	hjort	t
V	V	voiced labiodental fricative	v år	f
W	w	labial-velar approximant	a u la, air w ays	u
Ŋ	x\	voiceless palatal-v elar fricative	sj uk	k
Vowels				
Ø	2	close-mid front rounded vowel	f ö ll, f ö rr	0
Ø	2:	long close-mid front rounded vowel	f ö l, n ö t, f ö r	0
Θ	8	close-mid central rounded vowel	buss, full	0

IPA	X-SAMPA	Description	Example	Viseme
ə	@	mid central vowel	pojk e n	@
u ː	} :	long close central rounded vowel	hus, ful	u
a	a	open front unrounded vowel	hall, matt	a
æ	{	near-open front unrounded vowel	herr	a
a:	A:	long open back unrounded vowel	hal, mat	a
e:	e:	long close-mid front unrounded vowel	vet, hel	е
3	E	open-mid front unrounded vowel	vett, r ä tt, hetta, h ä ll	Е
ε:	E:	long open-mid front unrounded vowel	s ä l, h ä l, h ä r	E:
i:	i:	long close front unrounded vowel	vit, sil	i:
I	1	near-close near- front unrounded vowel	vitt, sill	i
0:	o:	long close-mid back rounded vowel	hål, mål	0

IPA	X-SAMPA	Description	Example	Viseme
Э	0	open-mid back rounded vowel	h å ll, m o ll	0
u:	u:	long close back rounded vowel	s o l, b o t	u
ឋ	U	near-close near- back rounded vowel	b o tt	u
У	У	close front rounded vowel	b y tt	u
y:	y:	long close front rounded vowel	s y l, s y l	u
Addition	al Symbols			
ı	п	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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

IPA	X-SAMPA	Description	Example	Viseme	
Consonants					
ſ	4	alveolar flap	durum	t	

IPA	X-SAMPA	Description	Example	Viseme
ĵ	4_0_r	voiceless fricated alveolar flap	bi r	t
Į	4_r	fricated alveolar flap	raf	t
b	b	voiced bilabial plosive	raf	р
С	С	voiceless palatal plosive	k edi	k
d	d	voiced alveolar plosive	d e d e	t
dz	dZ	voiced postalveo lar affricate	c am	S
f	f	voiceless labiodent al fricative	f are	f
g	g	voiced velar plosiv	g alibi	k
h	h	voiceless glottal fricative	h asta	k
j	j	palatal approxima nt	y at	i
Ĵ	٦/	voiced palatal plosive	g enç	J
k	k	voiceless velar plosive	a k ıl	k
l	l	alveolar lateral approximant	lale	t

IPA	X-SAMPA	Description	Example	Viseme
†	5	velarized alveolar lateral approxima nt	labirent	t
m	m	bilabial nasal	m aaş	p
n	n	alveolar nasal	anı	t
p	р	voiceless bilabial plosive	ip	p
S	S	voiceless alveolar fricative	s es	S
ſ	S	voiceless postalveo lar fricative	aşı	S
t	t	voiceless alveolar plosive	ü t ü	t
Ą	tS	voiceless postalveo lar affricate	ç aba	S
V	V	voiced labiodental fricative	ekvator, kahveci, akvaryum, isveçli, teşviki, cetvel	f
Z	Z	voiced alveolar fricative	v er	S
3	Z	voiced postalveo lar fricative	azık	S
Vowels				
Ø	2	close-mid front rounded vowel	g ö l	0

IPA	X-SAMPA	Description	Example	Viseme
œ	9	open-mid front rounded vowel	banliy ö	0
а	a	open front unrounded vowel	k a l	a
a:	a:	long open front unrounded vowel	d ava cı	a
æ	{	near-open front unrounded vowel	özlem, güvenlik, gürel, somersault	a
е	е	close-mid front unrounded vowel	k e çi	е
3	E	open-mid front unrounded vowel	ded e	Е
i	i	close front unrounded vowel	bir	i
i:	i:	long close front unrounded vowel	izah	i
I	1	near-close near- front unrounded vowel	keçi	i
ш	М	close back unrounded vowel	kıl	i
0	o	close-mid back rounded vowel	k o l	О
o:	o:	long close-mid back rounded vowel	d o lar	0

IPA	X-SAMPA	Description	Example	Viseme
u	u	close back rounded vowel	durum	u
u:	u:	long close back rounded vowel	r u hum	u
ឋ	U	near-close near- back rounded vowel	dol u	u
У	У	close front rounded vowel	g ü venlik	u
Y	Υ	near-close near- front rounded vowel	aşı	u
Addition	al Symbols			
1	п	primary stress	Ala ba ma	
1	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

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.

Phoneme/Viseme Table

IPA	X-SAMPA	Description	Example	Viseme
Consonai	nts			

IPA	X-SAMPA	Description	Example	Viseme
b	b	voiced bilabial plosive	b aban	p
d	d	voiced alveolar plosive	d eg	t
dz	dZ	voiced postalveo lar affricate	gare j	S
ð	D	voiced dental fricative	deu dd eg	Т
f	f	voiceless labiodent al fricative	ff acs	f
g	g	voiced velar plosive	g adael	k
h	h	voiceless glottal fricative	h aearn	k
j	j	palatal approxima nt	astudio	i
k	k	voiceless velar plosive	c ant	k
l	l	alveolar lateral approximant	lan	t
4	K	voiceless alveolar lateral fricative	llan	t
m	m	bilabial nasal	m ae	p
m	m_0	voiceless bilabial nasal	y mh en	р

IPA	X-SAMPA	Description	Example	Viseme
n	n	alveolar nasal	n aw	t
ů	n_0	voiceless alveolar nasal	a nh awster	t
ŋ	N	velar nasal	argyfw ng	k
ŋ°	N_0	voiceless velar nasal	a ng henion	k
p	p	voiceless bilabial plosive	p ump	p
r	r	alveolar trill	rh oi	r
ţ	r_0	voiceless alveolar trill	garw	r
S	S	voiceless alveolar fricative	s aith	S
ſ	S	voiceless postalveo lar fricative	si awns	S
t	t	voiceless alveolar plosive	t egan	t
Ą	tS	voiceless postalveo lar affricate	cy ts ain	S
θ	Т	voiceless dental fricative	aber th	Т
V	V	voiced labiodental fricative	praw f	f
W	w	labial-velar approximant	rhag w eld	u

IPA	X-SAMPA	Description	Example	Viseme
Χ	X	voiceless uvular fricative	chwech	k
Z	Z	voiced alveolar fricative	aid s	S
3	Z	voiced postalveo lar fricative	rou ge	S
Vowels				
ə	@	mid central vowel	y chwaneg a	@
a	a	open front unrounded vowel	a cen	a
ai	ai	diphthong	d au	a
au	au	diphthong	aw dur	a
a:	A:	long open back unrounded vowel	m a b	a
a: i	A:1	diphthong	ae lod	a
e:	e:	long close-mid front unrounded vowel	peth	е
3	Е	open-mid front unrounded vowel	pedwar	Е
εί	Ei	diphthong	b ei c	Е
i:	it	long close front unrounded vowel	tri	i

IPA	X-SAMPA	Description	Example	Viseme
I	I	near-close near- front unrounded vowel	miliwn	i
i u	1u	diphthong	unigr yw	i
0:	o:	long close-mid back rounded vowel	o ddi	0
Э	0	open-mid back rounded vowel	o ddieithr	0
oi	Oi	diphthong	tr oi	0
ou	Ou	diphthong	r ow nd	0
u:	u:	long close back rounded vowel	cwch	u
ឋ	U	near-close near- back rounded vowel	ac w stig	u
ชเ่	Ui	diphthong	wy th	u
Additiona	al Symbols			
1	п	primary stress	Ala ba ma	
ī	%	secondary stress	A labama	
		syllable boundary	A.la.ba.ma	

Amazon Polly voice engines

Amazon Polly has four voice engines that convert input text into life-like speech. These include: **Generative**, **Long-form**, **Neural**, and **Standard**. To use an Amazon Polly voice, select an engine and a speech synthesis API operation. Then provide input text for the engine to synthesize, and select an audio output format. Given these inputs, Amazon Polly synthesizes the provided text into a high-quality speech audio stream.

The following sections include details about the voice engines offered by Amazon Polly.

Topics

- · Generative voices
- Long-form voices
- Neural voices
- Standard voices
- Choosing a voice engine

Generative voices

Amazon Polly's **generative** text-to-speech (TTS) engine offers the most human-like, emotionally engaged, and adaptive conversational voices available for the use via the Amazon Polly console.

The **Generative engine** is the largest Amazon Polly TTS model to-date. It deploys a billion-parameter transformer that converts raw text into speech codes, followed by a convolution-based decoder that converts these speech codes into waveforms in an incremental, streamable manner. This method shows the widely-reported emergent abilities of Large Language Models (LLMs) when trained on increasing volumes of publicly available and proprietary data comprising a variety of voices, languages, and styles.

The Generative engine creates synthetic speech which is emotionally engaged, assertive, and highly colloquial in a way that is remarkably similar to a human voice. You can use these voices as a knowledgeable customer assistant, a virtual trainer, or an advertiser with a near-human synthetic speech.

Generative engine 193



Note

The state-of-the-art technology underlying these voices falls within the paradigm of generative AI for language and voice modelling. A side effect of the technology is that any updates to the training data and the model could result in slight variations to the way the voices sound, even in case when their overall quality improves with model updates. This could have an impact on use cases with different content parts synthesized over a long time period – for example, a season of podcasts.

Available generative voices

Amazon Polly currently offers 20 voices in a generative variant. These generative voices are also available in a conversational NTTS variant.

	Language	Language code	Name/ID	Gender
1	English (Australian)	en-AU	Olivia	Female
2	English (Indian)	en-IN	Kajal	Female
3	English (UK)	en-GB	Amy	Female
4	English (US)	en-US	Danielle	Female
5	English (US)	en-US	Joanna	Female
6	English (US)	en-US	Matthew	Male
7	English (US)	en-US	Ruth	Female
8	English (US)	en-US	Stephen	Male
9	English (South African)	en-ZA	Ayanda	Female
10	French (France)	fr-FR	Léa	Female
11	French (France)	fr-FR	Rémi	Male

Available generative voices 194

	Language	Language code	Name/ID	Gender
12	Spanish (Spain)	es-ES	Lucia	Female
13	Spanish (Spain)	es-ES	Sergio	Male
14	Spanish (Mexican)	es-MX	Mía	Female
15	Spanish (Mexican)	es-MX	Andrés	Male
16	Spanish (US)	es-US	Lupe	Female
17	Spanish (US)	es-US	Pedro	Male
18	German (Germany)	de-DE	Vicki	Female
19	German (Germany)	de-DE	Daniel	Male
20	Italian (Italy)	it-IT	Bianca	Female



Note

Generative voices cost is specified on the Amazon Polly pricing information page.

Feature and region compatibility

Amazon Polly generative voices are available in the following regions:

- US East (N. Virginia): us-east-1
- Europe (Frankfurt): eu-central-1
- US West (Oregon): us-west-2
- Other Regions are not available

The following features are supported for generative voices:

- Real-time and asynchronous speech synthesis operations.
- Newscaster speaking style is not supported in the Generative engine.
- · Many (but not all) SSML tags are supported by Amazon Polly. For more information about NTTSsupported SSML tags, see Supported SSML tags
- 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 generative voices is 24 kHz. Amazon Polly supports MP3, OGG (Vorbis), and raw PCM audio stream formats.

Support for generating speech marks is currently not available.



Note

In the unlikely event of model hallucination, (and with the Generative engine's model behavior of rendering the speech token by token) an imposed emergency stop mechanism is in place. The built-in mechanism stops the model from rendering speech any further. This safety feature is based on data analysis where the model has the potential to hallucinate, usually at the end of the sentence.

There could be cases where the model thinks it is going to hallucinate and then might end up cutting a word during a generation step, thus rendering half the word. This could potentially generate inappropriate results.

Long-form voices

Amazon Polly has a **Long-form engine** that produces human-like, highly expressive, and emotionally adept voices. Long-form voices are designed to captivate listeners' attention for longer content, such as news articles, training materials, or marketing videos.

Amazon Polly Long-form voices are developed with a cutting-edge deep learning TTS technology. The model learns to replicate phonemes, prosody, intonation, and other phonetic and acoustic aspects of human language, resulting in a highly natural speech output.

196 Long-form engine

The Long-form engine uses text embeddings to interpret the meaning of a text. Using text embeddings, the Long-form engine can generate the correct emphasis, pauses, and tone of a natural voice. The result is a voice that combines the complete range of emotional elements present in human communication. This includes mimicking surprisal or differentiating dialogue from narration. Together, this creates a premium speech product that sounds like a live human being.



Note

The state-of-the-art technology underlying these voices falls within the paradigm of generative AI for language and voice modelling. A side effect of the technology is that any updates to the training data and the model could result in a slight variations to the way the voices sound, even in case when their overall quality improves with model updates. This could have an impact on use cases with different content parts synthesized over a long time period – for example, a season of podcasts.

Available long-form voices

Amazon Polly currently offers four en-US and two es-ES long-form voices. Both languages have female and male voices available. The English long-form voices Daniel, Gregory, and Ruth are also available in a conversational NTTS variant.

	Language	Language code	Name/ID	Gender
1	English (US)	en-US	Danielle	Female
			Gregory	Male
			Ruth	Female
			Patrick	Male
2	Spanish (Spain)	es-ES	Alba	Female
			Raúl	Male

Available long-form voices 197

Feature and region compatibility

Amazon Polly long-form voices are available in the following regions:

- US East (N. Virginia): us-east-1
- Other regions not available

The Amazon Polly Long-form engine supports the following features:

- Real-time and asynchronous speech synthesis operations.
- All speech marks.
- Many (but not all) SSML tags are supported by Amazon Polly. For more information about NTTSsupported SSML tags, see Supported SSML tags
- 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, long-form, and neural voices are: 8 kHz, 16 kHz, 22kHz, or 24 kHz. The default for standard voices is 22 kHz. The default for long-form and neural voices is 24 kHz. Amazon Polly supports MP3, OGG (Vorbis), and raw PCM audio stream formats.



