Batch synthesis properties for text to speech

Article • 09/12/2024

> **Important**
>
> The Batch synthesis API is generally available. The Long Audio API will be retired on April 1st, 2027. For more information, see Migrate to batch synthesis API.

The Batch synthesis API can synthesize a large volume of text input (long and short) asynchronously. Publishers and audio content platforms can create long audio content in a batch. For example: audio books, news articles, and documents. The batch synthesis API can create synthesized audio longer than 10 minutes.

Some properties in JSON format are required when you create a new batch synthesis job. Other properties are optional. The batch synthesis response includes other properties to provide information about the synthesis status and results. For example, the `outputs.result` property contains the location of the batch synthesis result files with audio output and logs.

Batch synthesis properties

Batch synthesis properties are described in the following table.

Expand table

Property	Description
`createdDateTime`	The date and time when the batch synthesis job was created. This property is read-only.
`customVoices`	The map of a custom voice name and its deployment ID. For example: `"customVoices": {"your-custom-voice-name": "502ac834-6537-4bc3-9fd6-140114daa66d"}`

Property	Description
	You can use the voice name in your `synthesisConfig.voice` (when the `inputKind` is set to `"PlainText"`) or within the SSML text of `inputs` (when the `inputKind` is set to `"SSML"`). This property is required to use a custom voice. If you try to use a custom voice that isn't defined here, the service returns an error.
`description`	The description of the batch synthesis. This property is optional.
`id`	The batch synthesis job ID you passed in path. This property is required in path.
`inputs`	The plain text or SSML to be synthesized. When the `inputKind` is set to `"PlainText"`, provide plain text as shown here: `"inputs": [{"text": "The rainbow has seven colors."}]`. When the `inputKind` is set to `"SSML"`, provide text in the Speech Synthesis Markup Language (SSML) as shown here: `"inputs": [{"text": "<speak version='\''1.0'\'' xml:lang='\''en-US'\''><voice xml:lang='\''en-US'\'' xml:gender='\''Female'\'' name='\''en-US-AvaMultilingualNeural'\''>The rainbow has seven colors.</voice></speak>"}]`. Include up to 1,000 text objects if you want multiple audio output files. Here's example input text that should be synthesized to two audio output files: `"inputs": [{"text": "synthesize this to a file"},{"text": "synthesize this to another file"}]`. However, if the `properties.concatenateResult` property is set to `true`, then each synthesized result is written to the same audio output file. You don't need separate text inputs for new paragraphs. Within any of the (up to 1,000) text inputs, you can specify new paragraphs using the "\r\n" (newline) string. Here's example input text with two paragraphs that should be synthesized to the same audio output file:

Property	Description
	`"inputs": [{"text": "synthesize this to a file\r\nsynthesize this to another paragraph in the same file"}]` There are no paragraph limits, but the maximum JSON payload size (including all text inputs and other properties) is 2 megabytes. This property is required when you create a new batch synthesis job. This property isn't included in the response when you get the synthesis job.
`internalId`	The internal batch synthesis job ID. This property is read-only.
`lastActionDateTime`	The most recent date and time when the `status` property value changed. This property is read-only.
`outputs.result`	The location of the batch synthesis result files with audio output and logs. This property is read-only.
`properties`	A defined set of optional batch synthesis configuration settings.
`properties.sizeInBytes`	The audio output size in bytes. This property is read-only.
`properties.billingDetails`	The number of words that were processed and billed by `customNeuralCharacters` versus `neuralCharacters` (prebuilt) voices. This property is read-only.
`properties.concatenateResult`	Determines whether to concatenate the result. This optional `bool` value ("true" or "false") is "false" by default.
`properties.decompressOutputFiles`	Determines whether to unzip the synthesis result files in the destination container. This property can only be set when the `destinationContainerUrl` property is set. This

