Gemini 3.8 text-to-speech says hello

23.9.2026 - Leland Rechis Alan Cowen | Google Italia

Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS are our most expressive audio generation models yet. Generate custom character voices and direct scene dialogue across Google AI Studio, Gemini API, Gemini Enterprise, Gemini Notebook, and Google Vids.

Today, we're introducing two new text-to-speech models to the Gemini family, transforming voice generation from static presets into a dynamic creative studio. These models enable creators, developers, and enterprises to create richer, more expressive audio experiences, while enabling improved user experiences in products like Gemini Notebook and Google Vids.

- Gemini 3.8 Flash TTS: Built for deep creative direction and character design. Create entirely new voices from scratch using natural language prompts to bring characters to life across gaming, immersive audiobooks, podcasts, and interactive media. Direct every performance line by line with granular control over acting cues, pacing, dialect shifts, and backchanneling.
- Gemini 3.8 Flash-Lite TTS: Built for high-volume, cost-efficient scale. Optimized for high-volume dubbing, audio content creation, and expressive voice agents with fine-grained control over tone, pacing, and expressive nuance.

These models complement our fast-growing Gemini Audio family, following 3.5 Live Translate, 3.5 Transcribe, 3.8 Live, and 3.8 Live Extended Thinking.

Create and customize your own voices

Scale up from 30 original voices to an infinite library. Whether you need an entirely original character voice or a consistent brand ambassador, our 3.8 Flash TTS model powers a full vocal studio. This enables you to create and use expressive, natural-sounding voices for every moment, while empowering developers and enterprises to easily build custom audio experiences.

- Generative voice design: With Gemini 3.8 Flash TTS, create bespoke voices from scratch by customizing role, accent and voice characteristics across more than 100 languages and dialects using natural language prompting — whether you're bringing a dramatic, fire-breathing dragon to life or crafting a charismatic narrator with a distinct regional cadence.

Hear how Gemini 3.8 Flash TTS generates a high-energy DJ voice from Melbourne.

Hear how Gemini 3.8 Flash TTS generates a super-tinny, monotone robot voice.

Hear how Gemini 3.8 Flash TTS brings a Japanese dragon to life.

- Expansive voice library: Access 2,000+ production-ready voices with broad language coverage — including regional varieties like Mexican Spanish, Quebec French, and Scots English.
- Voice replication: Recreate consistent vocal profiles from just a 30-second audio sample of your voice or a voice you have the rights to use, backed by built-in consent verification, SynthID watermarking, and C2PA credentials to protect both developers and their vocal talent.
- Save and scale: Save and manage the custom voices you designed to ensure consistent performance and minimal drift across ongoing projects.
- Voice remixing: Coming soon, pick a voice from our voice library and fine-tune timbre, pitch, pace, and accent. Use prompts to dial in characteristics (e.g. "add subtle Southern US accent"

or “soften the delivery”).

Direct the performance, line by line

Once you've selected your voices, both TTS models give you precise control over how each line is delivered.

- Direct performance line by line: Write your own stage directions or let Gemini steer delivery with natural script cues — from a calm customer service agent to a whispered suspense scene.

Hear how Gemini 3.8 Flash TTS enables natural, highly expressive conversations for interactive voice agents.

Watch and hear how Gemini 3.8 Flash TTS uses granular script control to build a deeply engaging, immersive audio experience.

- Long-form generation: Maintain high voice quality, natural pacing, and character timbre across hours of continuous audio with minimal speaker drift — ideal for podcasts and audiobooks.
- Native two-speaker scene staging: Direct multi-turn conversations seamlessly from a single script —whether for a podcast or dramatic storytelling—while keeping both voices distinctly separated with natural conversational turn-taking.
- Scripted vocal bursts & backchanneling: Add realistic conversational texture using non verbal cues (like <laughs>, <sigh>, <gasp> and active-listening interjections (like |mhm| or|yeah|) for precise comedic timing and reaction beats.

See how Gemini 3.8 Flash TTS turns natural language prompts into bespoke vocal personas from scratch.

Watch how Gemini 3.8 Flash TTS enables creators to design custom scenes to bring animated dialogue to life.

See how Gemini 3.8 Flash TTS turns scripts into fully performed dialogue scenes, letting creators direct vocal delivery, and natural turn-taking.

Get expressive high-quality speech generation built for global scale

Gemini 3.8 Flash TTS delivers leading voice customization capabilities, securing the #1 overall spot on Hume AI’s Voice Design Benchmark (71.4) and also leading in accent modeling (60.8).

Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS enable truly expressive performances without sacrificing reliability, also securing the #1 and #2 spots respectively on Hume AI’s Overall Quality Index. The model shows major improvements on a wide range of use cases such as long-form content and dual-speaker screenplay control compared to Gemini 3.1 Flash TTS.

In blind human preference evaluations on Voice Arena, Gemini 3.8 Flash and Flash-Lite TTS secure top positions amongst competitors in key global languages, including Japanese, Brazilian Portuguese, Vietnamese, Modern Standard Arabic (MSA), Mexican Spanish and Hindi. With support for over 100 languages, these models empower creators, developers, and enterprises to build high-quality, multilingual voice experiences worldwide.

Build with trust, consent, and transparency

We built our voice creation and replication capabilities with strict safeguards to help protect voice talent, respect identity, and ensure content transparency. For voice replication our system leverages consent verification: users must provide a verbal consent recording from the voice owner that matches the reference speaker before a voice can be created.

More broadly, every audio clip generated by our Gemini Audio models is watermarked with SynthID. This imperceptible watermark is woven directly into the audio output, ensuring AI-generated speech remains detectable to help prevent misinformation. For more details on our approach to safety and responsibility, review the model card.

Try our new Google AI Studio audio playground

Starting today, developers can experience these new speech generation capabilities in Google AI Studio. Built like a voice design workspace, you can prompt entirely new vocal identities from scratch or replicate your own voice [1] , then bring them directly into a dual-speaker screenplay editor to direct line-by-line delivery.

Try voice replication in Google AI Studio.

Deploy high-performance voice interfaces with ease

By using the Gemini API, developer platforms such as Agora, LiveKit, Pipecat, Vercel enable developers to build and deploy high-performance speech generation experiences with ease.

We're partnering with companies like Figma, HeyGen, Linguana, Wondercraft, 99.co, and Ollang, who are integrating our latest TTS models to help accelerate global dubbing, localize media with nuanced regional accents, and power conversational voice agents at scale.

Start using our latest Gemini Audio models:

Gemini 3.8 Flash TTS is rolling out starting today:

- For developers: In the Gemini API and Google AI Studio
- For enterprises: Coming soon via API in Gemini Enterprise
- For everyone: In Gemini Notebook.

Gemini 3.8 Flash-Lite TTS is rolling out starting today:

- For developers: In the Gemini API and Google AI Studio
- For enterprises: Coming soon via API in Gemini Enterprise
- For everyone: In Google Vids

https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-8-text-to-speech