

# 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

### 65 languages and 147 voices

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

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

The most natural and dynamic text-to-speech user experience

# Cerence TTS

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

Voice Operating Point (VOP)	Flash size (excl. Code)	RAM Usage
<b>Embedded Pro</b> High-quality TTS optimized for navigation, in-car infotainment readout; basic SMS reading capabilities	Average: 56MB Max: 125 MB	Average: 16 MB Max: 35 MB
<b>Embedded High</b> High-quality TTS read-out for SMS, news, e-mail reading on embedded targets, suitable for all types of applications and use cases	Average: 120 MB Max: 338 MB	Average: 25 MB Max: 73 MB
<b>Embedded Premium</b> Highest-quality deep learning based concatenative synthesis available for selected voices	Average: 245 MB Max: 559 MB	Average: 139 MB Max: 252 MB

**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

<b>PC</b>	Windows: 32-bit and 64-bit Linux x86: 32-bit and 64-bit
<b>Devices</b>	Linux ARM: ARM32 Hardfp, ARM32 Softfp, ARM64

## Feature

## Benefits

<b>Emotional TTS</b>	Developers can choose from four different speaking styles: neutral, lively, forceful, and apologetic
<b>Gilded speech databases</b>	Speaking styles are enhanced by selecting expressive pre-recorded prompts (incl. non-verbal) from a "gilded speech" database accompanying the TTS voice
<b>Gender neutral voice</b>	New timbre markup tag controls the perceived age, gender, or physical size of a TTS voice
<b>Multi-lingual support</b>	Automatic language identification, foreign language dictionaries, and high-quality acoustic extensions provide unparalleled multi-lingual readout
<b>Domain intelligence</b>	Use markup tags to improve the readout of an address, date, phone number, number, spelling, currency, and more
<b>Prosody control</b>	Volume, pitch, speaking rate, and timbre can be changed at run time for more dynamic and lively effects
<b>Phonetic input</b>	Optimize quality using phonetic information from external content sources like music or map data
<b>User text rules</b>	Write regular expressions to expand custom abbreviations and text patterns
<b>User dictionaries</b>	Create your own dictionaries for out-of-vocabulary words
<b>Prompt sculpting</b>	Change unit selection results manually to increase expressivity and remove glitches (requires Cerence TTS Designer)
<b>Seamless prompt insertion</b>	Recorded audio prompts or tuned prompts are seamlessly blended with dynamic text-to-speech using automatic text matching (active prompt mechanism)
<b>Cerence TTS Designer</b>	A comprehensive Windows tool to create TTS tuning resources like user text rules, user dictionaries, and prompt databases
<b>Languages and Voices</b>	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
<b>Accuracy</b>	High linguistic accuracy offers correct readout for all types of text, including a large set of personal names
<b>Scalability</b>	A wide range of embedded footprints scaling from 20 to over 550 MB ensures optimal performance from low cost platforms to powerful multimedia systems
<b>SSML</b>	Speech Synthesis Markup Language (SSML) allows for TTS vendor-independent markup
<b>Daily dictionary updates</b>	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