Note

Long-form voices cost is specified on the Amazon Polly pricing information page.

Neural voices

Amazon Polly has a **Neural text-to-speech (NTTS) engine** that can produce even higher quality voices than its standard voices. Standard TTS voices use concatenative synthesis. The standard engine concatenates 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 NTTS engine 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. (Spectograms are snapshots of the energy levels in different frequency bands.)

• A vocoder — that converts spectrograms into a nearly 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 datasets used to build general-purpose concatenative-synthesis systems, this sequence-to-sequence approach will yield higher-quality, more natural-sounding voices.

Available neural voices

Neural voices are available in 35 languages and language variants. The following table lists the voices.

	Language and language variants	Language code	Name/ID	Gender
1	Arabic (Gulf)	ar-AE	Hala	Female
			Zayd	Male
2	Belgian Dutch (Flemish)	nl-BE	Lisa	Female
3	Catalan	ca-ES	Arlet	Female
4	Czech	cs-CZ	Jitka	Female
5	Chinese (Cantonese)	yue-CN	Hiujin	Female
6	Chinese (Mandarin)	cmn-CN	Zhiyu	Female

	Language and language variants	Language code	Name/ID	Gender
7	Danish	da-DK	Sofie	Female
8	Dutch	nl-NL	Laura	Female
9	English (Australian)	en-AU	Olivia	Female
10	English (British)	en-GB	Amy* Emma Brian Arthur	Female Female Male Male
11	English (Indian)	en-IN	Kajal	Female
12	English (Irish)	en-IE	Niamh	Female
13	English (New Zealand)	en-NZ	Aria	Female
14	English (South African)	en-ZA	Ayanda	Female

	Language and language variants	Language code	Name/ID	Gender
15	English (US)	en-US	Danielle	Female
			Gregory	Male
			lvy	Female (child)
			Joanna*	Female
			Kendra	Female
			Kimberly	Female
			Salli	Female
			Joey	Male
			Justin	Male (child)
			Kevin	Male (child)
			Matthew*	Male
			Ruth	Female
			Stephen	Male
16	Finnish	fi-FI	Suvi	Female
17	French (Belgian)	fr-BE	Isabelle	Female
18	French	fr-CA	Gabrielle	Female
	(Canadian)		Liam	Male
19	French	fr-FR	Léa	Female
			Rémi	Male

	Language and language variants	Language code	Name/ID	Gender
20	German	de-DE	Vicki	Female
			Daniel	Male
21	German (Austrian)	de-AT	Hannah	Female
22	German (Swiss)	de-CH	Sabrina	Female
23	Hindi	hi-IN	Kajal	Female
24	Italian	it-IT	Bianca	Female
			Adriano	Male
25	Japanese	ja-JP	Takumi	Male
			Kazuha	Female
			Tomoko	Female
26	Korean	ko-KR	Seoyeon	Female
27	Norwegian	nb-NO	lda	Female
28	Polish	pl-PL	Ola	Female
29	Portuguese	pt-BR	Camila	Female
	(Brazilian)		Vitória/Vitoria	Female
			Thiago	Male
30	Portuguese (European)	pt-PT	Inês/Ines	Female

	Language and language variants	Language code	Name/ID	Gender
31	Spanish (Spain)	es-ES	Lucia Sergio	Female Male
32	Spanish	es-MX	Mia	Female
32	(Mexican)	CSTIA	Andrés	Male
33	Spanish (US)	es-US	Lupe*	Female
			Pedro	Male
34	Swedish	sv-SE	Elin	Female
35	Turkish	tr-TR	Burcu	Female

^{*}The Amy, Joanna, Lupe, and Matthew voices can be used with the Newscaster speaking style. For more information, see Applying the newscaster voice.

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
- Africa (Cape Town): af-south-1
- Asia Pacific (Tokyo): ap-northeast-1
- Asia Pacific (Seoul): ap-northeast-2
- Asia Pacific (Osaka): ap-northeast-3
- Asia Pacific (Mumbai): ap-south-1
- Asia Pacific (Singapore): ap-southeast-1

- Asia Pacific (Sydney): ap-southeast-2
- Canada (Central): ca-central-1
- Europe (Frankfurt): eu-central-1
- Europe (Ireland): eu-west-1
- Europe (London): eu-west-2
- Europe (Paris): eu-west-3
- AWS GovCloud (US-West): us-gov-west-1

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 <u>Applying the</u> newscaster voice.
- All speech marks.
- Many (but not all) of the SSML tags that are supported by Amazon Polly. For more information about NTTS-supported SSML tags, see Supported Tags.

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.

Standard voices

Amazon Polly has a **standard** engine that use concatenative synthesis. The standard engine concatenates phonemes of recorded speech, producing very natural-sounding synthesized speech.

Available Standard voices

Amazon Polly currently offers 40 female and 20 male standard voices in 29 language and language variants.

Standard engine 204

	Language	Language code	Name/ID	Gender
1	Arabic	arb	Zeina	Female
2	Chinese (Mandarin)	cmn-CN	Zhiyu	Female
3	Danish	da-DK	Naja Mads	Female Male
4	Dutch	nl-NL	Lotte Ruben	Female Male
5	English (Australian)	en-AU	Nicole Russell	Female Male
6	English (British)	en-GB	Amy Emma Brian	Female Female Male
7	English (Indian)	en-IN	Aditi Raveena	Female Female
8	English (US)	en-US	Ivy Joanna Kendra Kimberly Salli Joey Kevin	Female Female Female Female Male Male

Available Standard voices 205

	Language	Language code	Name/ID	Gender
9	English (Welsh)	en-GB-WLS	Geraint	Male
10	French	fr-FR	Céline/Celine	Female
			Léa	Female
			Mathieu	Male
11	French (Canadian)	fr-CA	Chantal	Female
12	German	de-DE	Marlene	Female
			Vicki	Female
			Hans	Male
13	Hindi	hi-IN	Aditi	Female
14	Icelandic	is-IS	Dóra/Dora	Female
			Karl	Male
15	Italian	it-IT	Carla	Female
			Bianca	Female
			Giorgio	Male
16	Japanese	ja-JP	Mizuki	Female
			Takumi	Male
17	Korean	ko-KR	Seoyeon	Female
18	Norwegian	nb-NO	Liv	Female

Available Standard voices 206

	Language	Language code	Name/ID	Gender
19	Polish	pl-PL	Ewa	Female
			Maja	Female
			Jacek	Male
			Jan	Male
20	Portuguese	pt-BR	Camila	Female
	(Brazilian)		Vitória/Vitoria	Female
			Ricardo	Male
21	Portuguese (European)	pt-PT	Inês/Ines	Female
			Cristiano	Male
22	Romanian	ro-RO	Carmen	Female
23	Russian	ru-RU	Tatyana	Female
			Maxim	Male
24	Spanish (Spain)	es-ES	Conchita	Female
			Lucia	Female
			Enrique	Male
25	Spanish (Mexican)	es-MX	Mia	Female
26	Spanish (US)	es-US	Lupe	Female
			Penélope/	Female
			Penelope	Male
			Miguel	

Available Standard voices 207

	Language	Language code	Name/ID	Gender
27	Swedish	sv-SE	Astrid	Female
28	Turkish	tr-TR	Filiz	Male
29	Welsh	cy-GB	Gwyneth	Female

Feature and region compatibility

Amazon Polly standard voices are available in the following Amazon Polly regions:

- US East (N. Virginia): us-east-1
- US East (Ohio): us-east-2
- US West (N. California): us-west-1
- US West (Oregon): us-west-2
- Africa (Cape Town): af-south-1
- Asia Pacific (Hong Kong): ap-east-1
- Asia Pacific (Tokyo): ap-northeast-1
- Asia Pacific (Seoul): ap-northeast-2
- Asia Pacific (Osaka): ap-northeast-3
- Asia Pacific (Mumbai): ap-south-1
- Asia Pacific (Singapore): ap-southeast-1
- Asia Pacific (Sydney): ap-southeast-2
- China (Ningxia): cn-northwest-1;
- Canada (Central): ca-central-1
- Europe (Frankfurt): eu-central-1
- Europe (Ireland): eu-west-1
- Europe (London): eu-west-2
- Europe (Paris): eu-west-3
- Europe (Stockholm): eu-north-1
- Middle East (Bahrain): me-south-1
- South America (São Paulo): sa-east-1

AWS GovCloud (US-West): us-gov-west-1

Endpoints and protocols for these Regions are identical to those used for Neural voices. For more information, see Amazon Polly endpoints and quotas.

The Amazon Polly standard engine supports the following features (TBD):

- Real-time and asynchronous speech synthesis operations.
- All speech marks.
- Many (but not all) SSML tags are supported by Amazon Polly. For more information about NTTSsupported SSML tags, see Supported SSML tags.
- You can choose from various sampling rates to optimize the bandwidth and audio quality for your application. The default sampling rates for standard voices are 22 kHz. Amazon Polly supports MP3, OGG (Vorbis), and raw PCM audio stream formats.



Note

Standard voices cost is specified on the Amazon Polly pricing information page.

Choosing a voice engine

You can access Amazon Polly voices through the Amazon Polly console or AWS CLI.

To choose a voice engine on the console

- Open the Amazon Polly console at https://console.aws.amazon.com/polly/.
- 2. From the Amazon Polly console, choose the desired voice engine.
- 3. Choose the desired voice from the voice drop-down menu.
- Generate TTS audio with text of your choice. 4.

To choose a voice engine in the AWS CLI, specify the Engine and VoiceId in the SyntheszieSpeech or StartSpeechSynthesisTask API operations. For some examples, see the quick-start code samples and the Python examples.

209 Choosing a voice engine

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.

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

Topics

- Speech mark types
- Visemes and Amazon Polly
- Speech mark output
- Requesting speech marks
- Speech marks without SSML example
- Speech marks with SSML example

Speech mark types

You request speech marks using the <u>SpeechMarkTypes</u> option for either the <u>SynthesizeSpeech</u> or <u>StartSpeechSynthesisTask</u> 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.

Speech mark types 210

- 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.

ssml – Describes a <mark> element from the SSML input text. For more information, see
 Generating speech from SSML documents.

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 Languages in Amazon Polly.

IPA	X-SAMPA	Description	Example	Viseme	
Consonants					
b	b	Voiced bilabial plosive	b ed	р	
d	d	Voiced alveolar plosive	d ig	t	
d̃ʒ	dZ	Voiced postalveo lar affricate	jump	S	

Visemes and Amazon Polly 211

IPA	X-SAMPA	Description	Example	Viseme
ð	D	Voiced dental fricative	th en	Т
f	f	Voiceless labiodent al fricative	five	f
g	g	Voiced velar plosive	g ame	k
h	h	Voiceless glottal fricative	house	k

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":

Speech mark output 212

```
{"time":373,"type":"word","start":5,"end":8,"value":"had"}
```

The described word ("had") begins 373 milliseconds after the audio stream begins, and starts at byte 5 and ends at byte 8 of the input text.



Note

This metadata is for the Joanna voice-id. If you use another voice with the same input text, the metadata might differ.

Requesting speech marks

You can use the console or the synthesize-speech command to request speech marks from Amazon Polly. You can then view the metadata or save it to a file.

Console

To generate speech marks on the console

- 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.
- Turn on **SSML** to use SSML.
- 4. Type or paste your text into the input box.
- 5. For **Language**, choose the language of your text.
- 6. For **Voice**, choose the voice you want to use.
- 7. To change text pronunciation, expand **Additional settings**, turn on **Customize pronunciation**, and for **Apply lexicon**, choose the desired lexicon.
- To verify the speech, choose **Listen**.
- Turn on **Speech file format settings**.



Note

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

Requesting speech marks 213

- 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.

12. Choose **Download**.

AWS CLI

In addition to 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.

```
--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.

```
--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.

```
--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

Requesting speech marks 214

Specify the output file to which the metadata is written.

```
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.

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

Speech marks without SSML example

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.

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

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":6,"type":"viseme","value":"p"}
```

```
{"time":73, "type": "viseme", "value": "E"}
{"time":180, "type":"viseme", "value":"r"}
{"time":292, "type": "viseme", "value": "i"}
{"time":373,"type":"word","start":5,"end":8,"value":"had"}
{"time":373, "type": "viseme", "value": "k"}
{"time":460, "type":"viseme", "value":"a"}
{"time":521, "type": "viseme", "value": "t"}
{"time":604,"type":"word","start":9,"end":10,"value":"a"}
{"time":604, "type":"viseme", "value":"@"}
{"time":643,"type":"word","start":11,"end":17,"value":"little"}
{"time":643, "type": "viseme", "value": "t"}
{"time":739, "type":"viseme", "value":"i"}
{"time":769, "type": "viseme", "value": "t"}
{"time":799, "type": "viseme", "value": "t"}
{"time":882,"type":"word","start":18,"end":22,"value":"lamb"}
{"time":882,"type":"viseme","value":"t"}
{"time":964, "type": "viseme", "value": "a"}
{"time":1082, "type": "viseme", "value": "p"}
```

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

- The sentence "Mary had a little lamb."
- 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.

Speech marks with SSML example

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 don't 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"}
```

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.

