info@readspeaker.com
www.readspeaker.com

ReadSpeaker® speechCloud API

ReadSpeaker speechCloud API (SCAPI) is a server-based web service that converts text sent to the ReadSpeaker server into speech/audio using text-to-speech technology (TTS), which is then sent back to the customer's server.

Product Description

The Customer implements SCAPI using the implementation instructions provided. This results in SCAPI being integrated into the customer's online application, which can send text with a number of parameters to the ReadSpeaker server (request) and receive audio and timing information from the ReadSpeaker server. The customer-specific API key must always be included in the request. The audio data received may be streamed from the customer's application to the end user for playback in accordance with the order confirmation. This audio data may not be stored, reused, or further distributed online or offline unless specifically stated in the Order Confirmation.

The order confirmation specifies the licensed capacity for SCAPI, measured in one or more of the following ways: number of characters, requests per month, speech time (in hours), number of concurrent real-time channels, and generation speed. The order confirmation also states the time frame during which the specified capacity can be consumed. If no timeframe is specified, the default time frame is one year. If any capacity limitation is exceeded, SCAPI will respond with an error message instead of audio data.

In case of generation speed limitations, the speech output is returned to the customer's application as fast as the ReadSpeaker server allows. The channel will then be blocked for the remaining duration of generated audio play time.

The audio is generated using ReadSpeaker text-to-speech technology. The customer can purchase licenses for one or multiple languages and for one or multiple voices in each language.

Features

Text to speech

SCAPI can generate audio from text using state-of-the-art text to speech (TTS) in 50+ languages and 200+ voices. For a complete list of available languages go to our languages web page. The languages and voices available to the customer are specified in the order confirmation. The customer can see these languages and voices in the SCAPI customer portal or the customer can send a request to the REST API which will return which languages and voices are available for use.

Settings for audio generation

The text sent to the ReadSpeaker server can be accompanied by a number of parameters, through which the following characteristics of the audio output can be determined:

- language
- voice
- volume
- pitch
- speaking rate
- audio format
- audio container
- sample rate

- bit depth
- bitrate (for mp3)
- streaming
  value 1 (default) = you get a redirect to a streaming server that starts to stream the audio as soon as possible.
  value 0 = the audio is sent to you after the audio file is completely generated;
- whether or not to use the customer-specific pronunciation dictionary

Pronunciation dictionaries

The dictionaries SCAPI uses can be customized to pronounce words or sequences of words in a specific way to increase the quality of the audio. The dictionaries are voice-specific. The linguists at ReadSpeaker can correct mispronunciations in the dictionaries for a specific language used upon request. If a different pronunciation is customer specific, it will be edited in the customer's personalized dictionaries. The customer also has access to these editable customized pronunciation dictionaries via the SCAPI customer portal.

Timing information

The customer's server can also send a request to the ReadSpeaker server for timing information to be sent. This can be used to couple the audio output SCAPI generated to the text input the audio was generated from. This timing information allows the customer to develop useful features for end users such as highlighting - the customer's application can highlight text as it is being read aloud - and bookmarks - to hold the spot in the text where the user last listened to continue from there later.

Usage statistics

The customer can monitor the consumed capacity via the SCAPI customer portal or via the REST API (see below).

SCAPI customer portal

The customer can access a web application service platform (WASP) using login credentials supplied by ReadSpeaker to access the SCAPI customer portal. The portal shows account information, usage information and clear implementation and user instructions. It also provides access to the editable personalized pronunciation dictionary and allows the user to generate a new API key if there has been suspected misuse of the key.

REST API - available languages/voices/credits

The customer can send a request to a REST API to:

- receive information in JSON on which languages and voices are available for use
- receive information in JSON on usage
- receive information in JSON on how many remaining credits are available

Product support

ReadSpeaker has an excellent support team to assist the customer where necessary.