Property	Description
	optional `bool` value ("true" or "false") is "false" by default.
`properties.destinationContainerUrl`	The batch synthesis results can be stored in a writable Azure container. If you don't specify a container URI with shared access signatures (SAS) token, the Speech service stores the results in a container managed by Microsoft. SAS with stored access policies isn't supported. When the synthesis job is deleted, the result data is also deleted. This optional property isn't included in the response when you get the synthesis job.
`properties.destinationPath`	The prefix path where batch synthesis results can be stored with. If you don't specify a prefix path, the default prefix path is `YourSpeechResourceId/YourSynthesisId`. This optional property can only be set when the `destinationContainerUrl` property is set.
`properties.durationInMilliseconds`	The audio output duration in milliseconds. This property is read-only.
`properties.failedAudioCount`	The count of batch synthesis inputs to audio output failed. This property is read-only.
`properties.outputFormat`	The audio output format. For information about the accepted values, see audio output formats. The default output format is `riff-24khz-16bit-mono-pcm`.
`properties.sentenceBoundaryEnabled`	Determines whether to generate sentence boundary data. This optional `bool` value ("true" or "false") is "false" by default. If sentence boundary data is requested, then a corresponding `[nnnn].sentence.json` file is included in the results data ZIP file.

Property	Description
`properties.succeededAudioCount`	The count of batch synthesis inputs to audio output succeeded. This property is read-only.
`properties.timeToLiveInHours`	A duration in hours after the synthesis job is created, when the synthesis results will be automatically deleted. This optional setting is `744` (31 days) by default. The maximum time to live is 31 days. The date and time of automatic deletion (for synthesis jobs with a status of "Succeeded" or "Failed") is equal to the `lastActionDateTime + timeToLiveInHours` properties. Otherwise, you can call the delete synthesis method to remove the job sooner.
`properties.wordBoundaryEnabled`	Determines whether to generate word boundary data. This optional `bool` value ("true" or "false") is "false" by default. If word boundary data is requested, then a corresponding `[nnnn].word.json` file is included in the results data ZIP file.
`status`	The batch synthesis processing status. The status should progress from "NotStarted" to "Running", and finally to either "Succeeded" or "Failed". This property is read-only.
`synthesisConfig`	The configuration settings to use for batch synthesis of plain text. This property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.backgroundAudio`	The background audio for each audio output. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.backgroundAudio.fadein`	The duration of the background audio fade-in as milliseconds. The default value is `0`, which is the equivalent to no fade in. Accepted values: `0` to `10000`

Property	Description
	inclusive. For information, see the attributes table under add background audio in the Speech Synthesis Markup Language (SSML) documentation. Invalid values are ignored. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.backgroundAudio.fadeout`	The duration of the background audio fade-out in milliseconds. The default value is `0`, which is the equivalent to no fade out. Accepted values: `0` to `10000` inclusive. For information, see the attributes table under add background audio in the Speech Synthesis Markup Language (SSML) documentation. Invalid values are ignored. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.backgroundAudio.src`	The URI location of the background audio file. For information, see the attributes table under add background audio in the Speech Synthesis Markup Language (SSML) documentation. Invalid values are ignored. This property is required when `synthesisConfig.backgroundAudio` is set.
`synthesisConfig.backgroundAudio.volume`	The volume of the background audio file. Accepted values: `0` to `100` inclusive. The default value is `1`. For information, see the attributes table under add background audio in the Speech Synthesis Markup Language (SSML) documentation. Invalid values are ignored. This optional property is only applicable when `inputKind` is set to `"PlainText"`.

Property	Description
`synthesisConfig.pitch`	The pitch of the audio output. For information about the accepted values, see the adjust prosody table in the Speech Synthesis Markup Language (SSML) documentation. Invalid values are ignored. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.rate`	The rate of the audio output. For information about the accepted values, see the adjust prosody table in the Speech Synthesis Markup Language (SSML) documentation. Invalid values are ignored. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.role`	For some voices, you can adjust the speaking role-play. The voice can imitate a different age and gender, but the voice name isn't changed. For example, a male voice can raise the pitch and change the intonation to imitate a female voice, but the voice name isn't changed. If the role is missing or isn't supported for your voice, this attribute is ignored. For information about the available styles per voice, see voice styles and roles. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.speakerProfileId`	The speaker profile ID of a personal voice. For information about available personal voice base model names, see integrate personal voice. For information about how to get the speaker profile ID, see language and voice support. This property is required when `inputKind` is set to `"PlainText"`.

