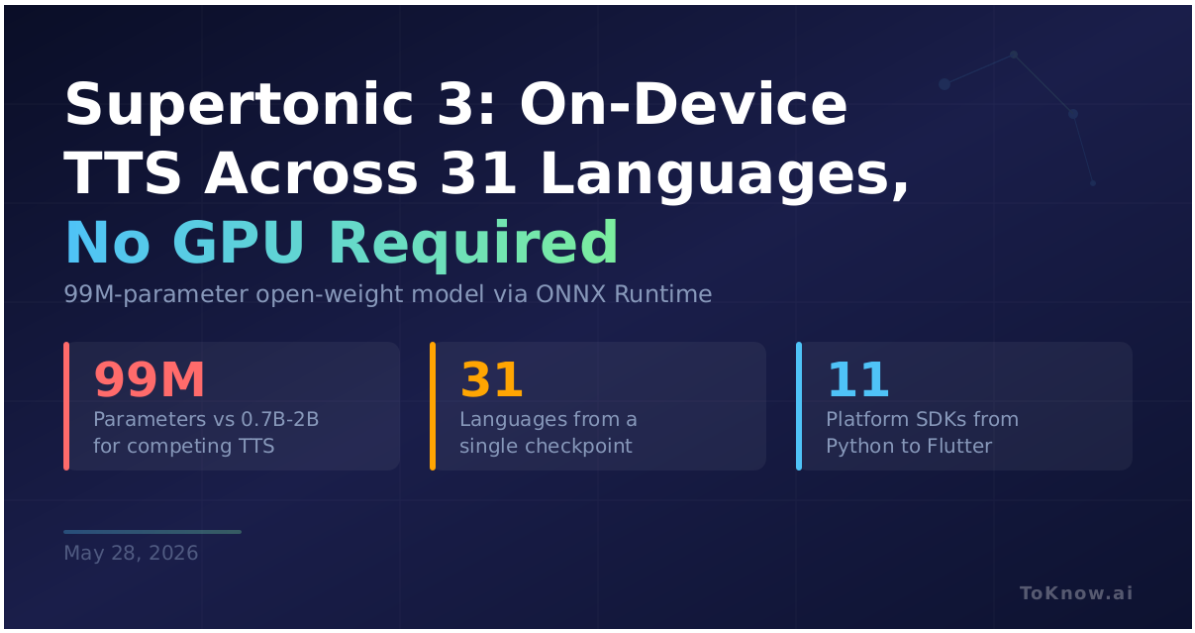
Supertonic 3: 99M-Parameter On-Device TTS Across 31 Languages, No GPU Required

Kabui, Charles

2026-05-28

Read at ToKnow.ai



Supertone released Supertonic 3, a 99-million-parameter text-to-speech model that runs entirely on-device through ONNX Runtime (a cross-platform inference engine). It supports 31 languages from a single checkpoint, outputs audio at 44.1kHz, and needs no GPU. The model runs on desktop, mobile, Raspberry Pi, and even e-readers in airplane mode. Ten inline expression tags (`<laugh>`, `<breath>`, `<sigh>`) add human nuance without reference

audio. A language-agnostic mode auto-detects input language when unspecified. SDKs cover 11 platforms including Python, Node.js, WebGPU browsers, Java, C++, Swift, Rust, and Flutter. The Python SDK includes a local HTTP server with an OpenAI-compatible `/v1/audio/speech` endpoint, so existing apps can switch by changing a URL. Weights are open under the OpenRAIL-M license.

Cloud TTS services charge per character and require an internet connection. Supertonic 3 runs offline, costs nothing after download, and fits in 99M parameters, a fraction of the 0.7B to 2B range typical of competing open TTS systems. A developer building a voice assistant can run `pip install supertonic` and generate speech locally in under a second. An e-reader can narrate books with zero network dependency. Any tool already using the OpenAI TTS API can switch to Supertonic's local server with one URL change.

On-device audio AI has closed the gap with cloud services faster than most expected. When a 99M-parameter model produces usable quality across 31 languages on a $35 Raspberry Pi, the argument for paying per-character API fees weakens for most use cases outside premium voice cloning. For more on the broader on-device AI trend, see Unsloth Studio + Qwen3.6: Running Frontier AI on 18GB RAM.

Sources:

- Supertonic 3 GitHub Repository
- Supertonic 3 Model on HuggingFace
- Supertonic 3 Audio Samples and Demo Page
- Supertonic Python SDK Documentation
- Supertonic 3 Interactive Demo on HuggingFace Spaces