Technical Specifications

Fully server-based, no software downloads or installation

The customer does not have to download or install any software for SCAPI to work. The customer does not need to host SCAPI since 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.

API Parameters and supported input and output formats

Requests sent to the ReadSpeaker server are accompanied by a number of parameters as defined in the product documentation. Those which determine characteristics of the audio output are mentioned above under “Settings for audio generation”. The remaining parameters specify characteristics of the input data:

- Text or SSML: SCAPI supports SSML. With SSML, the user has more control over how the text is read by the TTS engine, specifically the prosody of the speech. SSML can be used to insert pauses/breaks of certain length, insert phonetic transcriptions and switch voices and languages in the same text.
- Charset: the character encoding of the input data. All the most common encodings are supported, but using UTF-8 is encouraged.

Parameters or accepted values for the parameters may be added over time. Please refer to the product documentation for up-to-date information.

Supported audio output formats

SCAPI supports creating output audio in different formats:

- MP3 with bitrate 16, 24, 32, 48, 64, or 128 kbit/s
- Ogg variable bitrate
- U-Law
- A-Law
- PCM with sample rate 22050, 16000 or 8000 Hz - sample depth 16 bit
- WAV with sample rate 22050, 16000 or 8000 Hz - sample depth 16 bit

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

Timing information

The timing information can be requested in two different ways:

1. in a separate call to the server. This will return a json object with timing information. Note that this will withdraw the same amount of credits as the audio call does.

2. By requesting the audio format RSDS, which contains both the audio and the timing information. The RSDS response needs to be parsed by the customer to retrieve the audio and timing information separately. With this method, only one call to our server is necessary, so there's no extra cost in credits.

Sample code

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

Product Components

SaaS application

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

Enabling code

An API key which must be included with each API call. This key can be found in the SCAPI customer portal and a new key can be regenerated if there has been suspected misuse of the key.

Product documentation

Extensive implementation and user instructions are available to the customer via the SCAPI customer portal.

Access to the SCAPI customer portal

The customer receives a username and password and a url for the ReadSpeaker Web Application Service Platform (WASP) to login to the SCAPI customer portal. Here, the customer can access account information, information on the use of SCAPI and the remaining credits, the personalized pronunciation dictionary, and clear implementation and user instructions.

REST API

The communication between the customer's server and the ReadSpeaker server is via a REST API over the Internet using HTTP(s) protocol.

Terms of Use - speechCloud API

A document governing the terms of use of the product, which is available to the customer via the SCAPI customer portal.

Options

Options are features that can be added to SCAPI at a 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 to be converted to audio, which is included in the request sent to the ReadSpeaker server, has a default limit of 1000 characters (excluding white spaces) per request unless otherwise specified in the order confirmation.

Limited usage

SCAPI is not intended for high usage of over 25,000 requests per day or peaks of above 2,000 re.

Timing information and audio formats

If the audio format RSDS is requested in order to receive audio and timing information with one request, the audio can be extracted in mp3 format 48kbps only. Other audio formats are not available through RSDS. If another audio format is required, the timing information must be requested through a separate request.

Timing information and available voices

The RSDS format is currently not available for all languages and voices. Please check with your ReadSpeaker Account Manager for the available languages and voices.

Intended Use

Customers may only use SCAPI, which is a licensed service for text to speech, in accordance with this product specification. It may be used to send text originating from an online application from the customer's server to the ReadSpeaker server and receive audio generated from that text. This audio may be played back to end users of the customer's applications with limitations set forth in the Terms of Use - speechCloud API or specifically defined in the order confirmation.
Any other use of SCAPI is not considered intended use. The audio data may not be stored, re-used, or further distributed online or offline unless specifically stated in the order confirmation which will contain a clause determining the context within which the audio files may be produced and the context within which they may be used.

Disclaimer

Features listed in this document are guaranteed only if SCAPI has been implemented according to the product documentation provided. Specifications and features as described in this product specification can be changed by the manufacturer without prior notice.