

ReadSpeaker® speechCloud API

ReadSpeaker speechCloud API (SCAPI) is a cloud-based API service that converts text provided by the customer's application into synthetic speech. Customer systems send text input and audio configuration parameters to SCAPI over secure HTTPS, and the service returns the generated speech audio, along with optional timing information, in the HTTPS response.

Product Description

ReadSpeaker® speechCloud API (SCAPI) allows customer applications to generate synthetic speech by sending text and configuration parameters to the SCAPI service. SCAPI returns the corresponding speech audio and, if requested, timing information that can be used for text synchronization features such as highlighting or bookmarking.

A customer-specific API key is required and must be included with every request. The audio returned by SCAPI may be streamed from the customer's application to end users in accordance with the terms defined in the order confirmation. Unless explicitly permitted in the order confirmation, the generated audio may not be stored, reused, or further distributed in any form.

The order confirmation specifies the licensed usage capacity for SCAPI. Capacity may be defined in one or more ways, such as number of characters, number of requests, total speech time, number of concurrent real-time channels, or generation speed.

It also specifies the time period during which this capacity can be consumed; if no period is stated, the default is one year. When a capacity limit is exceeded, SCAPI returns an error response instead of audio.

If a generation-speed limitation applies, SCAPI returns speech output as quickly as the service allows. The corresponding channel then remains occupied for the duration of the generated audio's playback time.

SCAPI uses ReadSpeaker text-to-speech technology. Customers may license one or multiple languages, and one or multiple voices per language, as specified in the order confirmation.

Features

Text to Speech

SCAPI generates natural-sounding synthetic speech from text and supports more than 50 languages and over 200 voices. The specific languages and voices licensed to a customer are defined in the order confirmation. Customers can view their available languages and voices in the SCAPI customer portal, or they may retrieve this information by sending a request to the API, which will return the languages and voices available to the customer. A complete list of available languages and voices is also published on the [ReadSpeaker website](#).

Settings for Audio Generation

Each request to SCAPI can include parameters that determine how the audio output is generated. The available parameters include:

- language
- voice
- volume
- pitch
- speaking rate

- audio format
- audio container
- sample rate
- bit depth
- bitrate (for mp3)
- streaming
 - value **1** (default): SCAPI returns a redirect to a streaming server that begins streaming the audio as soon as possible
 - value **0**: SCAPI returns the audio only after the full audio file has been generated
- use of a customer-specific pronunciation dictionary, if applicable

Pronunciation Dictionaries

SCAPI can use pronunciation dictionaries to adjust how specific words or word sequences are pronounced. These dictionaries are voice-specific. ReadSpeaker's linguistic team can correct mispronunciations in the standard dictionaries for supported languages upon request. Customer-specific pronunciation requirements can be added to personalized dictionaries. These personalized dictionaries are accessible to the customer and can be viewed and edited directly through the SCAPI customer portal.

Timing Information

The customer's server can send a request to SCAPI for timing information associated with the generated speech. This information links the audio to the corresponding segments of the input text. Timing data can be used to implement features such as highlighting text during playback or bookmarking so users can resume playback from a specific position.

Credits

To use SCAPI, customers require either production credits or trial credits. New accounts typically receive a limited amount of trial credits for initial testing. Customers may order production credits through their ReadSpeaker Account Manager; these credits are then added to the account by the ReadSpeaker support team.

Under the standard pricing model, one credit is charged per character when SCAPI converts text into audio. Characters are counted excluding whitespace and including any changes resulting from pronunciation dictionary entries. If a different pricing model has been agreed with the ReadSpeaker Account Manager, the order confirmation may specify an alternative rate or indicate that billing is based on the seconds of speech generated rather than on the number of characters.

Usage Statistics

Customers can monitor their consumed capacity through the SCAPI customer portal or retrieve usage information through the REST API.

SCAPI Customer Portal

Customers receive login credentials from ReadSpeaker to access the SCAPI customer portal. The portal provides account information, usage information, and implementation instructions. It also provides access to the customer's personalized pronunciation dictionary, which can be viewed and edited through the portal. If necessary, the customer can generate a new API key in the event of suspected misuse of the existing key.

REST API — Available Languages, Voices, and Credits

The customer's server can send a request to the REST API to retrieve:

- the languages and voices available to the customer, returned in JSON
- usage information, returned in JSON
- the number of remaining credits, returned in JSON

Product Support

ReadSpeaker's support team is available to assist customers with questions related to the implementation and use of SCAPI. The team can provide guidance on configuration, usage, and troubleshooting when needed. Customers may contact ReadSpeaker support using the channels provided in their onboarding information.

Technical Specifications

Fully Server-based

The customer does not have to download or install any software for SCAPI to work. The customer does not need to host SCAPI, as it is cloud-based and hosted by ReadSpeaker.