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 Synthesizing speech with Amazon Polly example 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.
```

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.

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

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

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
- Using SSML on the console
- Using SSML with the Synthesize-Speech command
- Synthesizing an SSML-enhanced document
- Supported SSML tags

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 as follows:

Clfa displace code "ation mark (double quotation mark) &alanppersand &alanppersand

e or

Reserved characters 219

Cha dalquene

code

single

quotation

mark

& Hetss

than

sign

& greater

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&apos;re using the lawyer at Peabody &amp; 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. Additionallty, 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

Reserved characters 220

 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'
```

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"
```

Using SSML on the console

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 (Synthesizing Speech).

For more information on SSML, with examples, see Supported SSML tags.

Using SSML on the console 221

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/.

- If it isn't already displayed, choose the **Text-to-Speech** tab. 2.
- Turn on SSML. 3.
- 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
     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
     <amazon:effect name="whispered">"Nice job,"</amazon:effect> in a
    whisper.
</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.
- _{W3C} 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.
- To listen to the speech, choose **Listen**. 6.
- To save the speech file, choose **Download**. If you want to save it in a different format, expand 7. Additional settings, turn on Speech file format settings and choose the format that you want, then choose **Download**.

Using SSML on the console 222

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:

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 Specifying another language for specific words.

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.

Developer Guide Amazon Polly



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
```

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

Supported SSML tags

Amazon Polly supports the following SSML tags:

Action	SSML tag	Availabil ity with neural voices	Availabil ity with long-form voices	Availabil ity with generative voices
Adding a pause	 	Full availability	Full availability	Full availability
Emphasizing words	<emphasis></emphasis>	Not available	Not available	Not available
Specifying another language for specific words	<lang></lang>	Full availability	Full availability	Full availability

Supported SSML tags 225

Action	SSML tag	Availabil ity with neural voices	Availabil ity with long-form voices	Availabil ity with generative voices
Placing a custom tag in your text	<mark></mark>	Full availability	Full availability	Full availability
Adding a pause between paragraphs		Full availability	Full availability	Full availability
Using phonetic pronunciation	<pre><phoneme></phoneme></pre>	Full availability	Full availability	Not available
Controlling volume, speaking rate, and pitch	<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>	Partial availability	Partial availability	Not available
Setting a maximum duration for synthesized speech	<pre><pre><pre><pre><pre><pre>duration></pre></pre></pre></pre></pre></pre>	Not available	Not available	Not available
Adding a pause between sentences	<s></s>	Full availability	Full availability	Full availability
Controlling how special types of words are spoken	<say-as></say-as>	Partial availability	Full availability	Full availability
Identifying SSML-enhanced text	<speak></speak>	Full availability	Full availability	Full availability
Pronouncing acronyms and abbreviations		Full availability	Full availability	Full availability
Improving pronunciation by specifying parts of speech	<w></w>	Full availability	Full availability	Full availability

Supported SSML tags 226

Action	SSML tag	Availabil ity with neural voices	Availabil ity with long-form voices	Availabil ity with generative voices
Adding the sound of breathing	<amazon:auto-breaths></amazon:auto-breaths>	Not available	Not available	Not available
Newscaster speaking style	<amazon:domain name="news"></amazon:domain 	Select neural voices only	Not available	Not available
Adding dynamic range compression	<amazon:effect name="drc"></amazon:effect 	Full availability	Full availability	Not available
Speaking softly	<amazon:effect phonation="soft"></amazon:effect>	Not available	Not available	Not available
Controlling timbre	<amazon:effect ct-length="" vocal-tra=""></amazon:effect>	Not available	Not available	Not available
Whispering	<amazon:effect name="whispered"></amazon:effect 	Not available	Not available	Not available



Note

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

Identifying SSML-enhanced text

<speak>

This tag is supported by generative, long-form, neural, and standard TTS formats.

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.

<speak>Mary had a little lamb.</speak>

Adding a pause

<bre><break>

This tag is supported by generative, long-form, neural, and standard TTS formats.

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:

Adding a pause 228

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

- (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:

<speak>I already told you I really like that person.</speak>

Emphasizing words 229

Specifying another language for specific words

<lang>

This tag is supported by generative, long-form, 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 in Amazon Polly.

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:

```
<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:

```
<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:

```
<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:

```
<speak>
    Mi piace <lang xml:lang="en-US">Bruce Springsteen.</lang>
</speak>
```

This tag can also be used as a substitute for the optional <u>DefaultLangCode</u> 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 generative, long-form, 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

>

This tag is supported by generative, long-form, neural, and standard TTS formats.

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

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

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

Using phonetic pronunciation

<phoneme>

This tag is supported by long-form, 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 <u>Languages in</u> Amazon Polly

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.

Using phonetic pronunciation 232

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>
```

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>
```

Japanese uses Yomigana and Pronunciation Kana.

Yomigana

```
<speak>
    ###<phoneme alphabet="x-amazon-yomigana" ph="####">##</phoneme>###

###<phoneme alphabet="x-amazon-yomigana" ph="####">##</phoneme>###

###<phoneme alphabet="x-amazon-yomigana" ph="Hirokazu">##</phoneme>###

</speak>
```

Pronunciation Kana

```
<speak>
    ###<phoneme alphabet="x-amazon-pron-kana" ph="##"#">##</phoneme>###
</speak>
```

Using phonetic pronunciation 233

Controlling volume, speaking rate, and pitch

osody>

Prosody tag attributes are fully supported by the standard TTS voices. Neural and long-form 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:

- 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 condy volume="loud">increase the volume
    for a specific speech.

</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.
```

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 compressed on compresse
```

• 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:

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

Or you could set it this way:

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

Setting a maximum duration for synthesized speech

orosody 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 control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when it is synthesized, use the control how long you want a speech to take when you want a sp

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 rosody 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:

```
orosody amazon:max-duration="time duration">
```

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:

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

duration> tag, and you can use multiple cody amazon:max-duration> tags within your text.

For example:

```
<speak>
    con:max-duration="2400ms">
       Human speech is a powerful way to communicate.
    <break strength="strong"/>
    consody amazon:max-duration="5100ms">
       Even a simple 'Hello' can convey a lot of information depending on the pitch,
intonation, and tempo.
    <break strength="strong"/>
    consody 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.
    </speak>
```

Using the 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 oprosody amazon:max-duration> tag and in how it works with other SSML tags:

- You can't nest con:max-duration tags. If you put one con: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:

• You can't use the condy tags with the rate attribute within a condy amazon:maxduration tag. This is because both affect the speed at which text is spoken.

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

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:

```
</speak>
```

Adding a pause between sentences

<<u>5</u>>

This tag is supported by generative, long-form, 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

'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>

The <say-as> tag is supported by generative, long-form, neural, and standard TTS engines.

Note, however, that if Amazon Polly is using a neural voice and encounters the <say-as> tag with the characters option at runtime, the affected sentence will be synthesized using the related standard voice. However, the affected sentence will still be billed as if it uses a 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 for neural voices. If you're using a neural voice and 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 a 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 1/2inch, or by just a unit, as in 1meter.
- 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 don't 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.
- 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":

```
<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 generative, long-form, 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.

```
<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 generative, long-form, 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:

```
<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 preposition.
- 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:

```
<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:

```
<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:

```
<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 breaths
 <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"/> 
volume="loud">
    Wow! <amazon:breath duration="long" volume="loud"/> </prosody> That was quite
fast. <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.

<speak>

<amazon:auto-breaths>Amazon Polly is a service that turns text into lifelike

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).

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

• 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:

```
<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, has 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

an iceberg off the Grand Banks of Newfoundland. By wireless telegraphy she sent out signals of distress,

Newscaster speaking style 250

```
and several liners were near enough to catch and respond to the call.
</amazon:domain>
</speak>
```

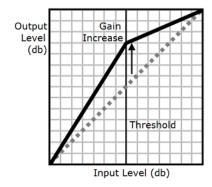
Adding dynamic range compression

<amazon:effect name="drc">

This tag is supported by long-form, 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>

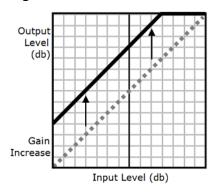


Note

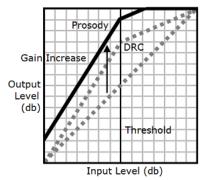
When you use "drc" in the amazon: effect syntax, it is case-sensitive.

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:

```
<speak>
    oody 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

<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:

```
<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>

Speaking softly 253

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:

```
    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

Controlling timbre 254

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.

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>
```

Whispering 255

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:

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.

Whispering 256

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.



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

- Using multiple lexicons
- Uploading a lexicon
- Applying lexicons (Synthesizing Speech)
- Filtering the lexicon list on the console
- Downloading lexicons on the console

Deleting a lexicon

Using 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>
    <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

```
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."

Using multiple lexicons 258

Example - Applying LexB before LexA

```
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 <u>Applying lexicons</u> (Synthesizing Speech).

Uploading a lexicon

The lexicons you use must conform to the Pronunciation Lexicon Specification (PLS) W3C recommendation. For more information, see <u>Pronunciation Lexicon Specification (PLS) Version 1.0</u> on the W3C website.

Console - Lexicons tab

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

- 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 **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.

Console - TTS tab

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.

AWS CLI - one lexeme

With Amazon Polly, you can use <u>PutLexicon</u> 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 <u>SynthesizeSpeech</u> request that you want to apply before the service starts synthesizing the text. For more information, see <u>Managing lexicons</u>.

Consider the following W3C PLS-compliant lexicon.

```
<?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>
</lexicon>
```

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:

- Save the lexicon as example.pls.
- Run the put-lexicon AWS CLI command to store the lexicon (with the name w3c), in the us-east-2 region.

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

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

```
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.

```
<?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>pecan</grapheme>
        <phoneme>p##k##n</phoneme>
        </lexeme>
</lexeme>
</lexeme></lexecon>
```

Follow the same steps to test this lexicon. Make sure you specify input text that has the word "pecan" (for example, "Pecan pie is delicious").

See the following resources for additional code samples for the PutLexicon API operation:

Java Sample: PutLexicon

Python (Boto3) Sample: PutLexicon

AWS CLI - multiple lexemes

With Amazon Polly, you can use <u>PutLexicon</u> 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 <u>SynthesizeSpeech</u> request that you want to apply before the service starts synthesizing the text. For more information, see <u>Managing lexicons</u>.

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 version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"</pre>
      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 first <lexeme> element defines an alias (World Wide Web Consortium).
- 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:

- 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.

```
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.

```
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.

```
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.

See the following resources for additional code samples for the PutLexicon API operation:

Java Sample: PutLexicon

• Python (Boto3) Sample: PutLexicon

Applying lexicons (Synthesizing Speech)

The lexicons you use must conform to the Pronunciation Lexicon Specification (PLS) W3C recommendation. For more information, see <u>Pronunciation Lexicon Specification (PLS) Version 1.0</u> on the W3C website.

Console

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 <u>Uploading a lexicon</u>. 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 <u>Using SSML on the console</u> 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"/>
In the middle of the 10/3/2014 W3C meeting
he shouted, "Score!" quite loudly.</speak>
```

- 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 **Uploading a lexicon**.

- 6. To listen to the speech immediately, choose **Listen**.
- 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.

AWS CLI

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 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"</pre>
```

```
alphabet="ipa" xml:lang="en-US">
    <lexeme>
        <grapheme>W3C</grapheme>
        <alias>World Wide Web Consortium</alias>
        </lexeme>
        </lexicon>
```

• Lexicon 2: w3cAlternate.pls

```
<?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.

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

• Upload the w3cAlternate.pls lexicon on the service as w3cAlternate.

```
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.

```
aws polly synthesize-speech \
--text 'PLS is a W3C recommendation' \
--voice-id Joanna \
--output-format mp3 \
--lexicon-names '["w3c","w3cAlternative"]' \
speech.mp3
```

4. Test the resulting speech.mp3. It should read as follows:

```
PLS is a World Wide Web Consortium recommendation
```

Filtering the lexicon list on the console

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

Console

To filter the lexicons listed by language

- 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.

AWS CLI

Amazon Polly provides the <u>ListLexicons</u> 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": [
        {
            "Attributes": {
                "LanguageCode": "en-US",
                "LastModified": 1474222543.989,
                "Alphabet": "ipa",
                "LexemesCount": 1,
                "LexiconArn": "arn:aws:polly:aws-region:account-id:lexicon/w3c",
                "Size": 495
            },
            "Name": "w3c"
        },
        {
            "Attributes": {
                "LanguageCode": "en-US",
                "LastModified": 1473099290.858,
                "Alphabet": "ipa",
                "LexemesCount": 1,
                "LexiconArn": "arn:aws:polly:aws-region:account-id:lexicon/tomato",
                "Size": 645
            },
            "Name": "tomato"
        }
    ]
}
```

The following resources contain additional information for the ListLexicons operation:

Java Sample: <u>ListLexicons</u>

Python (Boto3) Sample: ListLexicon

Downloading lexicons on 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.

Console

To download one or more lexicons

- 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.

AWS CLI

Amazon Polly provides the <u>GetLexicon</u> 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 <u>Uploading a lexicon</u>.

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
}
```

The following resources contain additional code samples for the GetLexicon operation:

• Java Sample: GetLexicon

Python (Boto3) Sample: <u>GetLexicon</u>

Deleting 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.

Console

To delete a lexicon

- 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**.
- Enter confirmation text and then choose **Delete** to remove the lexicon from the Region or Cancel to keep it.

Deleting a lexicon 271

AWS CLI

Amazon Polly provides the <u>DeleteLexicon</u> 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.

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

The following resources contain additional information for the DeleteLexicon operation:

Java Sample: DeleteLexicon

• Python (Boto3) Sample: DeleteLexicon

Deleting a lexicon 272

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 <u>Amazon S3 bucket encryption</u>, 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
- Creating long audio files

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:

- use of new Amazon Polly operations
- writing to the output S3 bucket
- publishing to the status SNS topic [optional]

The following policy grants only the necessary permissions required for asynchronous synthesis and can be attached to the IAM user.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "polly:StartSpeechSynthesisTask",
        "polly:GetSpeechSynthesisTask",
        "polly:ListSpeechSynthesisTasks"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "s3:PutObject",
      "Resource": "arn:aws:s3:::bucket-name/*"
    },
      "Effect": "Allow",
      "Action": "sns:Publish",
      "Resource": "arn:aws:sns:region:account:topic"
    }
  ]
}
```

Creating long audio files

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.

Console

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. Select **Long Form** as the engine if appropriate.
- 3. With **SSML** on or off, type or paste your text into the input box.
- 4. Choose the language, region, and voice for your text.
- 5. Choose Save to S3.



Both the **Download** and **Listen** options are greyed out if the text length is above the 3,000 character limit 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.



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**.

AWS 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 notifications 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.

```
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:

```
"SynthesisTask":
{
    "OutputFormat": "mp3",
    "OutputUri": "https://s3.us-east-2.amazonaws.com/your-bucket-name/optional/
prefix/path/file.<task_id>.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:

```
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.

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 <u>How Amazon Polly works</u> and follow the steps described in <u>Getting started with Amazon Polly</u>.

