Batch synthesis API for text to speech

Article • 09/12/2024

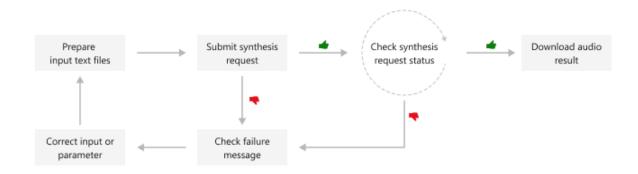
The Batch synthesis API can synthesize a large volume of text input (long and short) asynchronously. Publishers and audio content platforms can create long audio content in a batch. For example: audio books, news articles, and documents. The batch synthesis API can create synthesized audio longer than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

(i) Important

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

The batch synthesis API is asynchronous and doesn't return synthesized audio in real-time. You submit text files to be synthesized, poll for the status, and download the audio output when the status indicates success. The text inputs must be plain text or Speech Synthesis Markup Language (SSML) text.

This diagram provides a high-level overview of the workflow.



You can also use the <u>Speech SDK</u> to create synthesized audio longer than 10 minutes by iterating over the text and synthesizing it in chunks. For a C# example, see **GitHub** .

You can use the following REST API operations for batch synthesis:



Operation	Method	REST API call
Create batch synthesis	PUT	texttospeech/batchsyntheses/YourSynthesisId
Get batch synthesis	GET	texttospeech/batchsyntheses/YourSynthesisId
List batch synthesis	GET	texttospeech/batchsyntheses
Delete batch synthesis	DELETE	texttospeech/batchsyntheses/YourSynthesisId

For code samples, see GitHub

Create batch synthesis

To submit a batch synthesis request, construct the HTTP PUT request path and body according to the following instructions:

- Set the required inputKind property.
- If the inputKind property is set to "PlainText", then you must also set the voice property in the synthesisConfig. In the example below, the inputKind is set to "SSML", so the synthesisConfig isn't set.
- Optionally you can set the description, timeToLiveInHours, and other properties. For more information, see batch synthesis properties.

① Note

The maximum JSON payload size that's accepted is 2 megabytes.

Set the required YourSynthesisId in path. The YourSynthesisId must be unique. It must be 3-64 long, contains only numbers, letters, hyphens, underscores and dots, starts and ends with a letter or number.

Make an HTTP PUT request using the URI as shown in the following example. Replace YourSpeechKey with your Speech resource key, replace YourSpeechRegion with your Speech resource region, and set the request body properties as previously described.

Azure CLI

```
curl -v -X PUT -H "Ocp-Apim-Subscription-Key: YourSpeechKey" -H "Content-Type:
application/json" -d '{
    "description": "my ssml test",
    "inputKind": "SSML",
    "inputs": [
            "content": "<speak version=\"1.0\" xml:lang=\"en-US\"><voice</pre>
name=\"en-US-JennyNeural\">The rainbow has seven colors.</voice></speak>"
    ],
    "properties": {
        "outputFormat": "riff-24khz-16bit-mono-pcm",
        "wordBoundaryEnabled": false,
        "sentenceBoundaryEnabled": false,
        "concatenateResult": false,
        "decompressOutputFiles": false
    }
}'
"https://YourSpeechRegion.api.cognitive.microsoft.com/texttospeech/batchsynthes
es/YourSynthesisId?api-version=2024-04-01"
```

You should receive a response body in the following format:

```
JSON
  "id": "YourSynthesisId",
  "internalId": "7ab84171-9070-4d3b-88d4-1b8cc1cb928a",
  "status": "NotStarted",
  "createdDateTime": "2024-03-12T07:23:18.0097387Z",
  "lastActionDateTime": "2024-03-12T07:23:18.0097388Z",
  "inputKind": "SSML",
  "customVoices": {},
  "properties": {
    "timeToLiveInHours": 744,
    "outputFormat": "riff-24khz-16bit-mono-pcm",
    "concatenateResult": false,
    "decompressOutputFiles": false,
    "wordBoundaryEnabled": false,
    "sentenceBoundaryEnabled": false
  }
}
```

The status property should progress from NotStarted status, to Running, and finally to Succeeded or Failed. You can call the GET batch synthesis API periodically until the returned status is Succeeded or Failed.

