

WebAssembly Text to Speech

Cerence's technology transforms the speech dialog experience by offering highly natural text to speech for every use case. Create speech enabled web applications with our WebAssembly SDK. Develop once and run in all major browsers including desktop and mobile versions of Safari, Chrome and Firefox.

Feature	Benefit
JavaScript API	A simple-to-use JavaScript API is provided to integrate text to speech into your applications.
Efficiency and speed	WebAssembly aims to execute at native speed by taking advantage of common hardware capabilities available on a wide range of platforms.
Safety	WebAssembly describes a memory-safe, sandboxed execution environment. When embedded in the web, WebAssembly will enforce the same-origin and permissions security policies of the browser.
Expressivity and naturalness	Improved expressive speech gives voices personality for the most natural and engaging experience possible.
Multi-lingual support	Automatic language identification, foreign language dictionaries, and high-quality acoustic extensions provide unparalleled multi-lingual readout.
Built-in domain intelligence	Optimization settings provide extra control options for special use cases such as SMS reading.
Flexible speech generation	Volume, pitch, and speaking rate can be changed at run time for more dynamic and lively effects.
Direct phonetic input	Optimal read-out of off-line phonetic databases such as navigation map data or music titles.
User text rules	Customized read-out of application-specific abbreviations and text pattern using a user text processing rule set.
User dictionaries	Application specific lexica can be phonetically optimized for accurate readout of exceptional pronunciations.
Improved prompt tuning	With off-line tuning options any prompt set can be further optimized and customized for maximum quality.
Seamless prompt insertion	Recorded audio prompts or tuned prompts are seamlessly blended with dynamic text-to-speech using automatic text matching (active prompt mechanism).
Languages and Voices	A truly universal voice portfolio offers more than 60 languages and more than 140 voices to facilitate the creation of global solutions using a single engine. The portfolio is continually expanding.
Accuracy	High linguistic accuracy offers correct readout for all types of text including a large set of personal names.
Scalability	A wide range of footprints scaling from 8 to over 900 MB ensures optimal performance from small mobile devices to powerful multi-media systems.
Speech Synthesis Markup Language (SSML)	Support of SSML allows for TTS vendor-independent markup which is correctly interpreted by Vocalizer.

Data Sheet

Cerence's complete speech output solution generates high quality speech through a seamless blending of dynamic text to speech, pre-recorded audio and optimized tuned text to speech. This new engine is optimized for reading long text in a much more natural sounding way. New signal processing algorithms improve overall smoothness of the voice output and advanced syntactical analysis gives the spoken text a natural prosody, resulting in a unique experience.

What is WebAssembly?

WebAssembly is the resulting work of members representing the four major web browsers. It allows the execution of platform-independent code inside a web browser.

By using WebAssembly text to speech will be run inside the browser without the need of a server generating the audio.

What browsers are supported?

Chrome Firefox
Edge Safari

What frameworks are supported?

In addition to running inside a web browser, you can develop off-line applications using well-known frameworks such as Ionic, Electron, Node.js, making your apps run on Android, Windows, iOS and MacOS without changes.

Use cases

- * Web sites using text to speech.
- * Voice enabled multiplatform mobile and desktop applications.

WebAssembly Text to Speech

With a broad range of options, the Cerence technology offers an excellent quality trade-off for a variety of use cases and applications.

Code

The code size for a fully featured WebAssembly text to speech engine is 5MB

Voice and Language Data

Voice Operating Point (VOP)	Flash size (excl. Code)	RAM usage
Embedded Compact – small versatile TTS suited for constrained platforms.	Average: 8 MB Max: 22 MB	Average: 5 MB Max: 18 MB
Embedded Pro – high quality TTS optimized for navigation, in-car infotainment readout; basic SMS reading capabilities.	Average: 45 MB Max: 71 MB	Average: 11 MB Max: 23 MB
Embedded High – high quality TTS read-out for SMS, news, and e-mail reading on embedded targets, suitable for all types of applications and use cases.	Average: 90 MB Max: 283 MB	Average: 17 MB Max: 45 MB
Embedded Premium – highest quality concatenative synthesis, only available for selected voices.	Average: 300 MB Max: 550 MB	Average: 120 MB Max: 250 MB

Multi-lingual voices include recorded material for one or more foreign languages. They are released for Embedded Pro and Embedded High and require 50% more memory (flash and RAM) for Embedded Pro and 100% more memory for Embedded High.

RAM usage includes code, language data, selected voice data and dynamic RAM. RAM requirement for Embedded Premium is 120 MB if platform supports memory mapping, otherwise this increases to 180 MB.

Browser and framework minimum versions required

- Chrome: 57
- Firefox: 52
- Safari: 11
- Safari Mobile: 11
- Microsoft Edge: 16
- Opera: 44
- Samsung Internet: 7.0
- Node.js: 8.0.0

Languages

Cerence TTS offers the world's largest language and voice portfolio with 64 languages and 141 voices:

- Arabic Gulf and Levantine
- Argentinean Spanish
- Australian English
- Basque
- Belgian Dutch
- Bengali
- Bhojpuri
- Brazilian Portuguese
- British English
- Bulgarian
- Canadian French
- Cantonese
- Catalan
- Chilean Spanish
- Chinese Mandarin
- Croatian
- Colombian Spanish
- Czech
- Danish
- Dongbei
- Dutch
- Farsi
- Finnish
- French
- Galician
- German
- Greek
- Hebrew
- Hindi
- Hungarian
- Indian English
- Indonesian
- Irish English
- Italian
- Japanese
- Kannada
- Korean
- Malay
- Marathi
- Mexican Spanish
- Norwegian
- Polish
- Portuguese
- Romanian
- Russian
- Scottish English
- Shaanxi
- Shanghainese
- Sichuanese
- Slovak
- Slovenian
- South African English
- Spanish
- Swedish
- Taiwanese Mandarin
- Tamil
- Telugu
- Thai
- Turkish
- Ukrainian
- US English
- Valencian
- Vietnamese