Topics

- Java samples
- Python samples
- Java example
- Python example (HTML5 Client and Python Server)
- iOS example
- Android example

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
- DescribeVoices
- GetLexicon
- ListLexicons
- PutLexicon
- StartSpeechSynthesisTask

Java samples 279

- Speech Marks
- SynthesizeSpeech

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 <u>DeleteLexicon</u> API.

```
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.

DeleteLexicon 280

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);
        } catch (Exception e) {
            System.err.println("Exception caught: " + e);
        }
    }
}
```

DescribeVoices 281

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.

```
package com.amazonaws.polly.samples;
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
import com.amazonaws.services.polly.model.GetLexiconRequest;
import com.amazonaws.services.polly.model.GetLexiconResult;
public class GetLexiconSample {
    private String LEXICON_NAME = "SampleLexicon";
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();
    public void getLexicon() {
        GetLexiconRequest getLexiconRequest = new
 GetLexiconRequest().withName(LEXICON_NAME);
        try {
            GetLexiconResult getLexiconResult = client.getLexicon(getLexiconRequest);
            System.out.println("Lexicon: " + getLexiconResult.getLexicon());
        } catch (Exception e) {
            System.err.println("Exception caught: " + e);
        }
    }
}
```

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.

```
package com.amazonaws.polly.samples;
import com.amazonaws.services.polly.AmazonPolly;
import com.amazonaws.services.polly.AmazonPollyClientBuilder;
```

GetLexicon 282

```
import com.amazonaws.services.polly.model.LexiconAttributes;
import com.amazonaws.services.polly.model.LexiconDescription;
import com.amazonaws.services.polly.model.ListLexiconsRequest;
import com.amazonaws.services.polly.model.ListLexiconsResult;
public class ListLexiconsSample {
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();
    public void listLexicons() {
        ListLexiconsRequest listLexiconsRequest = new ListLexiconsRequest();
        try {
            String nextToken;
            do {
                ListLexiconsResult listLexiconsResult =
 client.listLexicons(listLexiconsRequest);
                nextToken = listLexiconsResult.getNextToken();
                listLexiconsRequest.setNextToken(nextToken);
                for (LexiconDescription lexiconDescription :
 listLexiconsResult.getLexicons()) {
                    LexiconAttributes attributes = lexiconDescription.getAttributes();
                    System.out.println("Name: " + lexiconDescription.getName()
                            + ", Alphabet: " + attributes.getAlphabet()
                            + ", LanguageCode: " + attributes.getLanguageCode()
                            + ", LastModified: " + attributes.getLastModified()
                            + ", LexemesCount: " + attributes.getLexemesCount()
                            + ", LexiconArn: " + attributes.getLexiconArn()
                            + ", Size: " + attributes.getSize());
                }
            } 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.

PutLexicon 283

```
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 {
    AmazonPolly client = AmazonPollyClientBuilder.defaultClient();
    private String LEXICON_CONTENT = "<?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>test1</grapheme><alias>test2</alias></lexeme>" +
            "</lexicon>";
    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.

```
package com.amazonaws.parrot.service.tests.speech.task;
import com.amazonaws.parrot.service.tests.AbstractParrotServiceTest;
import com.amazonaws.services.polly.AmazonPolly;
```

StartSpeechSynthesisTask 284

```
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 String SNS_TOPIC_ARN = "arn:aws:sns:eu-
west-2:123456789012:synthesize-finish-topic";
    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())
                );
    }
```

StartSpeechSynthesisTask 285

```
private static SynthesisTask getSynthesisTask(String taskId) {
        GetSpeechSynthesisTaskRequest getSpeechSynthesisTaskRequest = new
 GetSpeechSynthesisTaskRequest()
                .withTaskId(taskId);
        GetSpeechSynthesisTaskResult result
 =AMAZON_POLLY_CLIENT.getSpeechSynthesisTask(getSpeechSynthesisTaskRequest);
        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.

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;
```

Speech Marks 286

```
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);
        }
    }
}
```

SynthesizeSpeech

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.

```
package com.amazonaws.polly.samples;
```

SynthesizeSpeech 287

```
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);
        }
    }
}
```

SynthesizeSpeech 288

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 Snipppets

- DeleteLexicon
- GetLexicon
- ListLexicon
- PutLexicon
- StartSpeechSynthesisTask
- SynthesizeSpeech

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 Setting up the AWS CLI.

For more information on this operation, see the reference for the DeleteLexicon API.

```
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
```

Python samples 289

```
# 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 Setting up the AWS CLI.

For more information on this operation, see the reference for the GetLexicon API.

```
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")
```

GetLexicon 290

```
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)
else:
    # 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 Setting up the AWS CLI.

ListLexicon 291

For more information on this operation, see the reference for the ListLexicons API.

```
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.

PutLexicon 292

The following code example uses default credentials stored in the AWS SDK configuration file. For information about creating the configuration file, see Setting up the AWS CLI.

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))
```

PutLexicon 293

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 Setting up the AWS CLI.

For more information, see the reference for StartSpeechSynthesisTask API.

```
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 SynthesizeSpeech 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 Generating speech from SSML documents.

```
import boto3
```

StartSpeechSynthesisTask 294

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 Mayen environment variables to run Mayen 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%;%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:

```
project xmlns="http://maven.apache.org/POM/4.0.0"
                   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/
maven-4.0.0.xsd">
<modelVersion>4.0.0</modelVersion>
<groupId>com.amazonaws.polly
<artifactId>java-demo</artifactId>
<version>0.0.1-SNAPSHOT</version>
<dependencies>
 <!-- https://mvnrepository.com/artifact/com.amazonaws/aws-java-sdk-polly -->
 <dependency>
  <groupId>com.amazonaws
  <artifactId>aws-java-sdk-polly</artifactId>
  <version>1.11.77</version>
 </dependency>
 <!-- https://mvnrepository.com/artifact/com.googlecode.soundlibs/jlayer -->
 <dependency>
```

```
<groupId>com.googlecode.soundlibs</groupId>
  <artifactId>jlayer</artifactId>
  <version>1.0.1-1
 </dependency>
</dependencies>
<build>
 <plugins>
  <plugin>
   <groupId>org.codehaus.mojo
   <artifactId>exec-maven-plugin</artifactId>
   <version>1.2.1
   <executions>
    <execution>
     <goals>
      <goal>java</goal>
     </goals>
    </execution>
   </executions>
   <configuration>
    <mainClass>com.amazonaws.demos.polly.PollyDemo</mainClass>
   </configuration>
  </plugin>
 </plugins>
</build>
</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:

```
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();
 // 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:

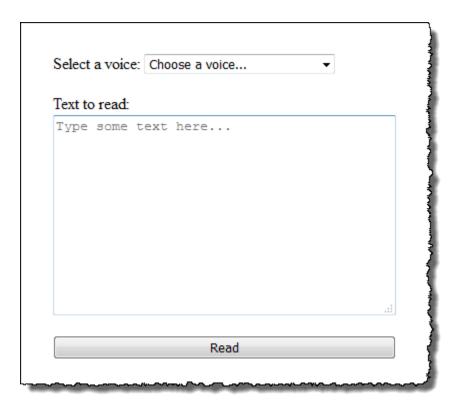
```
mvn clean compile exec:java
```

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):

Python example 299



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

- Save the server code as server.py. For the code, see <u>Python example: Python Server Code</u> (server.py).
- 2. Save the HTML5 client code as index.html. For the code, see Python example: HTML5 User Interface (index.html).

Python example 300

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).

```
$ 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.

You can pass the address and port for the application server to use as a parameter to server.py. For more information, run python server.py -h.

- 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).

```
<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+
 * - 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;
                // Parse the JSON in the response
                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'),
        submit = document.getElementById('submit'),
        supportedFormats = 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) {</pre>
        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);
                });
        });
    </script>
    <style>
        #input {
            min-width: 100px;
            max-width: 600px;
            margin: 0 auto;
            padding: 50px;
        }
        #input div {
            margin-bottom: 20px;
        }
        #text {
            width: 100%;
            height: 200px;
            display: block;
        }
        #submit {
            width: 100%;
    </style>
</head>
<body>
    <form id="input" method="GET" action="/read">
        <div>
            <label for="voice">Select a voice:</label>
            <select id="voice" name="voiceId" disabled>
                <option value="">Choose a voice...</option>
            </select>
        </div>
        <div>
            <label for="text">Text to read:</label>
            <textarea id="text" maxlength="1000" minlength="1" name="text"
                    placeholder="Type some text here..."></textarea>
        </div>
```

```
<input type="submit" value="Read" id="submit" />
    </form>
    <audio id="player"></audio>
</body>
</html>
```

Python example: Python Server Code (server.py)

This section provides the code for the Python server described in <u>Python example (HTML5 Client</u> and Python Server).

```
.....
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'
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")
class HTTPStatusError(Exception):
    """Exception wrapping a value from http.server.HTTPStatus"""
    def __init__(self, status, description=None):
```

```
11 11 11
        Constructs an error instance from a tuple of
        (code, message, description), see http.server.HTTPStatus
        super(HTTPStatusError, self).__init__()
        self.code = status.code
        self.message = status.message
        self.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")
        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"))
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,
                                                              ROUTE_INDEX))
    try:
        # Listen for requests indefinitely
        server.serve_forever()
    except KeyboardInterrupt:
        # A request to terminate has been received, stop the server
        print("\nShutting down...")
        server.socket.close()
```

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

iOS example 312

```
// Region of Amazon Polly.
let AwsRegion = AWSRegionType.usEast1

// Cognito pool ID. Pool needs to be 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

Synthesize Speech

```
// 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
```

iOS example 313

```
// 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 = AWSPollyVoiceId.joanna
// Create an task to synthesize speech using the given synthesis input
let builder = AWSPollySynthesizeSpeechURLBuilder.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";
```

Android example 314

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);
```

