Cerence TTS (Text-to-Speech)

Cerence’s new generation of cloud and embedded text-to-speech solutions

Cerence TTS transforms the voice assistant experience by offering the most natural text-to-speech for every cloud and embedded use case.

Cerence TTS is offered through Cerence Cloud Services and embedded SDKs for Windows and Linux. Android and iOS-OSX are supported via Cerence SDK, with additional platforms supported through Cerence Professional Services.

Cerence TTS is a suite of speech output solutions to generate high-quality speech, with seamless blending of dynamic text-to-speech, pre-recorded audio, and tuned prompts. Cerence TTS is optimized to read long texts in a natural, human way. New, deep learning-based algorithms deliver higher smoothness and more natural prosody, resulting in a unique voice experience. Cerence TTS is available through distribution partners for automotive and non- automotive applications:

**Transportation**
- Route guidance
- Passenger information
- Infotainment

**Accessibility**
- Screen readers for PC and mobile phones
- Daisy book readers
- Talking kiosks & ATMs

**Consumer electronics**
- Phones
- e-Book readers
- Toys and gaming
- Electronic dictionaries

**Industrial**
- Warehouse logistics

**Public Announcements**
- Arabic Gulf & Levantine
- Argentinian Spanish
- Australian English
- Basque
- Belgian Dutch
- Belgian French
- 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
- US English
- Ukrainian

65 languages and 147 voices

Cerence TTS offers the world’s largest language and voice portfolio:

- Arabic Gulf & Levantine
- Argentinian Spanish
- Australian English
- Basque
- Belgian Dutch
- Belgian French
- 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
- US English
- Ukrainian

The most natural and dynamic text-to-speech user experience
With a broad range of options, Cerence TTS offers an excellent quality trade-off for a variety of platforms and applications.

**Embedded Code Size**
The code size for a fully featured Cerence TTS Embedded engine is 7 to 13.7 MB depending on the target platform. This can be optimized based on required language set, features and compiler choices.

**Voice and Language Data**

<table>
<thead>
<tr>
<th>Voice Operating Point (VOP)</th>
<th>Flash size (excl. Code)</th>
<th>RAM Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Pro</td>
<td>Average: 56 MB</td>
<td>Average: 16 MB</td>
</tr>
<tr>
<td></td>
<td>Max: 125 MB</td>
<td>Max: 35 MB</td>
</tr>
<tr>
<td>High-quality TTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Embedded High</strong></td>
<td>Average: 120 MB</td>
<td>Average: 25 MB</td>
</tr>
<tr>
<td>High-quality TTS readout for SMS, news, e-mail reading on embedded targets, suitable for all types of applications and use cases</td>
<td>Max: 338 MB</td>
<td>Max: 73 MB</td>
</tr>
<tr>
<td><strong>Embedded Premium</strong></td>
<td>Average: 245 MB</td>
<td>Average: 139 MB</td>
</tr>
<tr>
<td>Highest-quality deep learning based concatenative synthesis available for selected voices</td>
<td>Max: 559 MB</td>
<td>Max: 252 MB</td>
</tr>
</tbody>
</table>

**Multi-lingual** voices include recorded material for one or several foreign languages. They are released for all operating points except Embedded Compact and require up to 50% more memory (flash and RAM) compared to the numbers above.

**Embedded SDK Platforms**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Operating System</th>
<th>Code size (excl. Code)</th>
<th>RAM Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC</strong></td>
<td>Windows: 32-bit and 64-bit</td>
<td>Average: 56 MB</td>
<td>Average: 16 MB</td>
</tr>
<tr>
<td></td>
<td>Linux x86: 32-bit and 64-bit</td>
<td>Max: 125 MB</td>
<td>Max: 35 MB</td>
</tr>
<tr>
<td><strong>Devices</strong></td>
<td>Linux ARM: ARM32 Hardfp, ARM32 Softfp, ARM64</td>
<td>Average: 120 MB</td>
<td>Average: 25 MB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max: 338 MB</td>
<td>Max: 73 MB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average: 245 MB</td>
<td>Average: 139 MB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max: 559 MB</td>
<td>Max: 252 MB</td>
</tr>
</tbody>
</table>

**Feature** | **Benefits**
---|---
Emotional TTS | Developers can choose from four different speaking styles: neutral, lively, forceful, and apologetic.
Gilded speech databases | Speaking styles are enhanced by selecting expressive pre-recorded prompts (incl. non-verbal) from a “gilded speech” database accompanying the TTS voice.
Gender neutral voice | New timbre markup tag controls the perceived age, gender, or physical size of a TTS voice.
Multi-lingual support | Automatic language identification, foreign language dictionaries, and high-quality acoustic extensions provide unparalleled multi-lingual readout.
Domain intelligence | Use markup tags to improve the readout of an address, date, phone number, number, spelling, currency, and more.
Prosody control | Volume, pitch, speaking rate, and timbre can be changed at run time for more dynamic and lively effects.
Phonetic input | Optimize quality using phonetic information from external content sources like music or map data.
User text rules | Write regular expressions to expand custom abbreviations and text patterns.
User dictionaries | Create your own dictionaries for out-of-vocabulary words.
Prompt sculpting | Change unit selection results manually to increase expressivity and remove glitches (requires Cerence TTS Designer).
Seamless prompt insertion | Recorded audio prompts or tuned prompts are seamlessly blended with dynamic text-to-speech using automatic text matching (active prompt mechanism).
Cerence TTS Designer | A comprehensive Windows tool to create TTS tuning resources like user text rules, user dictionaries, and prompt databases.
Languages and Voices | A truly universal voice portfolio offers 65 languages and 147 voices for the creation of global solutions using a single engine. The language and voice 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 embedded footprints scaling from 20 to over 550 MB ensures optimal performance from low cost platforms to powerful multimedia systems.
SSML | Speech Synthesis Markup Language (SSML) allows for TTS vendor-independent markup.
Daily dictionary updates | Daily dictionary updates are offered through Cerence Cloud, guaranteeing accurate pronunciation of trending words/names in the news domain. 10 languages (en-US, en-UK, fr-FR, fr-CA, it-IT, de-DE, es-ES, es-MX, pt-PT, pt-BR) are supported.