Get batch synthesis

To get the status of the batch synthesis job, make an HTTP GET request using the URI as shown in the following example. Replace YourSpeechKey with your Speech resource key, and replace YourSpeechRegion with your Speech resource region.

```
curl -v -X GET
  "https://YourSpeechRegion.api.cognitive.microsoft.com/texttospeech/batchsynthes
  es/YourSynthesisId?api-version=2024-04-01" -H "Ocp-Apim-Subscription-Key:
  YourSpeechKey"
```

You should receive a response body in the following format:

```
JSON
  "id": "YourSynthesisId",
  "internalId": "7ab84171-9070-4d3b-88d4-1b8cc1cb928a",
  "status": "Succeeded",
  "createdDateTime": "2024-03-12T07:23:18.0097387Z",
  "lastActionDateTime": "2024-03-12T07:23:18.7979669",
  "inputKind": "SSML",
  "customVoices": {},
  "properties": {
    "timeToLiveInHours": 744,
    "outputFormat": "riff-24khz-16bit-mono-pcm",
    "concatenateResult": false,
    "decompressOutputFiles": false,
    "wordBoundaryEnabled": false,
    "sentenceBoundaryEnabled": false,
    "sizeInBytes": 120000,
    "succeededAudioCount": 1,
    "failedAudioCount": 0,
    "durationInMilliseconds": 2500,
    "billingDetails": {
      "neuralCharacters": 29
    }
  },
  "outputs": {
    "result": "https://stttssvcuse.blob.core.windows.net/batchsynthesis-
output/29f2105f997c4bfea176d39d05ff201e/YourSynthesisId/results.zip?SAS_Token"
  }
}
```

From outputs.result, you can download a ZIP file that contains the audio (such as 0001.wav), summary, and debug details. For more information, see batch synthesis results.

List batch synthesis

To list all batch synthesis jobs for the Speech resource, make an HTTP GET request using the URI as shown in the following example. Replace YourSpeechKey with your Speech resource key and replace YourSpeechRegion with your Speech resource region. Optionally, you can set the skip and maxpagesize (up to 100) query parameters in URL. The default value for skip is 0 and the default value for maxpagesize is 100.

```
curl -v -X GET
  "https://YourSpeechRegion.api.cognitive.microsoft.com/texttospeech/batchsynthes
  es?api-version=2024-04-01&skip=1&maxpagesize=2" -H "Ocp-Apim-Subscription-Key:
  YourSpeechKey"
```

You should receive a response body in the following format:

```
JSON
  "value": [
      "id": "my-job-03",
      "internalId": "5f7e9ab6-2c92-4dcb-b5ee-ec0983ee4db0",
      "status": "Succeeded",
      "createdDateTime": "2024-03-12T07:28:32.5690441Z",
      "lastActionDateTime": "2024-03-12T07:28:33.0042293",
      "inputKind": "SSML",
      "customVoices": {},
      "properties": {
        "timeToLiveInHours": 744,
        "outputFormat": "riff-24khz-16bit-mono-pcm",
        "concatenateResult": false,
        "decompressOutputFiles": false,
        "wordBoundaryEnabled": false,
        "sentenceBoundaryEnabled": false,
        "sizeInBytes": 120000,
        "succeededAudioCount": 1,
        "failedAudioCount": 0,
        "durationInMilliseconds": 2500,
        "billingDetails": {
          "neuralCharacters": 29
```

```
}
      },
      "outputs": {
        "result": "https://stttssvcuse.blob.core.windows.net/batchsynthesis-
output/29f2105f997c4bfea176d39d05ff201e/my-job-03/results.zip?SAS_Token"
      }
    },
    {
      "id": "my-job-02",
      "internalId": "5577585f-4710-4d4f-aab6-162d14bd7ee0",
      "status": "Succeeded",
      "createdDateTime": "2024-03-12T07:28:29.6418211Z",
      "lastActionDateTime": "2024-03-12T07:28:30.0910306",
      "inputKind": "SSML",
      "customVoices": {},
      "properties": {
        "timeToLiveInHours": 744,
        "outputFormat": "riff-24khz-16bit-mono-pcm",
        "concatenateResult": false,
        "decompressOutputFiles": false,
        "wordBoundaryEnabled": false,
        "sentenceBoundaryEnabled": false,
        "sizeInBytes": 120000,
        "succeededAudioCount": 1,
        "failedAudioCount": 0,
        "durationInMilliseconds": 2500,
        "billingDetails": {
          "neuralCharacters": 29
        }
      },
      "outputs": {
        "result": "https://stttssvcuse.blob.core.windows.net/batchsynthesis-
output/29f2105f997c4bfea176d39d05ff201e/my-job-02/results.zip?SAS Token"
    }
  ],
  "nextLink":
"https://YourSpeechRegion.api.cognitive.microsoft.com/texttospeech/batchsynthes
es?skip=3&maxpagesize=2&api-version=2024-04-01"
}
```