Android example 315

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();
    }
});
```

Android example 316

Quotas in Amazon Polly

Amazon Polly applies quotas to customer traffic by rejecting excessive requests. The default quota for the SynthesizeSpeech request with standard voices is 80 transactions per second (tps), in a single region, for a single AWS account. If limits did not increase, and if you generated 100 SynthesizeSpeech requests per second using a standard voice, 80 requests per second would succeed, and 20 requests per second would be throttled by Amazon Polly. These requests would return a response with HTTP status 400, and a response header indicating ThrottlingException. Amazon Polly also throttles traffic to all operations based on the request rate.

Speech synthesis limit examples

- Synthesize the first 24 letters of the English alphabet one letter at a time. If the synthesis of each letter took less than 50 milliseconds, with an operation limit of eight tps, synthesizing 24 letters would take at least three seconds. During that time, you could synthesize up to eight letters per second. Any further requests would be throttled. As the requests last a short time, they would be synthesized serially without overlap.
- Synthesize 16 paragraphs of text. If each paragraph was synthesized and fully received on the client side in two seconds or less, with an operation limit of eight concurrent requests, it would take at least four seconds to synthesize all 16 articles. In the first second, you could start up to eight requests. During concurrent requests, any attempt to start a new synthesis would be throttled due to the concurrency limit. You could synthesize the remaining eight paragraphs after the first two seconds, after the first batch of requests finishes.

Keep the following limits in mind when using Amazon Polly.

Topics

- Supported regions
- Quotas and throttle rates
- Pronunciation lexicons
- SynthesizeSpeech API operations
- SpeechSynthesisTask API operations
- Speech Synthesis Markup Language (SSML)

Supported regions

For a list of AWS Regions where Amazon Polly is available, see <u>Amazon Polly Endpoints and Quotas</u> in the *Amazon Web Services General Reference*.

- For Regions that support generative voices, see Generative voices.
- For Regions that support long-form voices, see Long-form voices.
- For Regions that support neural voices, see <u>the section called "Feature and region compatibility"</u> for neural TTS.

Quotas and throttle rates

The following table defines throttle rates per Amazon Polly operation. You can use the AWS Management Console to request quota increases for the adjustable quotas when needed.

Operation	Limit
Lexicon	
DeleteLexicon	Any 2 transactions per second (tps) from these operations combined.
PutLexicon	
GetLexicon	Maximum allowed burst of 4 tps.
ListLexicons	
Speech	
DescribeVoices	80 tps with a burst limit of 100 tps
SynthesizeSpeech	Generative voice: 8 tps
	Long-form voice: 8 tps with a burst limit of 10 tps
	Neural voice: 8 tps with a burst limit of 10 tps
	Standard voice: 80 tps with a burst limit of 100 tps

Supported regions 318

Operation	Limit
StartSpeechSynthes isTask	Generative voice: 1 tps
	Long-form voice: 1 tps
	Neural voice: 1 tps
	Standard voice: 10 tps with a burst limit of 12 tps
GetSynthesizeSpeec hTask and ListSynth esizeSpeechTask	Maximum allowed 10 tps combined

Concurrent requests

For **generative voice**, Amazon Polly supports up to 26 concurrent requests. For **long-form voice**, Amazon Polly supports up to 26 concurrent requests. For **neural voice**, Amazon Polly supports 8 tps with a burst limit of 10 tps, for up to 18 concurrent requests. Amazon Polly also supports limits for concurrent requests. For **standard voice**, Amazon Polly supports 80 tps for up to 80 concurrent requests.

Best practices to mitigate throttling

- Retry throttles with backoff and jitter so you can spread the load over a short period of time, and handle unexpected peaks in usage without compromising availability. AWS Code Sample Catalog is already configured to do this by default in many programming languages. Visit <u>feature</u> retry behavior to see the details.
- **Use <u>Amazon Polly metrics</u>**. Amazon Polly automatically publishes to CloudWatch to analyze your current usage and forecast usage growth.

Note

Before requesting a quota increase (where applicable), calculate your tps needs following the guidelines on this page. Amazon Polly secures only the required computational resources according to customer demand in order to keep your costs low.

Concurrent requests 319

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 40,000 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.

SynthesizeSpeech API operations

When estimating the usage of SynthesizeSpeech, keep in mind that the audio produced by Amazon Polly, especially for interactive applications, usually takes at least several seconds to be played. This reduces the rate of requests to SynthesizeSpeech, even for a large number of concurrent consumers. Additionally, Amazon Polly throttles SynthesizeSpeech requests by the number of concurrent requests that it synthesizes. There is no separate setting for concurrent requests. The concurrent requests limit has always the same value as the number of tps allowed and scales with it.

Short story example application. You can use Amazon Polly to build an application that plays a series of short stories. With this kind of app, the first story would start playing, and then the next, and so on, until a user quit the application. Each story would take around 0.5 seconds to synthesize and 10 seconds to play. In this scenario, you could expect one call to SynthesizeSpeech for every 10 seconds that the customer spent using the application. This would translate to one call per second for every 10 customers who were concurrently using the application. If you had 1000 customers concurrently using the application, you could expect an average call rate to SynthesizeSpeech of only 100 transactions per second.

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.

Pronunciation lexicons 320

For more information, see SynthesizeSpeech.



Note

Some limitations of the SynthesizeSpeech API operation can be bypassed using the StartSythensizeSpeechTask API operation. For more information, see Long audio files.

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 rosody> tag doesn't support values for the rate attribute lower than -80%.

For more information, see Generating speech from SSML documents.

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 <u>shared responsibility model</u> 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 <u>AWS</u>
 <u>Compliance Programs</u>. 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're 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
- Identity and Access Management in Amazon Polly
- Logging and Monitoring in Amazon Polly
- Compliance Validation for Amazon Polly
- Resilience in Amazon Polly
- Infrastructure Security in Amazon Polly
- Security Best Practices for Amazon Polly
- Using Amazon Polly with interface VPC endpoints

Data Protection in Amazon Polly

Amazon Polly conforms to the AWS <u>shared responsibility model</u>, 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 users 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 <u>AWS Shared Responsibility Model and GDPR</u> 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 <a href="the-section called "SynthesizeSpeech" operation is stateless and is not associated with a customer identity. You can't retrieve it from Amazon Polly later.

Data Protection 323

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 <u>digital signatures</u> 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
- Authenticating with identities
- Managing access using policies
- How Amazon Polly works with IAM
- Identity-based policy examples for Amazon Polly
- Amazon Polly API Permissions: Actions, Permissions, and Resources Reference
- Troubleshooting Amazon Polly identity and access

Audience

How you use AWS Identity and Access Management (IAM) differs, depending on the work that you do in Amazon Polly.

Service user – If you use the Amazon Polly service to do your job, then 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 can

Encryption in Transit 324

help you request the right permissions from your administrator. If you cannot access a feature in Amazon Polly, see Troubleshooting Amazon Polly identity and access.

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 service users should access. You must 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.

IAM administrator – If you're an IAM administrator, you might want to learn details about how you can write policies to manage access to Amazon Polly. To view example Amazon Polly identity-based policies that you can use in IAM, see Identity-based policy examples for Amazon Polly.

Authenticating with identities

Authentication is how you sign in to AWS using your identity credentials. You must be *authenticated* (signed in to AWS) as the AWS account root user, as an IAM user, or by assuming an IAM role.

You can sign in to AWS as a federated identity by using credentials provided through an identity source. AWS IAM Identity Center (IAM Identity Center) users, your company's single sign-on authentication, and your Google or Facebook credentials are examples of federated identities. When you sign in as a federated identity, your administrator previously set up identity federation using IAM roles. When you access AWS by using federation, you are indirectly assuming a role.

Depending on the type of user you are, you can sign in to the AWS Management Console or the AWS access portal. For more information about signing in to AWS, see How to sign in to your AWS account in the AWS Sign-In User Guide.

If you access AWS programmatically, AWS provides a software development kit (SDK) and a command line interface (CLI) to cryptographically sign your requests by using your credentials. If you don't use AWS tools, you must sign requests yourself. For more information about using the recommended method to sign requests yourself, see <u>AWS Signature Version 4 for API requests</u> in the *IAM User Guide*.

Regardless of the authentication method that you use, you might 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 Multi-factor authentication in

Authenticating with identities 325

the AWS IAM Identity Center User Guide and AWS Multi-factor authentication in IAM in the IAM User Guide.

AWS account root user

When you create an AWS account, you begin with one 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 don't use the root user for your everyday tasks. Safeguard your root user credentials and use them to perform the tasks that only the root user can perform. For the complete list of tasks that require you to sign in as the root user, see <u>Tasks that require root user credentials</u> in the *IAM User Guide*.

Federated identity

As a best practice, require human users, including users that require administrator access, to use federation with an identity provider to access AWS services by using temporary credentials.

A federated identity is a user from your enterprise user directory, a web identity provider, the AWS Directory Service, the Identity Center directory, or any user that accesses AWS services by using credentials provided through an identity source. When federated identities access AWS accounts, they assume roles, and the roles provide temporary credentials.

For centralized access management, we recommend that you use AWS IAM Identity Center. You can create users and groups in IAM Identity Center, or you can connect and synchronize to a set of users and groups in your own identity source for use across all your AWS accounts and applications. For information about IAM Identity Center, see What is IAM Identity Center? in the AWS IAM Identity Center User Guide.

IAM users and groups

An <u>IAM user</u> is an identity within your AWS account that has specific permissions for a single person or application. Where possible, we recommend relying on temporary credentials instead of creating IAM users who have long-term credentials such as passwords and access keys. However, if you have specific use cases that require long-term credentials with IAM users, we recommend that you rotate access keys. For more information, see <u>Rotate access keys regularly for use cases that require long-term credentials</u> in the <u>IAM User Guide</u>.

An <u>IAM group</u> 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

Authenticating with identities 326

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 <u>Use cases for IAM users</u> in the *IAM User Guide*.

IAM roles

An <u>IAM role</u> 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. To temporarily assume an IAM role in the AWS Management Console, you can <u>switch from a user to an IAM role (console)</u>. 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 <u>Methods to assume a role</u> in the <u>IAM User Guide</u>.

IAM roles with temporary credentials are useful in the following situations:

- Federated user access To assign permissions to a federated identity, you create a role and define permissions for the role. When a federated identity authenticates, the identity is associated with the role and is granted the permissions that are defined by the role. For information about roles for federation, see Create a role for a third-party identity provider (federation) in the IAM User Guide. If you use IAM Identity Center, you configure a permission set. To control what your identities can access after they authenticate, IAM Identity Center correlates the permission set to a role in IAM. For information about permissions sets, see Permission sets in the AWS IAM Identity Center User Guide.
- **Temporary IAM user permissions** An IAM user or role can assume an IAM role to temporarily take on different permissions for a specific task.
- 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 crossaccount 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 Cross account resource access in IAM 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.

Authenticating with identities 327

• Forward access sessions (FAS) – When you use an IAM user or role to perform actions in AWS, you are considered a principal. When you use some services, you might perform an action that then initiates another action in a different service. FAS uses the permissions of the principal calling an AWS service, combined with the requesting AWS service to make requests to downstream services. FAS requests are only made when a service receives a request that requires interactions with other AWS services or resources to complete. In this case, you must have permissions to perform both actions. For policy details when making FAS requests, see Forward access sessions.

- Service role A service role is an <u>IAM role</u> 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 <u>Create a role to delegate permissions to an AWS service</u> 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 AWS 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 <u>Use an IAM role to grant permissions to applications running on Amazon EC2 instances</u> in the *IAM User Guide*.

Managing access using policies

You control access in AWS by creating policies and attaching them to AWS identities or resources. A policy is an object in AWS that, when associated with an identity or resource, defines their permissions. AWS evaluates these policies when a principal (user, root user, or role session) makes a request. 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**.

By default, users and roles have no permissions. To grant users permission to perform actions on the resources that they need, an IAM administrator can create IAM policies. The administrator can then add the IAM policies to roles, and users can assume the roles.

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 <u>Define custom IAM permissions with customer managed policies</u> 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 Choose between managed policies and inline policies in the IAM User Guide.

Resource-based policies

Resource-based policies are JSON policy documents that you attach to a resource. Examples of resource-based policies are IAM *role trust policies* and Amazon S3 *bucket policies*. In services that support resource-based policies, service administrators can use them to control access to a specific resource. For the resource where the policy is attached, the policy defines what actions a specified principal can perform on that resource and under what conditions. You must <u>specify a principal</u> in a resource-based policy. Principals can include accounts, users, roles, federated users, or AWS services.

Resource-based policies are inline policies that are located in that service. You can't use AWS managed policies from IAM in a resource-based policy.

Access control lists (ACLs)

Access control lists (ACLs) control which principals (account members, users, or roles) have permissions to access a resource. ACLs are similar to resource-based policies, although they do not use the JSON policy document format.

Amazon S3, AWS WAF, and Amazon VPC are examples of services that support ACLs. To learn more about ACLs, see <u>Access control list (ACL) overview</u> in the *Amazon Simple Storage Service Developer Guide*.

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 an 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 Service
 control policies in the AWS Organizations User Guide.
- Resource control policies (RCPs) RCPs are JSON policies that you can use to set the maximum available permissions for resources in your accounts without updating the IAM policies attached to each resource that you own. The RCP limits permissions for resources in member accounts and can impact the effective permissions for identities, including the AWS account root user, regardless of whether they belong to your organization. For more information about Organizations and RCPs, including a list of AWS services that support RCPs, see Resource control policies (RCPs) 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, learn what IAM features are available to use with Amazon Polly.

IAM features you can use with Amazon Polly

IAM feature	Amazon Polly support
Identity-based policies	Yes
Resource-based policies	No
Policy actions	Yes
Policy resources	Yes
Policy condition keys (service-specific)	No
ACLs	No
ABAC (tags in policies)	No
Temporary credentials	Yes
Forward access sessions (FAS) for Amazon Polly	Yes
Service roles	No
Service-linked roles	No

To get a high-level view of how Amazon Polly and other AWS services work with most IAM features, see AWS services that work with IAM in the IAM User Guide.

Identity-based policies for Amazon Polly

Supports identity-based policies: Yes

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 Define custom IAM permissions with customer managed policies in the IAM User Guide.

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. You can't specify the principal in an identity-based policy because it applies to the user or role to which it is attached. To learn about all of the elements that you can use in a JSON policy, see IAM JSON policy elements reference in the IAM User Guide.

Identity-based policy examples for Amazon Polly

To view examples of Amazon Polly identity-based policies, see <u>Identity-based policy examples for Amazon Polly.</u>

Resource-based policies within Amazon Polly

Supports resource-based policies: No

Resource-based policies are JSON policy documents that you attach to a resource. Examples of resource-based policies are IAM *role trust policies* and Amazon S3 *bucket policies*. In services that support resource-based policies, service administrators can use them to control access to a specific resource. For the resource where the policy is attached, the policy defines what actions a specified principal can perform on that resource and under what conditions. You must <u>specify a principal</u> in a resource-based policy. Principals can include accounts, users, roles, federated users, or AWS services.

To enable cross-account access, you can specify an entire account or IAM entities in another account as the principal in a resource-based policy. Adding a cross-account principal to a resource-based policy is only half of establishing the trust relationship. When the principal and the resource

are in different AWS accounts, an IAM administrator in the trusted account must also grant the principal entity (user or role) permission to access the resource. They grant permission by attaching an identity-based policy to the entity. However, if a resource-based policy grants access to a principal in the same account, no additional identity-based policy is required. For more information, see Cross account resource access in IAM in the IAM User Guide.

Policy actions for Amazon Polly

Supports policy actions: Yes

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.

To see a list of Amazon Polly actions, see <u>Actions defined by Amazon Polly</u> in the *Service Authorization Reference*.

Policy actions in Amazon Polly use the following prefix before the action:

```
polly
```

To specify multiple actions in a single statement, separate them with commas.

```
"Action": [
    "polly:action1",
    "polly:action2"
]
```

To view examples of Amazon Polly identity-based policies, see <u>Identity-based policy examples for Amazon Polly.</u>

Policy resources for Amazon Polly

Supports policy resources: Yes

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 Resource JSON policy element specifies the object or objects to which the action applies. Statements must include either a Resource or a NotResource element. As a best practice, specify a resource using its Amazon Resource Name (ARN). You can do this for actions that support a specific resource type, known as resource-level permissions.

For actions that don't support resource-level permissions, such as listing operations, use a wildcard (*) to indicate that the statement applies to all resources.

