



TEXT-TO-SPEECH · NO HYPE

The Honest Buyer's Guide to AI Voice

What to actually check before you pay for a text-to-speech tool — the seven questions, the cost traps, and a 10-minute test you can run today.

Prepared by **Helge Thorsen** · Independent review · 2026 edition

Why this guide exists

Most articles about AI voice are written to sell you one tool. This one isn't. I make my living reviewing software honestly, and I'd rather you buy the *right* tool (even if that's not the one I recommend) than the loudest one.

Text-to-speech has crossed a line in the last two years: the best voices are now genuinely hard to distinguish from a human read. That's exciting — and it's exactly why so much of the marketing around it is inflated. Vendors show you their single best voice, in their best language, reading a script it was tuned on. Real projects don't work like that.

So here is the guide I wish existed: a short, practical checklist for choosing an AI voice tool that fits **your** use case, budget and legal situation — not the demo.

How to use this guide: read the seven checks once, then use the one-page scorecard on the last page to rate any two or three tools side by side. Fifteen minutes of scoring will save you weeks of buyer's remorse.

Where AI voice genuinely shines

- **Scale and speed:** narration for dozens of videos, e-learning modules or product updates, where hiring voice talent for each would be slow and costly.
- **Iteration:** when the script changes weekly, regenerating audio in seconds beats re-booking a studio.
- **Accessibility & localization:** turning written content into audio, or into many languages, quickly.
- **Drafts & prototyping:** hearing a script read aloud before you commit to a human recording.

Where it still falls short (be honest with yourself)

- **High-emotion, brand-defining reads** — a flagship ad, a moving story. A great human still wins, and audiences can tell.
- **Unusual names, jargon and acronyms** — expect to fix pronunciation by hand.
- **Long-form consistency** — tone can drift across a 40-minute audiobook.

If your project lives mostly in the first list, AI voice is likely a strong fit. If it lives in the second, budget for a human — or a hybrid.

The 7 checks before you choose

1 Naturalness — and the 10-minute ear test

Demos are cherry-picked. Instead, paste *your own* real script — including a hard sentence, a number, a date and a brand name — and listen on the device your audience will use (often a phone speaker, not studio headphones).

Ask: does it still sound natural on *my* text, or only on their demo script?

2 Languages, accents & pronunciation control

Coverage counts, but control matters more. Can you fix a mispronounced word with phonetics or a custom dictionary? Can you control emphasis, pauses and pacing (often via SSML)? A tool with 100 languages but no pronunciation control will fight you on every proper noun.

Ask: can I correct a wrong pronunciation without re-recording the whole clip?

3 Voice cloning — and the consent line

Cloning a voice is powerful and legally sensitive. Only ever clone a voice you own or have **written consent** to use. Check the vendor's safeguards (do they require proof of consent?) and your local rules. A tool that lets anyone clone anyone is a liability, not a feature.

Ask: whose voice is this, and can I prove I'm allowed to use it?

4 Latency & real-time vs batch

Pre-rendering a video? Latency barely matters. Building a live agent, IVR or app that speaks back to users? Then time-to-first-audio is everything — and you'll want a streaming API, not just file export.

Ask: is my use case batch (files) or real-time (streaming)? They need different tools.

5 The real price (not the sticker price)

Pricing is where surprises hide. Map it before you commit:

- **Per-character vs per-second vs per-seat** — a "cheap" per-character plan gets expensive at volume; a per-seat plan wastes money if only one person generates.
- **Overage rates** — what happens when you exceed the quota? This is the number that actually bites.
- **What "unlimited" excludes** — often cloning, premium voices, or commercial rights sit on a higher tier.

Ask: what will my real monthly volume cost, including overage?

6 Commercial usage & licensing rights

Can you actually **monetize** the audio — put it in a paid course, a client's ad, a monetized YouTube video? Some plans grant personal use only. Read the license, not the landing page. This single clause has sunk more projects than audio quality ever has.

Ask: does my plan explicitly allow the commercial use I have in mind?

7 Integration & workflow fit

The best voice is worthless if it doesn't fit how you work. Check for the export formats you need, a clean API if you'll automate, and integrations with your editor, LMS or CMS. A tool that forces manual copy-paste for every clip will quietly cost you hours.

Ask: does this drop into my existing workflow, or bolt awkwardly onto the side?

Cost traps to avoid

- Annual plans bought before you've tested real volume for a month.
- "Unlimited" tiers that throttle speed or quality once you're heavy.
- Premium voices priced separately from the base quota.
- Commercial rights locked behind the top tier — check before, not after.
- Per-seat pricing when only one person actually generates audio.

REALITY CHECK

When *not* to use AI voice

Saying this out loud is the whole point of an honest guide:

- **Your flagship, emotion-led piece.** If one video carries your brand, hire a human.
- **Regulated or high-trust contexts** where a synthetic voice could mislead — disclose, or don't use it.
- **Tiny volume.** If you need one 30-second read, a freelancer may be cheaper and better than a subscription.
- **When you can't secure the rights** to the voice or the commercial use.

Everywhere else — scale, iteration, localization, accessibility, drafts — modern AI voice is genuinely good, and getting better fast.

The one-page scorecard

Rate each tool 1–5 on every check. Add the column. The highest total that also clears your must-haves (rights, budget, real-time if you need it) wins.

CHECK	TOOL A	TOOL B	TOOL C
1 · Naturalness on <i>your</i> script	—	—	—
2 · Languages & pronunciation control	—	—	—
3 · Voice cloning & consent safeguards	—	—	—
4 · Latency (real-time vs batch)	—	—	—
5 · Real price at your volume	—	—	—
6 · Commercial usage rights	—	—	—
7 · Integration & workflow fit	—	—	—
Total (out of 35)	—	—	—

Where Fish Audio fits

In the spirit of this guide: no tool wins every row for every buyer. Here's my honest read on **Fish Audio**, and who it's for.

What it does well

- Strong, natural-sounding voices that hold up on real scripts, not just demos (check 1).
- Multilingual coverage — useful if you localize (check 2).
- A developer-friendly API and streaming, so it suits both batch narration and real-time, speak-back use cases (check 4, check 7).
- Voice-cloning capability — powerful when you have the rights and consent (check 3).

Who should look elsewhere

- If your entire need is a single, emotion-led brand read — hire a human.
- If you want the absolute cheapest option for tiny, occasional use — a pay-per-use freelancer may beat any subscription.

The honest move is to run the 10-minute ear test (check 1) with your own script, confirm the commercial rights you need (check 6), and score it against one alternative using the sheet on the previous page. If it wins on the rows that matter to *you*, it's a strong pick.

Try Fish Audio with your own script

Run the ear test yourself — paste your real text, listen on a phone speaker, and see if it clears your must-haves.

→ fish.audio

Disclosure: the Fish Audio link above is an affiliate link. If you sign up through it, I may earn a commission — at no extra cost to you. It doesn't change the advice in this guide: the seven checks and the scorecard work for any tool, and I'd rather you pick the one that fits your project.

Thanks for reading. Score two tools this week — the fifteen minutes it takes is the cheapest insurance you'll buy all year.