ReadSpeaker
pioneering voice technology

info@readspeaker.com
www.readspeaker.com

ReadSpeaker® speechServer

Network-based text-to-speech solution allowing multiple *client applications in a client–server architecture* to request speech synthesis.

Product Description

ReadSpeaker speechServer ("speechServer") is a client–server text-to-speech (TTS) solution that makes the ReadSpeaker TTS engine, server-side APIs, and tools available to applications over the network, enabling scalable, dynamic multithreaded synthetic speech generation. Client applications can communicate with speechServer through a TCP/IP communication protocol, a command-line tool, or a REST API (when deployed behind a CGI-enabled web server).

The solution provides a complete server-side TTS runtime consisting of:

- The ReadSpeaker TTS engine
- A Server API and supporting tools for integration and control
- Documentation

speechServer supports multiple simultaneous synthetic speech requests, with concurrency limited only by the licensed number of ports and available system resources. It delivers natural-sounding, high-quality synthetic speech and is suitable for IVR systems, virtual assistants, alert and notification platforms, educational applications, and other environments requiring efficient, deterministic, server-based text-to-speech output.

Features

High-Quality Synthetic Speech

ReadSpeaker's text-to-speech voices deliver clear, natural, human-like synthetic speech and are continually optimized to ensure consistent quality and reliable output.

150+ Voices in 50+ Languages

speechServer supports ReadSpeaker's catalog of more than 150 synthetic voices across over 50 languages. New voices and languages are added regularly. A complete list is available on the ReadSpeaker website.

Modifiable Speaking Rate, Pitch, and Volume (DSP)

The speaking rate, pitch, and volume can be configured as desired.

Voice and Language Switch

speechServer supports switching to another language or to another voice in the same language during text-to-speech conversion, based on SSML instructions in the input.

User Dictionary and IPA Support

Each licensed language includes a user dictionary where customers can define custom pronunciations for words or patterns. This improves the reading accuracy for domain-specific vocabulary. Multiple user dictionaries per language and IPA transcription input are supported.

Audio Clip Insertion Supported

speechServer allows for links to audio clips (files) to be inserted in the input text so that the referenced audio is inserted at the correct position within the synthetic speech output streamed back to the client application.

Mark Information for Event Notification

speechServer allows for mark information to be created alongside the output audio to allow:

- Event triggers at specific text positions
- Device or interface synchronization
- Text highlighting aligned with audio playback

Technical Specifications

Supported Operating Systems

Windows

- Windows Server 2019 and higher (most recently tested: Windows Server 2025)

Linux

- CentOS Stream 9.x, 10.x
- Ubuntu 18.x – 24.x
- RHEL 8.x, 9.x
- Rocky Linux 9.x

CPU Requirements

- Intel x64, 1 GHz or higher
  Note: If the AVX instruction set is not supported, please consult with ReadSpeaker.

RAM Requirements

- 4 GB or more recommended (depending on the number of channels)

Voice Footprint

- 9 MB – 35 MB per voice (Deep Neural Networks technology)

Runtime Memory Usage

- 100 MB for basic load
- 35 MB per voice
- 2 MB – 90 MB per channel

Development Languages

- C / C++
- Java
- C# / .NET (Windows)

Supported Input Formats

- Plain text
- SSML

Supported Character Encoding for Text Input

- Plain text: Multibyte and UTF-8 encodings
- SSML: UTF-8 encoding

Supported Audio Formats

- 16-bit linear PCM
- 16-bit linear PCM Wave
- 8-bit A-law PCM
- 8-bit A-law PCM Wave
- 8-bit μ-law PCM
- 8-bit μ-law PCM Wave

- 8-bit unsigned linear PCM Wave
- 4-bit Dialogic ADPCM
- mp3 (requires downloading and installating the LAME package)
- OGG

TCP/IP Protocol

speechServer supports the use of TCP/IP protocol for communication between the server application and the TTS engine/speechServer.

Command Line Interface

speechServer allows the server application to communicate with the TTS engine/speechServer through a command-line tool.

REST API

speechServer supports communication via a REST API when deployed behind a CGI-enabled web server (for example, Apache).

Voice and Language Switch

Switching to another language or another voice in the same language can also be performed using SSML as the input text.

Simple Implementation

ReadSpeaker speechServer is easy to install using the speechServer installer. Any licensed ReadSpeaker voice can be installed and used in speechServer.

Installation and Implementation Support

speechServer includes installation and implementation support from the ReadSpeaker Support Team. The team assists customers in understanding and carrying out the installation according to the provided instructions, as well as the implementation necessary for the client application to communicate with speechServer.

Configuration Settings

speechServer supports the use of a configuration file in which default values for key parameters can be set, including:

- Port
- Volume
- Speaking rate
- Pitch
- Sentence pause time
- Comma pause time
- Default user dictionary

Licensing

ReadSpeaker speechServer use is governed by a license agreement. See the Licensing section under Requirements and Limitations for more information. The license agreement is technically enforced by a license file

The license file specifies and enforces:

- The text-to-speech voices that are licensed
- The number of concurrent TTS ports permitted
- The permitted synthesis speed
- The number of servers covered
- The license term (duration of permitted use of speechServer)

Product Components

ReadSpeaker speechServer

A TCP/IP network-based speech synthesis solution that processes text-to-speech requests from the client application.

Client SDK

Client-side SDKs are available for C, Java, and .NET, enabling communication between the client application and speechServer.

ReadSpeaker TTS Engine

speechServer includes a TTS engine for each licenced language. speechServer invokes the TTS Engine to convert text into synthetic speech. The ReadSpeaker TTS Engine includes a pronunciation dictionary for the licensed language.

User Dictionary

Allows the user to create one or more user dictionaries for each licensed language to customize the pronunciation. The user dictionary files are used by the TTS Engine for the voice synthesis.

User Documentation

User documentation is provided and includes the APIs required to integrate the client application with speechServer, as well as the installation and implementation instructions needed to set up and configure the solution.

Requirements and Limitations

Licensing

A license agreement governs and restricts the use of ReadSpeaker speechServer.

The license agreement defines:

- The permitted usage context for speechServer, typically identified by the name and description of the speech-enabled server application
- The text-to-speech voices that are licensed
- The number of concurrent TTS ports permitted
- The permitted synthesis speed

- The number of servers covered
- The license term (duration of permitted use of speechServer)

Intended Use

Customers may use ReadSpeaker speechServer only in accordance with this product specification and the required license agreement.

Any use outside the scope defined in this specification and the license agreement is not considered intended use.

Disclaimer

The features listed in this document are guaranteed only when ReadSpeaker speechServer is implemented according to the provided instructions.

Specifications and features may be modified by ReadSpeaker at any time without prior notice.