```
"Resource": "*"
```

To see a list of Amazon Polly resource types and their ARNs, see <u>Resources defined by Amazon</u> <u>Polly</u> in the *Service Authorization Reference*. To learn with which actions you can specify the ARN of each resource, see <u>Actions defined by Amazon Polly</u>.

To view examples of Amazon Polly identity-based policies, see <u>Identity-based policy examples for Amazon Polly.</u>

Policy condition keys for Amazon Polly

Supports service-specific policy condition keys: No

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 Condition element (or Condition *block*) lets you specify conditions in which a statement is in effect. The Condition element is optional. You can create conditional expressions that use <u>condition operators</u>, such as equals or less than, to match the condition in the policy with values in the request.

If you specify multiple Condition elements in a statement, or multiple keys in a single Condition element, AWS evaluates them using a logical AND operation. If you specify multiple

values for a single condition key, AWS evaluates the condition using a logical OR operation. All of the conditions must be met before the statement's permissions are granted.

You can also use placeholder variables when you specify conditions. For example, you can grant an IAM user permission to access a resource only if it is tagged with their IAM user name. For more information, see IAM policy elements: variables and tags in the IAM User Guide.

AWS supports global condition keys and service-specific condition keys. To see all AWS global condition keys, see AWS global condition context keys in the *IAM User Guide*.

To see a list of Amazon Polly condition keys, see <u>Condition keys for Amazon Polly</u> in the *Service Authorization Reference*. To learn with which actions and resources you can use a condition key, see Actions defined by Amazon Polly.

To view examples of Amazon Polly identity-based policies, see <u>Identity-based policy examples for Amazon Polly</u>.

ACLs in Amazon Polly

Supports ACLs: No

Access control lists (ACLs) control which principals (account members, users, or roles) have permissions to access a resource. ACLs are similar to resource-based policies, although they do not use the JSON policy document format.

ABAC with Amazon Polly

Supports ABAC (tags in policies): No

Attribute-based access control (ABAC) is an authorization strategy that defines permissions based on attributes. In AWS, these attributes are called *tags*. You can attach tags to IAM entities (users or roles) and to many AWS resources. Tagging entities and resources is the first step of ABAC. Then you design ABAC policies to allow operations when the principal's tag matches the tag on the resource that they are trying to access.

ABAC is helpful in environments that are growing rapidly and helps with situations where policy management becomes cumbersome.

To control access based on tags, you provide tag information in the <u>condition element</u> of a policy using the aws:ResourceTag/*key-name*, aws:RequestTag/*key-name*, or aws:TagKeys condition keys.

If a service supports all three condition keys for every resource type, then the value is **Yes** for the service. If a service supports all three condition keys for only some resource types, then the value is **Partial**.

For more information about ABAC, see <u>Define permissions with ABAC authorization</u> in the *IAM User Guide*. To view a tutorial with steps for setting up ABAC, see <u>Use attribute-based access control</u> (ABAC) in the *IAM User Guide*.

Using temporary credentials with Amazon Polly

Supports temporary credentials: Yes

Some AWS services don't work when you sign in using temporary credentials. For additional information, including which AWS services work with temporary credentials, see <u>AWS services that</u> work with IAM in the *IAM User Guide*.

You are using temporary credentials if you sign in to the AWS Management Console using any method except a user name and password. For example, when you access AWS using your company's single sign-on (SSO) link, that process automatically creates temporary credentials. You also automatically create temporary credentials when you sign in to the console as a user and then switch roles. For more information about switching roles, see Switch from a user to an IAM role (console) in the IAM User Guide.

You can manually create temporary credentials using the AWS CLI or AWS API. You can then use those temporary credentials to access AWS. AWS recommends that you dynamically generate temporary credentials instead of using long-term access keys. For more information, see Temporary security credentials in IAM.

Cross-service forward access sessions (FAS) for Amazon Polly

Supports forward access sessions (FAS): Yes

When you use an IAM user or role to perform actions in AWS, you are considered a principal. When you use some services, you might perform an action that then initiates another action in a different service. FAS uses the permissions of the principal calling an AWS service, combined with the requesting AWS service to make requests to downstream services. FAS requests are only made when a service receives a request that requires interactions with other AWS services or resources to complete. In this case, you must have permissions to perform both actions. For policy details when making FAS requests, see Forward access sessions.

Service roles for Amazon Polly

Supports service roles: No

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 Create a role to delegate permissions to an AWS service in the IAM User Guide.

Marning

Changing the permissions for a service role might break Amazon Polly functionality. Edit service roles only when Amazon Polly provides guidance to do so.

Service-linked roles for Amazon Polly

Supports service-linked roles: No

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 AWS account and are owned by the service. An IAM administrator can view, but not edit the permissions for service-linked roles.

For details about creating or managing service-linked roles, see AWS services that work with IAM. Find a service in the table that includes a Yes in the Service-linked role column. Choose the Yes link to view the service-linked role documentation for that service.

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 <u>Access Management</u> 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 <u>Identities (Users, Groups, and Roles)</u> in the *IAM User Guide*.

Identity-based policy examples for Amazon Polly

By default, users and roles don't have permission to create or modify Amazon Polly resources. They also can't perform tasks by using the AWS Management Console, AWS Command Line Interface

(AWS CLI), or AWS API. To grant users permission to perform actions on the resources that they need, an IAM administrator can create IAM policies. The administrator can then add the IAM policies to roles, and users can assume the roles.

To learn how to create an IAM identity-based policy by using these example JSON policy documents, see Create IAM policies (console) in the IAM User Guide.

For details about actions and resource types defined by Amazon Polly, including the format of the ARNs for each of the resource types, see <u>Actions, resources, and condition keys for Amazon Polly</u> in the *Service Authorization Reference*.

Topics

- Policy best practices
- Using the Amazon Polly console
- · Allow users to view their own permissions
- AWS managed (predefined) policies for Amazon Polly
- Customer-managed policy examples

Policy best practices

Identity-based policies 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 with AWS managed policies and move toward least-privilege permissions To
 get started granting permissions to your users and workloads, use the AWS managed policies
 that grant permissions for many common use cases. They are available in your AWS account. We
 recommend that you reduce permissions further by defining AWS customer managed policies
 that are specific to your use cases. For more information, see <u>AWS managed policies</u> or <u>AWS</u>
 managed policies for job functions in the IAM User Guide.
- Apply least-privilege permissions When you set permissions with IAM policies, grant only the
 permissions required to perform a task. You do this by defining the actions that can be taken on
 specific resources under specific conditions, also known as least-privilege permissions. For more
 information about using IAM to apply permissions, see Policies and permissions in IAM in the
 IAM User Guide.
- Use conditions in IAM policies to further restrict access You can add a condition to your policies to limit access to actions and resources. For example, you can write a policy condition to

specify that all requests must be sent using SSL. You can also use conditions to grant access to service actions if they are used through a specific AWS service, such as AWS CloudFormation. For more information, see IAM JSON policy elements: Condition in the IAM User Guide.

- Use IAM Access Analyzer to validate your IAM policies to ensure secure and functional
 permissions IAM Access Analyzer validates new and existing policies so that the policies
 adhere to the IAM policy language (JSON) and IAM best practices. IAM Access Analyzer provides
 more than 100 policy checks and actionable recommendations to help you author secure and
 functional policies. For more information, see <u>Validate policies with IAM Access Analyzer</u> in the
 IAM User Guide.
- Require multi-factor authentication (MFA) If you have a scenario that requires IAM users or
 a root user in your AWS account, turn on MFA for additional security. To require MFA when API
 operations are called, add MFA conditions to your policies. For more information, see Secure API
 access with MFA in the IAM User Guide.

For more information about best practices in IAM, see <u>Security best practices in IAM</u> in the *IAM User Guide*.

Using the Amazon Polly console

To access the Amazon Polly console, you must have a minimum set of permissions. These permissions must allow you to list and view details about the Amazon Polly resources in your AWS account. If you create an identity-based policy that is more restrictive than the minimum required permissions, the console won't function as intended for entities (users or roles) with that policy.

You don't need to allow minimum console permissions for users that are making calls only to the AWS CLI or the AWS API. Instead, allow access to only the actions that match the API operation that they're trying to perform.

To ensure that users and roles can still use the Amazon Polly console, also attach the Amazon Polly ConsoleAccess or ReadOnly AWS managed policy to the entities. For more information, see Adding permissions to a user in the IAM User Guide.

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:

```
{
  "Version": "2012-10-17",
    "Statement": [{
        "Sid": "Console-AllowAllPollyActions",
```

```
"Effect": "Allow",
    "Action": [
          "polly:*"],
          "Resource": "*"
     }
]
```

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 programmatically using the AWS CLI or AWS API.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ViewOwnUserInfo",
            "Effect": "Allow",
            "Action": [
                "iam:GetUserPolicy",
                "iam:ListGroupsForUser",
                "iam:ListAttachedUserPolicies",
                "iam:ListUserPolicies",
                "iam:GetUser"
            ],
            "Resource": ["arn:aws:iam::*:user/${aws:username}"]
        },
        {
            "Sid": "NavigateInConsole",
            "Effect": "Allow",
            "Action": [
                "iam:GetGroupPolicy",
                "iam:GetPolicyVersion",
                "iam:GetPolicy",
                "iam:ListAttachedGroupPolicies",
                "iam:ListGroupPolicies",
                "iam:ListPolicyVersions",
                "iam:ListPolicies",
                "iam:ListUsers"
            ],
            "Resource": "*"
```

```
}
      ]
}
```

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're using AWS SDKs or the AWS CLI. When you're 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
- Example 2: Allow all Amazon Polly actions except DeleteLexicon
- Example 3: Allow DeleteLexicon
- Example 4: Allow Delete Lexicon in a specified Region
- Example 5: Allow DeleteLexicon for specified Lexicon

Example 1: Allow All Amazon Polly Actions

After you sign up (see <u>Getting started with Amazon Polly</u>) 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.

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.

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.

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"],
      "Resource": "arn:aws:polly:us-east-2:123456789012:lexicon/*"
   ]
}
```

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"],
      "Resource": "arn:aws:polly:us-east-2:123456789012:lexicon/myLexicon"
   ]
}
```

Amazon Polly API Permissions: Actions, Permissions, and Resources Reference

When you're 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/*.

Amazon Polly API and Required Permissions for Actions

API Operation: DeleteLexicon

Required Permissions (API Action): polly:DeleteLexicon

Resources: arn:aws:polly:region:account-id:lexicon/LexiconName

API Operation: DescribeVoices

Required Permissions (API Action): polly:DescribeVoices

Resources: arn:aws:polly:region:account-id:lexicon/voice-name

API Operation: GetLexicon

Required Permissions (API Action): polly:GetLexicon

Resources: arn:aws:polly:region:account-id:lexicon/voice-name

API Operation: ListLexicons

Required Permissions (API Action): polly:ListLexicons

Resources: arn: aws:polly:region:account-id:lexicon/*

API Operation: PutLexicon

Required Permissions (API Action): polly:ListLexicons

Resources: *

API Operation: SynthesizeSpeech

Required Permissions (API Action): polly:SynthesizeSpeech

Resources: *

Troubleshooting Amazon Polly identity and access

Use the following information to help you diagnose and fix common issues that you might encounter when working with Amazon Polly and IAM.

Topics

- I am not authorized to perform an action in Amazon Polly
- I am not authorized to perform iam:PassRole
- I want to allow people outside of my AWS account to access my Amazon Polly resources

I am not authorized to perform an action in Amazon Polly

If you receive an error that you're not authorized to perform an action, your policies must be updated to allow you to perform the action.

The following example error occurs when the mateojackson IAM user tries to use the console to view details about a fictional *my-example-widget* resource but doesn't have the fictional polly: *GetWidget* permissions.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform: polly:GetWidget on resource: my-example-widget
```

In this case, the policy for the mateojackson user must be updated to allow access to the *my-example-widget* resource by using the polly: *GetWidget* action.

If you need help, contact your AWS administrator. Your administrator is the person who provided you with your sign-in credentials.

I am not authorized to perform iam:PassRole

If you receive an error that you're not authorized to perform the iam: PassRole action, your policies must be updated to allow you to pass a role to Amazon Polly.

Some AWS services allow you to pass an existing role to that service instead of creating a new service role or service-linked role. To do this, you must have permissions to pass the role to the service.

Troubleshooting 347

The following example error occurs when an IAM user named marymajor tries to use the console to perform an action in Amazon Polly. However, the action requires the service to have permissions that are granted by a service role. Mary does not have permissions to pass the role to the service.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform: iam:PassRole
```

In this case, Mary's policies must be updated to allow her to perform the iam: PassRole action.

If you need help, contact your AWS administrator. Your administrator is the person who provided you with your sign-in credentials.

I want to allow people outside of my AWS account to access my Amazon Polly resources

You can create a role that users in other accounts or people outside of your organization can use to access your resources. You can specify who is trusted to assume the role. For services that support resource-based policies or access control lists (ACLs), you can use those policies to grant people access to your resources.

To learn more, consult the following:

- To learn whether Amazon Polly supports these features, see <u>How Amazon Polly works with IAM</u>.
- To learn how to provide access to your resources across AWS accounts that you own, see <u>Providing access to an IAM user in another AWS account that you own</u> in the *IAM User Guide*.
- To learn how to provide access to your resources to third-party AWS accounts, see Providing access to AWS accounts owned by third parties in the IAM User Guide.
- To learn how to provide access through identity federation, see <u>Providing access to externally</u> authenticated users (identity federation) in the *IAM User Guide*.
- To learn the difference between using roles and resource-based policies for cross-account access, see Cross account resource access in IAM in the IAM User Guide.

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.

Logging and Monitoring 348

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
don't 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.

• 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.

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.

For a list of AWS services in scope of specific compliance programs, see <u>AWS Services in Scope by</u> Compliance Program. For general information, see AWS Compliance Programs.

You can download third-party audit reports using AWS Artifact. For more information, see Downloading Reports in AWS Artifact.

Your compliance responsibility when using Amazon Polly is determined by the sensitivity of your data, your company's compliance objectives, and applicable laws and regulations. AWS provides the following resources to help with compliance:

- <u>Security and Compliance Quick Start Guides</u> These deployment guides discuss architectural
 considerations and provide steps for deploying security- and compliance-focused baseline
 environments on AWS.
- <u>Architecting for HIPAA Security and Compliance Whitepaper</u> This whitepaper describes how companies can use AWS to create HIPAA-compliant applications.
- <u>AWS Compliance Resources</u> This collection of workbooks and guides might apply to your industry and location.
- <u>Evaluating Resources with Rules</u> in the *AWS Config Developer Guide* The AWS Config service assesses how well your resource configurations comply with internal practices, industry guidelines, and regulations.

Compliance Validation 349

• <u>AWS Security Hub</u> – This AWS service provides a comprehensive view of your security state within AWS that helps you check your compliance with security industry standards and best practices.

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 <u>AWS Security Token Service</u> (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 <u>AWS Data Privacy FAQ</u>.

Amazon Polly does not retain the the content of text submissions.

Resilience 350

For a broad view of AWS security, including compliance, penetration testing, bulletins, and resources, visit the AWS Cloud Security website.

Using Amazon Polly with interface VPC endpoints

If you use Amazon Virtual Private Cloud (Amazon VPC) to host your AWS resources, you can establish a private connection between your VPC and Amazon Polly. You can use this connection to synthesize speech with Amazon Polly without traversing the public internet.

Amazon VPC is an AWS service that you can use to launch AWS resources in a virtual network that you define. With a VPC, you have control over your network settings, such the IP address range, subnets, route tables, and network gateways. To connect your VPC to Amazon Polly, you define an *interface VPC endpoint* for Amazon Polly. This type of endpoint enables you to connect your VPC to AWS services. The endpoint provides reliable, scalable connectivity to Amazon Polly without requiring an internet gateway, network address translation (NAT) instance, or VPN connection. For more information, see the What is Amazon VPC in the Amazon VPC User Guide.

Interface VPC endpoints are powered by AWS PrivateLink, an AWS technology that enables private communication between AWS services using an elastic network interface with private IP addresses. For more information, see New - AWS PrivateLink for AWS services.