Product Hosting

SCAPI is hosted by ReadSpeaker in a high-security environment with a high level of redundancy and failover, and with the following properties:

- Double redundant enterprise load balancing and failover
- 100% Gigabit network (internally and to the Internet)
- Redundant power supply

- Highly scalable and high-availability design
- Environmentally friendly
- UPS and diesel engine power backup
- 24/7 video and audio surveillance
- Level 3 security
- Argonite-based automatic fire-fighting system
- automatic server monitoring
- automatic daily backup

API Parameters and Supported Input and Output Formats

Requests sent to the ReadSpeaker server include parameters as defined in the product documentation. Parameters that determine audio output characteristics are described in *Settings for audio generation*. Additional parameters specify characteristics of the input data:

- **Text or SSML:** SCAPI supports SSML, which allows greater control over how the text is rendered by the TTS engine, including prosody, pauses, phonetic transcriptions, and switching voices or languages within the same text.
- **Charset:** The character encoding of the input text. Common encodings are supported, and the use of UTF-8 is encouraged.

Parameters or accepted parameter values may be updated over time. Customers should refer to the product documentation for the most current information.

Supported Audio Formats

Exported audio files are available in the following formats:

- mp3 with bitrates of 16, 24, 32, 48, 64, or 128 kbit/s
- Ogg Vorbis (variable bitrate)
- WAV (U-law 16 bit/8 KHz, A-law 16 bit/8 KHz)
- PCM with sample rates of 22050, 16000, or 8000 Hz - sample depth 16 bit
- WAV with sample rates of 22050, 16000, or 8000 Hz - sample depth 16 bit

Additional formats may become available and will be listed in the product documentation. Audio formats requiring a specific file header (such as WAV, U-law, and A-law) are fully generated before being sent to the customer's application. Other formats are sent in buffers for the shortest possible time-to-audio.

Timing Information

Timing information can be requested in two different ways:

1. **Separate request**

The customer's server can send a separate request to the ReadSpeaker server solely for timing information. This request returns a JSON object containing the timing data. This call withdraws the same amount of credits as an audio generation request.

2. **RSDS audio format**

The customer's server can request the **RSDS** audio format, which includes both the audio and the timing information in a single response. The RSDS response must be parsed by the customer to obtain the audio and timing data separately. Because only one request is required, no additional credits are charged beyond the cost of generating the audio.

Sample Code

Sample code for using SCAPI in various popular programming languages is available in the SCAPI customer portal.

Product Components

SaaS Application

The core of SCAPI is the ReadSpeaker Software as a Service (SaaS) application, hosted by ReadSpeaker.

Enabling Code

An API key must be included with every API call. The key is available in the SCAPI customer portal, where the customer can also generate a new key if misuse of the existing key is suspected.

Product Documentation

Implementation and user instructions are available to the customer through the SCAPI customer portal.

Access to the SCAPI Customer Portal

The customer receives a username, password, and URL to access the SCAPI customer portal. The portal provides account information, usage information, remaining credits, access to the personalized pronunciation dictionary, and implementation and user instructions.

REST API

Communication between the customer's server and the ReadSpeaker server occurs through a REST API over HTTPS.

Terms of Use - speechCloud API

The Terms of Use governing SCAPI are available to the customer through the SCAPI customer portal.

Options

Options are additional features that can be enabled for SCAPI at an added cost.

Requirements and Limitations

Enable Code

The customer-specific API key must be included in every request sent to the ReadSpeaker server.

Text Limit

The text included in a request sent to the ReadSpeaker server has a default limit of 5,000 characters (excluding whitespace) per request, unless otherwise specified in the order confirmation.

Limited Usage

SCAPI is not intended for usage exceeding 25,000 requests per day or for peaks exceeding 2,000 requests within a single hour.

Timing Information and Audio Formats

When requesting the RSDS audio format to obtain audio and timing information in a single response, the audio can only be extracted in MP3 format at 48 kbit/s. Other audio formats are not available through RSDS. If another audio format is required, timing information must be requested through a separate call.

Timing Information and Available Voices

The RSDS format is not available for all languages and voices. Customers should consult their ReadSpeaker Account Manager for information on supported languages and voices.

Intended Use

SCAPI is a licensed text-to-speech service and may only be used in accordance with this product specification. It may be used to send text originating from the customer's online application to the ReadSpeaker server and to receive audio generated from that text. This audio may be played back to end users of the customer's applications, subject to the limitations set out in the *Terms of Use — speechCloud API* or as specifically stated in the order confirmation.

Any other use of SCAPI is not considered intended use. The audio data may not be stored, reused, or distributed online or offline unless explicitly permitted in the order confirmation, which will specify the context in which audio may be produced and used.

Disclaimer

The features described in this document are guaranteed only when SCAPI is implemented in accordance with the product documentation provided. Specifications and features described in this product specification may be changed by ReadSpeaker without prior notice.