Property	Description
`synthesisConfig.style`	For some voices, you can adjust the speaking style to express different emotions like cheerfulness, empathy, and calm. You can optimize the voice for different scenarios like customer service, newscast, and voice assistant. For information about the available styles per voice, see voice styles and roles. This optional property is only applicable when `synthesisConfig.style` is set.
`synthesisConfig.styleDegree`	The intensity of the speaking style. You can specify a stronger or softer style to make the speech more expressive or subdued. The range of accepted values are: 0.01 to 2 inclusive. The default value is 1, which means the predefined style intensity. The minimum unit is 0.01, which results in a slight tendency for the target style. A value of 2 results in a doubling of the default style intensity. If the style degree is missing or isn't supported for your voice, this attribute is ignored. For information about the available styles per voice, see voice styles and roles. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`synthesisConfig.voice`	The voice that speaks the audio output. For information about the available prebuilt neural voices, see language and voice support. To use a custom voice, you must specify a valid custom voice and deployment ID mapping in the `customVoices` property. To use a personal voice, you need to specify the `synthesisConfig.speakerProfileId` property. This property is required when `inputKind` is set to `"PlainText"`.
`synthesisConfig.volume`	The volume of the audio output. For information about the accepted values, see the adjust prosody table in the Speech Synthesis Markup

Property	Description
	Language (SSML) documentation. Invalid values are ignored. This optional property is only applicable when `inputKind` is set to `"PlainText"`.
`inputKind`	Indicates whether the `inputs` text property should be plain text or SSML. The possible case-insensitive values are "PlainText" and "SSML". When the `inputKind` is set to `"PlainText"`, you must also set the `synthesisConfig` voice property. This property is required.

Batch synthesis latency and best practices

When using batch synthesis for generating synthesized speech, it's important to consider the latency involved and follow best practices for achieving optimal results.

Latency in batch synthesis

The latency in batch synthesis depends on various factors, including the complexity of the input text, the number of inputs in the batch, and the processing capabilities of the underlying hardware.

The latency for batch synthesis is as follows (approximately):

- The latency of 50% of the synthesized speech outputs is within 10-20 seconds.
- The latency of 95% of the synthesized speech outputs is within 120 seconds.

Best practices

When considering batch synthesis for your application, it's recommended to assess whether the latency meets your requirements. If the latency aligns with your desired performance, batch synthesis can be a suitable choice. However, if the latency doesn't meet your needs, you might consider using real-time API.

HTTP status codes

The section details the HTTP response codes and messages from the batch synthesis API.

HTTP 200 OK

HTTP 200 OK indicates that the request was successful.

HTTP 201 Created

HTTP 201 Created indicates that the create batch synthesis request (via HTTP POST) was successful.

HTTP 204 error

An HTTP 204 error indicates that the request was successful, but the resource doesn't exist. For example:

- You tried to get or delete a synthesis job that doesn't exist.
- You successfully deleted a synthesis job.

HTTP 400 error

Here are examples that can result in the 400 error:

- The `outputFormat` is unsupported or invalid. Provide a valid format value, or leave `outputFormat` empty to use the default setting.
- The number of requested text inputs exceeded the limit of 10,000.
- You tried to use an invalid deployment ID or a custom voice that isn't successfully deployed. Make sure the Speech resource has access to the custom voice, and the custom voice is successfully deployed. You must also ensure that the mapping of `{"your-custom-voice-name": "your-deployment-ID"}` is correct in your batch synthesis request.
- You tried to use a *F0* Speech resource, but the region only supports the *Standard* Speech resource pricing tier.

HTTP 404 error

The specified entity can't be found. Make sure the synthesis ID is correct.

HTTP 429 error

There are too many recent requests. Each client application can submit up to 100 requests per 10 seconds for each Speech resource. Reduce the number of requests per second.

HTTP 500 error

HTTP 500 Internal Server Error indicates that the request failed. The response body contains the error message.

HTTP error example

Here's an example request that results in an HTTP 400 error, because the `inputs` property is required to create a job.

Console

```
curl -v -X PUT -H "Ocp-Apim-Subscription-Key: YourSpeechKey" -H "Content-Type:
application/json" -d '{
    "inputKind": "SSML"
}'
"https://YourSpeechRegion.api.cognitive.microsoft.com/texttospeech/batchsynthes
es/YourSynthesisId?api-version=2024-04-01"
```

In this case, the response headers include `HTTP/1.1 400 Bad Request`.

The response body resembles the following JSON example:

JSON

```
{
  "error": {
    "code": "BadRequest",
    "message": "The inputs is required."
  }
}
```

Next steps

- Speech Synthesis Markup Language (SSML)
- Text to speech quickstart
- Migrate to batch synthesis

Feedback

Was this page helpful? Yes No

Provide product feedback | Get help at Microsoft Q&A