The following steps are for users of Amazon VPC. For more information, see <u>Getting Started</u> in the *Amazon VPC User Guide*.

Availability

VPC endpoints are supported in all the <u>Regions where Amazon Polly is supported</u>. For more information about AWS Regions and Availability Zones, see AWS Global Infrastructure.

Creating a VPC endpoint for Amazon Polly

To start using Amazon Polly with your VPC, create an interface VPC endpoint for Amazon Polly. The service to choose is **com.amazonaws.** Region.polly. You don't need to change any settings for Amazon Polly. For more information, see <u>Creating an Interface Endpoint</u> in the Amazon VPC User Guide.

Testing the connection between your VPC and Amazon Polly

After you create the endpoint, you can test the connection.

To test the connection between your VPC and your Amazon Polly endpoint

1. Connect to an Amazon EC2 instance that resides in your VPC. For information about connecting, see Connecting to your Windows instance in the Amazon EC2 documentation.

2. From the instance, use aws polly describe-voices from the AWS CLI to list available Amazon Polly voices.

If the response to the command includes the list of available Amazon Polly voices, the command has succeeded, and your VPC endpoint is working.

Controlling access to your Amazon Polly endpoint

A VPC endpoint policy is an IAM resource policy that you attach to an endpoint when you create or modify the endpoint. If you don't attach a policy when you create an endpoint, we attach a default policy for you that allows full access to the service. An endpoint policy doesn't override or replace IAM user policies or service-specific policies. It's a separate policy for controlling access from the endpoint to the specified service.

Endpoint policies must be written in JSON format.

For more information, see <u>Controlling Access to Services with VPC Endpoints</u> in the *Amazon VPC User Guide*.

The following is an example of an endpoint policy for Amazon Polly. This policy enables users connecting to Amazon Polly through the VPC to describe voices and synthesize speech with Amazon Polly, and prevents them from performing other Amazon Polly actions.

```
{
    "Statement": [
    {
        "Sid": "SynthesisAndDescribeVoicesOnly",
        "Principal": "*",
        "Action": [
            "polly:DescribeVoices",
            "polly:SynthesizeSpeech"
        ],
        "Effect": "Allow",
        "Resource": "*"
    }
}
```

}

To modify the VPC endpoint policy for Amazon Polly

- 1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc.
- 2. In the navigation pane, choose **Endpoints**.
- 3. If you have not already created the endpoint for Amazon Polly, choose **Create endpoint**. Then select **com.amazonaws.***Region.***polly** and choose **Create endpoint**.
- 4. Select the **com.amazonaws**. **Region**. **polly** endpoint, and choose the **Policy** tab in the lower half of the screen.
- 5. Choose **Edit Policy** and make the changes to the policy.

Support for VPC context keys

Amazon Polly supports the aws: SourceVpc and aws: SourceVpce context keys that can limit access to specific VPCs or specific VPC endpoints. These keys work only when the user is using VPC endpoints. For more information, see Keys Available for Some Services in the IAM user Guide.

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 <u>AWS CloudTrail</u> 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
- DescribeVoices
- GetLexicon
- GetSpeechSynthesisTask
- ListLexicons
- ListSpeechSynthesisTasks
- PutLexicon
- StartSpeechSynthesisTask
- SynthesizeSpeech

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 user 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.

```
"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"
                ],
                "engine": "neural",
                "outputFormat": "mp3",
                "sampleRate": "22050",
                "text": "********,
                "textType": "text",
                "voiceId": "Kendra"
            },
            "responseElements": null,
            "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"
            }
        }
    ]
}
```

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)

- 1. Open the CloudWatch console at https://console.aws.amazon.com/cloudwatch/.
- 2. In the navigation pane, choose **Metrics**.
- 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 on the AWS 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.

```
"Value": "SynthesizeSpeech"

}
],
    "MetricName": "ResponseLatency"

},
{
    "Namespace": "AWS/Polly",
    "Dimensions": [
        {
            "Name": "Operation",
            "Value": "SynthesizeSpeech"
        }
    ],
    "MetricName": "RequestCharacters"
}
```

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.

Metric	Description
RequestCharacters	The number of characters in the request. This is billable characters only and does not include SSML tags. Valid Dimension: Operation
	Valid Statistics: Minimum, Maximum, Average, SampleCount, Sum Unit: Count
ResponseLatency	The latency between when the request was made and the start of the streaming response. Valid Dimensions: Operation

Amazon Polly Metrics 358

Metric	Description
	Valid Statistics: Minimum, Maximum, Average, SampleCount
	Unit: milliseconds
2XXCount	HTTP 200 level code returned upon a successful response.
	Valid Dimensions: Operation
	Valid Statistics: Average, SampleCount, Sum
	Unit: Count
4XXCount	HTTP 400 level error code returned upon an error. For each successful response, a zero (0) is emitted.
	Valid Dimensions: Operation
	Valid Statistics: Average, SampleCount, Sum
	Unit: Count
5XXCount	HTTP 500 level error code returned upon an error. For each successful response, a zero (0) is emitted.
	Valid Dimensions: Operation
	Valid Statistics: Average, SampleCount, Sum
	Unit: Count

Dimensions for Amazon Polly Metrics

Amazon Polly metrics use the AWS/Polly namespace and provide metrics for the following dimension:

Dimension	Description
Operation	Metrics are grouped by the API method they refer to. Possible values are SynthesizeSpeech , PutLexicon , DescribeVoices , etc.

Amazon Polly API Reference

Amazon Polly is a web service that makes it easy to synthesize speech from text.

The Amazon Polly service provides API operations for synthesizing high-quality speech from plain text and Speech Synthesis Markup Language (SSML), along with managing pronunciations lexicons that enable you to get the best results for your application domain.

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
- Data Types
- Common Errors
- Common Parameters

Actions

The following actions are supported:

- DeleteLexicon
- DescribeVoices
- GetLexicon
- GetSpeechSynthesisTask
- ListLexicons
- ListSpeechSynthesisTasks
- PutLexicon
- StartSpeechSynthesisTask
- SynthesizeSpeech

Actions 361

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

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.

DeleteLexicon 362

Verify that the lexicon exists, is in the region (see ListLexicons) 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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

DeleteLexicon 363

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=LanguageTode
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

Engine

Specifies the engine (standard, neural, long-form or generative) used by Amazon Polly when processing input text for speech synthesis.

```
Valid Values: standard | neural | long-form | generative
```

IncludeAdditionalLanguageCodes

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

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.

```
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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

NextToken

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

}

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

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

A list of voices with their properties.

Type: Array of <u>Voice</u> objects

Errors

InvalidNextTokenException

The NextToken is invalid. Verify that it's spelled correctly, and then try again.

HTTP Status Code: 400

${\bf Service Failure Exception}$

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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- 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

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,
```

GetLexicon 368

```
"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

Lexicon object that provides name and the string content of the lexicon.

Type: Lexicon object

LexiconAttributes

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: <u>LexiconAttributes</u> 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 <u>ListLexicons</u>) 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

GetLexicon 369

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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

GetLexicon 370

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

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",
        "OutputUri": "string",
        "RequestCharacters": number,
```

GetSpeechSynthesisTask 371

```
"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

SynthesisTask object that provides information from the requested task, including output format, creation time, task status, and so on.

Type: SynthesisTask 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

2. 300

Synthesis Task Not Found Exception

The Speech Synthesis task with requested Task ID cannot be found.

HTTP Status Code: 400

GetSpeechSynthesisTask 372

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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

GetSpeechSynthesisTask 373

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

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

ListLexicons 374

```
],
"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

A list of lexicon names and attributes.

Type: Array of LexiconDescription objects

NextToken

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:

ListLexicons 375

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

ListLexicons 376

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

Maximum number of speech synthesis tasks returned in a List operation.

Valid Range: Minimum value of 1. Maximum value of 100.

NextToken

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

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

ListSpeechSynthesisTasks 377

```
{
   "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

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

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** objects

ListSpeechSynthesisTasks 378

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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

ListSpeechSynthesisTasks 379

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

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

Content of the PLS lexicon as string data.

Type: String

Required: Yes

PutLexicon 380

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

LexiconSizeExceededException

The maximum size of the specified lexicon would be exceeded by this operation.

HTTP Status Code: 400

MaxLexemeLengthExceededException

The maximum size of the lexeme would be exceeded by this operation.

HTTP Status Code: 400

${\bf MaxLexicons Number Exceeded Exception}$

The maximum number of lexicons would be exceeded by this operation.

HTTP Status Code: 400

ServiceFailureException

An unknown condition has caused a service failure.

HTTP Status Code: 500

UnsupportedPlsAlphabetException

The alphabet specified by the lexicon is not a supported alphabet. Valid values are x-sampa and ipa.

PutLexicon 381

HTTP Status Code: 400

UnsupportedPlsLanguageException

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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

PutLexicon 382

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

Specifies the engine (standard, neural, long-form or generative) 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 | long-form | generative

Required: No

LanguageCode

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.

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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

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: Yes

OutputS3BucketName

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

The Amazon S3 key prefix for the output speech file.

Type: String

Pattern: ^[0-9a-zA-Z\/\!\-_\.*\'\(\):;\\$@=+\,\?&]{0,800}\$

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". The default value for generative 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: a rn: aws(-(cn|iso(-b)?|us-gov))?:sns:[a-z0-9_-]{1,50}:\d{12}:[a-zA-Z0-9_-]{1,251}([a-zA-Z0-9_-]{0,5}|\.fifo)\$

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

Text

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

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 | Arlet | Hannah | Arthur | Daniel | Liam | Pedro | Kajal |
Hiujin | Laura | Elin | Ida | Suvi | Ola | Hala | Andres | Sergio | Remi
| Adriano | Thiago | Ruth | Stephen | Kazuha | Tomoko | Niamh | Sofie |
Lisa | Isabelle | Zayd | Danielle | Gregory | Burcu | Jitka | Sabrina
```

Required: Yes

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

SynthesisTask object that provides information and attributes about a newly submitted speech synthesis task.

Type: SynthesisTask 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) 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.

StartSpeechSynthesisTask 389

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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

StartSpeechSynthesisTask 390

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

```
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

Specifies the engine (standard, neural, long-form, or generative) for Amazon Polly to use when processing input text for speech synthesis. Provide an engine that is supported by the voice you select. If you don't provide an engine, the standard engine is selected by default. If a chosen voice isn't supported by the standard engine, this will result in an error. For information on Amazon Polly voices and which voices are available for each engine, see Available Voices.

Type: String

Valid Values: standard | neural | long-form | generative

Required: Yes

Type: String

Valid Values: standard | neural | long-form | generative

Required: No

LanguageCode

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: 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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

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. 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

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

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". The default value for generative voices is "24000".

Valid values for pcm are "8000" and "16000" The default value is "16000".

Type: String

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

Text

Input text to synthesize. If you specify ssml as the TextType, follow the SSML format for the input text.

Type: String

Required: Yes

TextType

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

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 | Arlet | Hannah | Arthur | Daniel | Liam | Pedro | Kajal |
Hiujin | Laura | Elin | Ida | Suvi | Ola | Hala | Andres | Sergio | Remi
| Adriano | Thiago | Ruth | Stephen | Kazuha | Tomoko | Niamh | Sofie |
Lisa | Isabelle | Zayd | Danielle | Gregory | Burcu | Jitka | Sabrina
```

Required: Yes

Response Syntax

HTTP/1.1 200

Content-Type: ContentType

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

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 application/x-jsonstream.

RequestCharacters

Number of characters synthesized.

The response returns the following as the HTTP body.

AudioStream

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) 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 v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

Data Types

The following data types are supported:

- Lexicon
- LexiconAttributes
- LexiconDescription
- SynthesisTask
- Voice

Data Types 397

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 Java V2
- AWS SDK for Ruby V3

Lexicon 398

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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

Required: No

LastModified

Date lexicon was last modified (a timestamp value).

Type: Timestamp

Required: No

LexemesCount

Number of lexemes in the lexicon.

Type: Integer

LexiconAttributes 399

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 Java V2
- AWS SDK for Ruby V3

LexiconAttributes 400

LexiconDescription

Describes the content of the lexicon.

Contents

Attributes

Provides lexicon metadata.

Type: LexiconAttributes 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 Java V2
- AWS SDK for Ruby V3

LexiconDescription 401

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, neural, long-form or generative) 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 | long-form | generative

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
```

```
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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

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

${\bf Request Characters}$

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". The default value for generative 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: a rn: aws(-(cn|iso(-b)?|us-gov))?:sns:[a-z0-9_-]{1,50}:\d{12}:[a-zA-Z0-9_-]{1,251}([a-zA-Z0-9_-]{0,5}|\.fifo)\$

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 | Arlet | Hannah | Arthur | Daniel | Liam | Pedro | Kajal |
Hiujin | Laura | Elin | Ida | Suvi | Ola | Hala | Andres | Sergio | Remi
| Adriano | Thiago | Ruth | Stephen | Kazuha | Tomoko | Niamh | Sofie |
Lisa | Isabelle | Zayd | Danielle | Gregory | Burcu | Jitka | Sabrina
```

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 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

```
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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

Required: No

Gender

Gender of the voice.

Type: String

Valid Values: Female | Male

Required: No

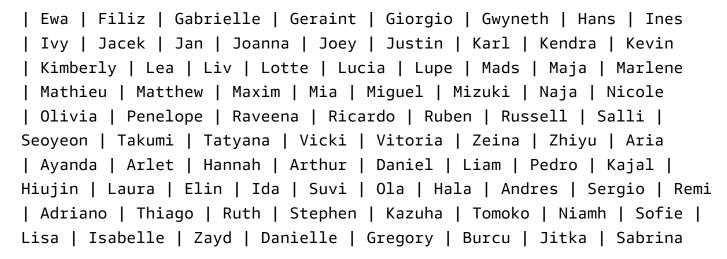
Id

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
```

Voice 407



Required: No

LanguageCode

Language code of the voice.

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 | ca-ES | de-AT | yue-CN | ar-AE | fi-FI | en-IE | nl-BE | fr-BE | cs-CZ | de-CH
```

Required: No

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

Voice 408

SupportedEngines

Specifies which engines (standard, neural, long-form or generative) are supported by a given voice.

