



Gemini 3.8 Audio (Flash TTS, Flash-Lite TTS) Model evaluation

Approach, methodology & results

Approach

We evaluated Gemini 3.8 Flash TTS, and Gemini 3.8 Flash-Lite TTS using the methodologies below.

Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS

Methodology

All evaluations for Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS were conducted using production checkpoints with default sampling settings and single-attempt generation (pass@1). Evaluations benchmark acoustic quality, instruction following, latency, and multilingual breadth across controlled listener panels and blind comparative arenas.

Capabilities / Benchmarks

- **Hume AI Text-to-Speech benchmarks:** These [benchmarks](#) evaluate speech synthesis quality across multiple dimensions—including overall conversational voice naturalness, expressiveness, instruction following (style, emotion, and pacing), inline vocal burst adherence, and multilingual coverage—using controlled human listener panels and side-by-side preference ratings.
- **Voice Arena:** Evaluated via [Voice Arena](#), this benchmark measures text-to-speech performance through public crowdsourced, blind pairwise preference evaluations across US English as well as international languages and locales to determine comparative Elo rankings.

Results

Gemini 3.8 Flash TTS and Gemini 3.8 Flash-Lite TTS results as of September 2026 are below:

Hume Evals

Text-to-Speech Quality Benchmark

Hume AI

	Gemini 3.8 Flash TTS	Gemini 3.8 Flash-Lite TTS	Gemini 3.1 Flash TTS	Eleven Labs v3	Eleven Labs v3 conversational	Cartesia Sonic 3.6	OpenAI gpt-4o-mini-tts	Inworld TTS-2
Overall Reliability x Expressiveness	0.920	0.914	0.783	0.706	0.769	0.840	0.740	0.576
Human-like variation Lower = flatter or wilder than a human across turns	4.58	4.51	3.95	5.00	4.97	3.40	4.22	3.76
Multispeaker	4.14	4.10	3.60	3.85	--	--	--	--
Style Tag Control Single Tag	4.34	4.32	4.31	3.89	3.97	3.37	--	4.15

Methodology: deepmind.google/models/evals-methodology/gemini-3-8-tts

Text-to-Speech Voice Design Leaderboard

Hume AI

	Gemini 3.8 Flash TTS	ElevenLabs Voice Design v3	Inworld Voice Design
Overall English	71.4	70.8	69.8
Multilingual	3.82	3.65	3.57
Accents	60.8	45.4	35.8
Voice Qualities	74.6	76.6	76.3

Methodology: deepmind.google/models/evals-methodology/gemini-3-8-tts

Voice Arena Evals

Text-to-Speech Leaderboard
Voice Arena

	Gemini 3.8 Flash TTS	Gemini 3.8 Flash-Lite TTS	Gemini 3.1 Flash TTS	Eleven Labs v3	Cartesia Sonic 3.6	OpenAI gpt-4o-mini- tts
English	1061	1087	1051	986	1068	940
Japanese	1232	1152	1148	1048	--	975
Brazilian Portuguese	1104	1134	1094	1040	1080	946
Vietnamese	1135	1156	1099	1043	--	839
Arabic MSA	1204	1181	1135	1020	--	911
Hindi	1106	1076	1086	1052	1104	843
Mexican Spanish	1152	1146	1092	1015	1089	880

Methodology: deepmind.google/models/evals-methodology/gemini-3-8-tts