From outputs.result, you can download a ZIP file that contains the audio (such as 0001.wav), summary, and debug details. For more information, see batch synthesis results.

The value property in the json response lists your synthesis requests. The list is paginated, with a maximum page size of 100. The "nextLink" property is provided as needed to get the next page of the paginated list.

Delete batch synthesis

Delete the batch synthesis job history after you retrieved the audio output results. The Speech service keeps batch synthesis history for up to 31 days, or the duration of the request timeToLiveInHours property, whichever comes sooner. The date and time of automatic deletion (for synthesis jobs with a status of "Succeeded" or "Failed") is equal to the lastActionDateTime + timeToLiveInHours properties.

To delete a batch synthesis job, make an HTTP DELETE request using the URI as shown in the following example. Replace YourSynthesisId with your batch synthesis ID, replace YourSpeechKey with your Speech resource key, and replace YourSpeechRegion with your Speech resource region.

```
curl -v -X DELETE
  "https://YourSpeechRegion.api.cognitive.microsoft.com/texttospeech/batchsynthes
  es/YourSynthesisId?api-version=2024-04-01" -H "Ocp-Apim-Subscription-Key:
  YourSpeechKey"
```

The response headers include HTTP/1.1 204 No Content if the delete request was successful.

Batch synthesis results

After you get a batch synthesis job with status of "Succeeded", you can download the audio output results. Use the URL from the outputs.result property of the batch synthesis GET response.

To get the batch synthesis results file, make an HTTP GET request using the URI as shown in the following example. Replace YourOutputsResultUrl with the URL from the outputs.result property of the batch synthesis GET response. Replace YourSpeechKey with your Speech resource key.

```
Azure CLI

curl -v -X GET "YourOutputsResultUrl" -H "Ocp-Apim-Subscription-Key:
YourSpeechKey" > results.zip
```

The results are in a ZIP file that contains the audio (such as <code>0001.wav</code>), summary, and debug details. The numbered prefix of each filename (shown below as <code>[nnnn]</code>) is in the same order as the text inputs used when you created the batch synthesis.

① Note

The [nnnn].debug.json file contains the synthesis result ID and other information that might help with troubleshooting. The properties that it contains might change, so you shouldn't take any dependencies on the JSON format.

The summary file contains the synthesis results for each text input. Here's an example summary.json file:

If sentence boundary data was requested ("sentenceBoundaryEnabled": true), then a corresponding [nnnn].sentence.json file is included in the results. Likewise, if word boundary data was requested ("wordBoundaryEnabled": true), then a corresponding [nnnn].word.json file is included in the results.

Here's an example word data file with both audio offset and duration in milliseconds:

```
JSON
```

```
"Text": "The",
    "AudioOffset": 50,
    "Duration": 137
  },
    "Text": "rainbow",
    "AudioOffset": 200,
    "Duration": 350
  },
    "Text": "has",
    "AudioOffset": 562,
    "Duration": 175
  },
    "Text": "seven",
    "AudioOffset": 750,
    "Duration": 300
  },
    "Text": "colors",
    "AudioOffset": 1062,
    "Duration": 625
  },
    "Text": ".",
    "AudioOffset": 1700,
    "Duration": 100
  }
]
```

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 batch synthesis create request (via HTTP PUT) 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 *FO* 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)
- Batch synthesis properties
- Migrate to batch synthesis

Feedback

Provide product feedback

| Get help at Microsoft Q&A