Type: Array of strings

Valid Values: standard | neural | long-form | generative

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 Java V2
- AWS SDK for Ruby V3

Common Errors

This section lists the errors common to the API actions of all AWS services. For errors specific to an API action for this service, see the topic for that API action.

${\bf Access Denied Exception}$

You do not have sufficient access to perform this action.

HTTP Status Code: 400

IncompleteSignature

The request signature does not conform to AWS standards.

HTTP Status Code: 400

InternalFailure

The request processing has failed because of an unknown error, exception or failure.

HTTP Status Code: 500

Common Errors 409

InvalidAction

The action or operation requested is invalid. Verify that the action is typed correctly.

HTTP Status Code: 400

InvalidClientTokenId

The X.509 certificate or AWS access key ID provided does not exist in our records.

HTTP Status Code: 403

NotAuthorized

You do not have permission to perform this action.

HTTP Status Code: 400

OptInRequired

The AWS access key ID needs a subscription for the service.

HTTP Status Code: 403

RequestExpired

The request reached the service more than 15 minutes after the date stamp on the request or more than 15 minutes after the request expiration date (such as for pre-signed URLs), or the date stamp on the request is more than 15 minutes in the future.

HTTP Status Code: 400

ServiceUnavailable

The request has failed due to a temporary failure of the server.

HTTP Status Code: 503

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 400

ValidationError

The input fails to satisfy the constraints specified by an AWS service.

Common Errors 410

HTTP Status Code: 400

Common Parameters

The following list contains the parameters that all actions use for signing Signature Version 4 requests with a query string. Any action-specific parameters are listed in the topic for that action. For more information about Signature Version 4, see <u>Signing AWS API requests</u> in the *IAM User Guide*.

Action

The action to be performed.

Type: string

Required: Yes

Version

The API version that the request is written for, expressed in the format YYYY-MM-DD.

Type: string

Required: Yes

X-Amz-Algorithm

The hash algorithm that you used to create the request signature.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Valid Values: AWS4-HMAC-SHA256

Required: Conditional

X-Amz-Credential

The credential scope value, which is a string that includes your access key, the date, the region you are targeting, the service you are requesting, and a termination string ("aws4_request"). The value is expressed in the following format: access_key/YYYYMMDD/region/service/aws4_request.

Common Parameters 411

For more information, see Create a signed AWS API request in the IAM User Guide.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

X-Amz-Date

The date that is used to create the signature. The format must be ISO 8601 basic format (YYYYMMDD'T'HHMMSS'Z'). For example, the following date time is a valid X-Amz-Date value: 20120325T120000Z.

Condition: X-Amz-Date is optional for all requests; it can be used to override the date used for signing requests. If the Date header is specified in the ISO 8601 basic format, X-Amz-Date is not required. When X-Amz-Date is used, it always overrides the value of the Date header. For more information, see Elements of an AWS API request signature in the *IAM User Guide*.

Type: string

Required: Conditional

X-Amz-Security-Token

The temporary security token that was obtained through a call to AWS Security Token Service (AWS STS). For a list of services that support temporary security credentials from AWS STS, see AWS services that work with IAM in the *IAM User Guide*.

Condition: If you're using temporary security credentials from AWS STS, you must include the security token.

Type: string

Required: Conditional

X-Amz-Signature

Specifies the hex-encoded signature that was calculated from the string to sign and the derived signing key.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Common Parameters 412

Type: string

Required: Conditional

X-Amz-SignedHeaders

Specifies all the HTTP headers that were included as part of the canonical request. For more information about specifying signed headers, see <u>Create a signed AWS API request</u> in the *IAM User Guide*.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

Common Parameters 413

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: November 14, 2024

Change	Description	Date
New voices added for generative text-to-speech	Amazon Polly now provides seven new generative voices: Pedro, Andrés, Sergio, Daniel, Kajal, Rémi, Bianca. See Generative voices for a list of generative TTS voices.	November 21, 2024
New long-form voices added	Amazon Polly now provides more long-form voices. One additional English (US) and two new Spanish (Spain) long-form voices were added: Patrick, Alba, and Raúl. See Long-form voices for a list of all long-form voices.	November 14, 2024
New voices added for generative text-to-speech	Amazon Polly now provides six new generative voices: Ayanda, Léa, Lucia, Mía, Lupe, and Vicki. See Generative voices for a list of generative TTS voices.	November 6, 2024
New voices added for generative text-to-speech	Amazon Polly now provides four new generative voices: Olivia, Danielle, Joanna, and Stephen. See Generative	October 10, 2024

	<u>voices</u> for a list of generative TTS voices.	
New languages added for text-to-speech	Amazon Polly now provides two new TTS languages: Czech (cs-CZ) and German (Swiss) (de-CH). See Supported languages for a list of languages.	September 26, 2024
New voices added for NTTS	Amazon Polly now provides two new NTTS voices: Jitka and Sabrina. See Neural voices for a list of NTTS voices.	August 27, 2024
New generative voice engine added	Amazon Polly now offers a generative voice engine designed for longer content, with three English voices in a generative variant: Amy, Matthew, and Ruth. See Generative voices for more information.	March 28, 2024
New voice added for NTTS	Amazon Polly now provides the NTTS Turkish voice Burcu. See Neural voices for a list of NTTS voices.	February 14, 2024
New long-form voice engine added	Amazon Polly now offers a long-form voice engine designed for longer content, with three en-US voices: Danielle, Gregory, and Ruth. See Long-form voices for more information.	November 16, 2023

New voices added for NTTS	Amazon Polly now provides two new NTTS US English voices: Danielle and Gregory. See Neural voices for a list of NTTS voices.	October 5, 2023
Amazon Polly for Windows	The Amazon Polly Windows Speech Application Programming Interface (SAPI) plugin will no longer be supported.	September 26, 2023
Updated quota guidance for Amazon Polly	Updated Amazon Polly quotas guide. Added examples and clarification of terms. Refer to Quotas in Amazon Polly for the updates.	August 17, 2023
New voice added for NTTS	Amazon Polly now provides the Gulf Arabic NTTS voice Zayd. See Neural voices for a list of NTTS voices.	August 16, 2023
New voice added for NTTS	Amazon Polly now provides the Belgian French NTTS voice Isabelle. See <u>Neural voices</u> for a list of NTTS voices.	August 1, 2023
New voice added for NTTS	Amazon Polly now provides the Belgian Dutch (Flemish) NTTS voice Lisa. See Neural voices for a list of NTTS voices.	June 7, 2023

New voices added for NTTS	Amazon Polly now provides two new NTTS voices: Irish English (Niamh), and Danish (Sofie). See <u>Neural voices</u> for a list of NTTS voices.	May 30, 2023
Updated the IAM guidance for Amazon Polly	Updated guide to align with the IAM best practices . For more information, see Security best practices in IAM.	April 19, 2023
WordPress update	The Amazon Polly WordPress plugin will no longer be supported.	April 6, 2023
New Region added	Amazon Polly is now available in the Asia Pacific (Osaka) AWS Region. This Region supports neural TTS (NTTS). For more information, see Feature and Region Compatibility for a list of regions that support NTTS.	April 5, 2023
New voices added for NTTS	Amazon Polly now provides two new Japanese NTTS voices: Kazuha and Tomoko. See Neural voices for a list of NTTS voices.	February 7, 2023
New voices added for NTTS	Amazon Polly now provides two new US English NTTS voices: Stephen and Ruth. See Neural voices for a list of NTTS voices.	January 31, 2023

New voices added for NTTS	Amazon Polly now provides new NTTS voices for: Brazilian Portuguese (Thiago), Castilian Spanish (Sergio), French (Rémi), Italian (Adriano), and Mexican Spanish (Andrés). See Neural voices for a list of NTTS voices.	January 24, 2023
New voices added for NTTS	Amazon Polly now provides NTTS voices for Arabic (Hala) and Polish (Ola). See Neural voices for a list of NTTS voices.	November 17, 2022
Release AWS PrivateLink support	Amazon Polly now provides AWS PrivateLink support. See Using Amazon Polly with VPC endpoints to learn more.	November 9, 2022
New voices and languages added for NTTS	Amazon Polly now provides NTTS voices for Finnish (Suvi), Norwegian (Ida), and Swedish (Elin). See Neural voices for a list of NTTS voices.	November 8, 2022
New voice added for NTTS	Amazon Polly now provides the Dutch NTTS voice Laura. See Neural voices for a list of NTTS voices.	November 2, 2022

New Region added	Amazon Polly is now available in the Europe (Paris) AWS Region. This Region supports neural TTS (NTTS). For more information, see Feature and Region Compatibility for a list of regions that support NTTS.	September 22, 2022
New voice and language added for NTTS	Amazon Polly now provides the Cantonese NTTS voice Hiujin. See <u>Neural voices</u> for a list of NTTS voices.	September 20, 2022
New Region added	Amazon Polly is now available in the Asia Pacific (Mumbai) AWS Region. This Region supports neural TTS (NTTS). For more information, see Feature and Region Compatibility for a list of regions that support NTTS.	September 1, 2022
New voice added for NTTS	Amazon Polly now provides the Mandarin voice Zhiyu as an NTTS voice. See Neural voices for a list of NTTS voices.	August 23, 2022
New voice added for NTTS	Amazon Polly now provides the Hindi NTTS voice Kajal. See Neural voices for a list of NTTS voices.	July 27, 2022

New voices added for NTTS	Amazon Polly now provides NTTS voices for US Spanish (Pedro), German (Daniel), Canadian French (Liam), and UK English (Arthur). See Neural voices for a list of NTTS voices.	June 28, 2022
New voice added for NTTS	Amazon Polly now provides the Portuguese (Brazilian) voice Vitória as an NTTS voice. See Neural voices for a list of NTTS voices.	April 27, 2022
New voice added for NTTS	Amazon Polly now provides the Portuguese (European) voice Inês as an NTTS voice. See Neural voices for a list of NTTS voices.	April 26, 2022
New voice and language added for NTTS	Amazon Polly now provides the German (Austrian) language and the NTTS voice Hannah. See Neural voices for a list of NTTS voices.	April 19, 2022
New voices and language added for NTTS	Amazon Polly now provides the Spanish (Mexican) voice Mia as an NTTS voice. A new language, Catalan, was added along with the NTTS voice Arlet. See Neural voices for a list of NTTS voices.	March 22, 2022

New voice added for NTTS	Amazon Polly now provides the Japanese voice Takumi as an NTTS voice. See Neural voices for a list of NTTS voices.	December 6, 2021
New voice added for NTTS	Amazon Polly now provides the French voice Léa as an NTTS voice. See <u>Neural voices</u> for a list of NTTS voices.	November 18, 2021
New voices added for NTTS	Amazon Polly now provides the Italian voice Bianca and the European Spanish voice Lucia as NTTS voices. See Neural voices for a list of NTTS voices.	November 8, 2021
New voice added for NTTS	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.	September 1, 2021
New Region added	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.	September 1, 2021

New language and voice added	Amazon Polly now supports New Zealand English (en- NZ). A new NTTS voice, Aria, speaks New Zealand English and a selection of Maori words.	August 24, 2021
New feature	Amazon Polly makes the conversational speaking style the default version for the neural Matthew and Joanna voices. We removed references to the conversat ional speaking style.	June 28, 2021
New voice added for NTTS	Amazon Polly now provides the German voice Vicki as an NTTS voice.	June 15, 2021
New voice added	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.	June 1, 2021
New voice added for NTTS	Amazon Polly now provides the Korean voice Seoyeon as an NTTS voice.	May 11, 2021

New Region added for NTTS

Amazon Polly now supports neural TTS (NTTS) in the Canada (Central) AWS Region. For more information, see <u>Feature and Region Compatibi</u> lity for NTTS. March 17, 2021

New voice available for newscaster style

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.

November 10, 2020

New Regions added for NTTS

In addition to the existing Regions for NTTS (us-east-1, us-west-2, eu-west-1, and apsoutheast-2), neural voices are now supported in four additional Regions: (ap-north east-1 (Tokyo), ap-southe ast-1 (Singapore), eu-centra l-1 (Frankfurt), and eu-west-2 (London). For more information, see Feature and Region Compatibility for NTTS.

September 3, 2020

New voice added

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: useast-1 (N. Virginia), us-west-2
(Oregon), eu-west-1 (Ireland),
and ap-southeast-2 (Sydney).
For more information, see
NTTS Voices.

June 16, 2020

New voice available for newscaster style

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 Newscaste r style. For more information, see NTTS Speaking Styles.

April 16, 2020

New feature

In addition to the Newscaste r 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.

November 25, 2019

New voices added

Two new voices added: Camila (female, Portuguese-Brazil) and Lupe (female, Spanish-US).

October 23, 2019

New feature added

Addition of Amazon Polly for Windows plugin to incorpora te the full range of Amazon Polly voices into Windows SAPI-compliant applications.

September 26, 2019

Major new feature	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.	July 30, 2019
New voices added	New voices added: Lucia (female, Spanish), and Bianca (female, Italian).	August 2, 2018
New language added	New language added: Mexican Spanish (es-MX). This language uses the female voice of Mia.	August 2, 2018
New language added	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.	August 2, 2018
New feature added	Addition of <u>Speech synthesis</u> of long text passages (up to 100,000 billed characters).	July 17, 2018
New SSML feature added	Addition of Maximum Duration for Synthesized Speech.	July 17, 2018

New voice added	New voice added: Léa (female, French).	June 5, 2018
Region expansion	Expansion of Amazon Polly service to all commercial regions.	June 4, 2018
New language added	New language added: Korean (ko-KR).	June 4, 2018
Expanded feature	The Amazon Polly WordPress Plugin feature, including addition of Amazon Translate capabilities.	June 4, 2018
New voices added	Two new voices added: Aditi (female, Indian English) and Seoyeon (female, Korean).	November 15, 2017
New feature	Addition of new <u>Speech</u> <u>Marks</u> feature, as well as an expansion of <u>SSML</u> capabilit ies	April 19, 2017
New guide	This is the first release of the Amazon Polly Developer Guide.	November 30, 2016