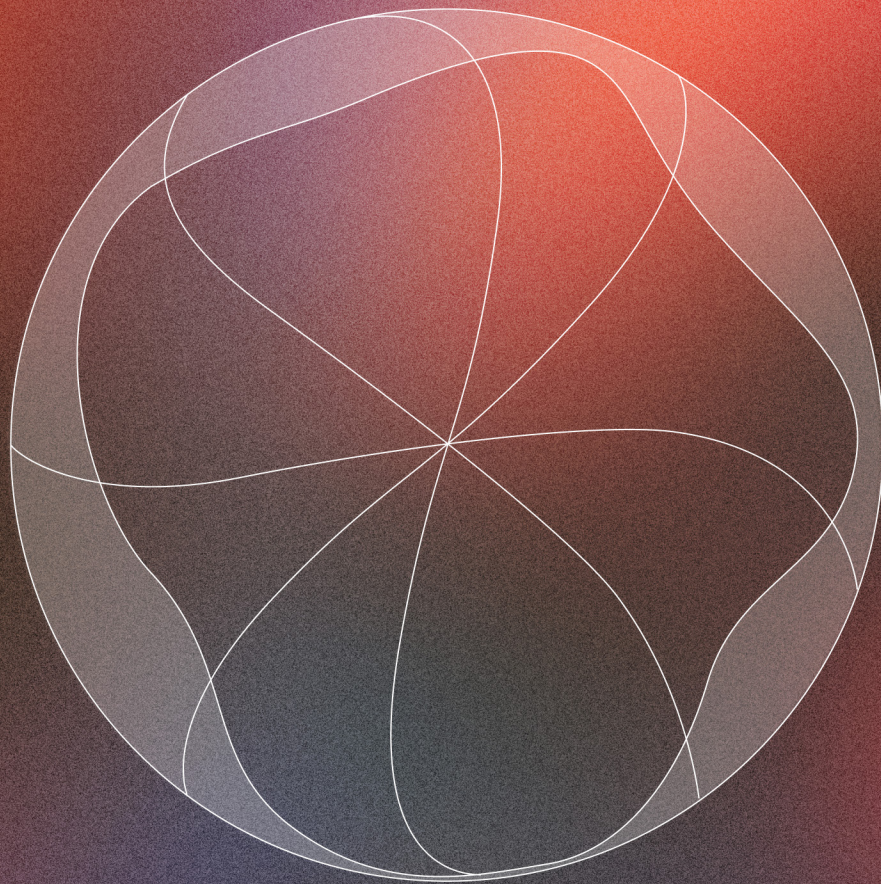


IIElevenLabs

The Voice Blueprint

Designing the Ideal AI Voice for Your Brand



Contents

- 03 Introduction
- 04 Defining your brand's AI voice
- 06 Selecting your voice
- 14 How leading companies found the right AI voice for their use case
- 19 Powering your voice
- 24 Delivering the next generation of voice experiences

Introduction

In every major industry - from telecom to finance, retail, and gaming - voice has become the defining interface of automation. Choosing the right voice to represent your brand is now central to enterprise success in deploying voice AI.

Voice automation delivers two key outcomes: it scales operations and creates human-like experiences that keep customers engaged through natural conversation. The first is straightforward - it transforms manual workflows into hands-free processes. The second - sustaining engagement - is more nuanced, and depends heavily on how a brand sounds, and whether that voice feels authentic to its audience.

Speak with a voice design expert today

Author



Mark Seidl
Voice operations
at ElevenLabs

Mark specializes in helping enterprises design and deploy proprietary voice systems that align with their brand identity and performance requirements. Mark oversees the end-to-end process of curating, sourcing, and producing voices across industries, working closely with professional voice actors, studios, and engineering teams to ensure technical precision and expressive realism in every deployment.

Defining your brand's AI voice

How a brand delivers its communications meaningfully shapes how customers perceive it. So, what constitutes the “perfect” AI voice for your brand? At ElevenLabs, where we support over 1,500 enterprises with AI voice solutions, we’ve found that the ideal AI voice depends heavily on the company’s use case, brand identity and audience. The best conversational agent voice may deliberately have a down-to-earth, hyperrealistic tone that feels authentic in telecom or customer support. In marketing, a voiceover may need a voice to sell a product that is emotive and engaging with a rich tone that draws in the listener.

For global enterprises, considering localization is equally vital. Extending a brand’s voice to every language their customers speak helps maintain stronger connections across regions. Whether it’s a fintech agent reassuring customers in Spanish or a retailer scaling multilingual support, the ability to preserve emotion and identity across accents ensures consistency and trust.

In response, we've focused at ElevenLabs on building technology that helps brands design and select voices with the same precision they apply to their visual identity. Since 2022, our team of world-leading audio AI engineers has developed the world's most realistic voice models - systems that capture the subtle nuances of the human voice, not just tone or timbre, but the imperfect qualities that make speech sound genuinely human. The hesitations, rhythm, and cadence - the authenticity of a rare accent or the emphasis in a phrase - are what differentiate a human brand voice from a robotic one.

With these innovations, enterprises in telecom, retail, financial services, and healthcare are already replacing legacy IVR systems with adaptive, human-sounding AI agents that reflect their brand personality.



Selecting your voice

With your brand's preferred voice identity defined, the next step is exploring the full range of ways to create it. At ElevenLabs, there are multiple ways to obtain the perfect voice for each of your use cases.

We'll dive deeper into each:

Professional content

Choose from 10,000+ Professional Voice Clones in 70 languages

Voice cloning

Clone a custom voice

Voice Design

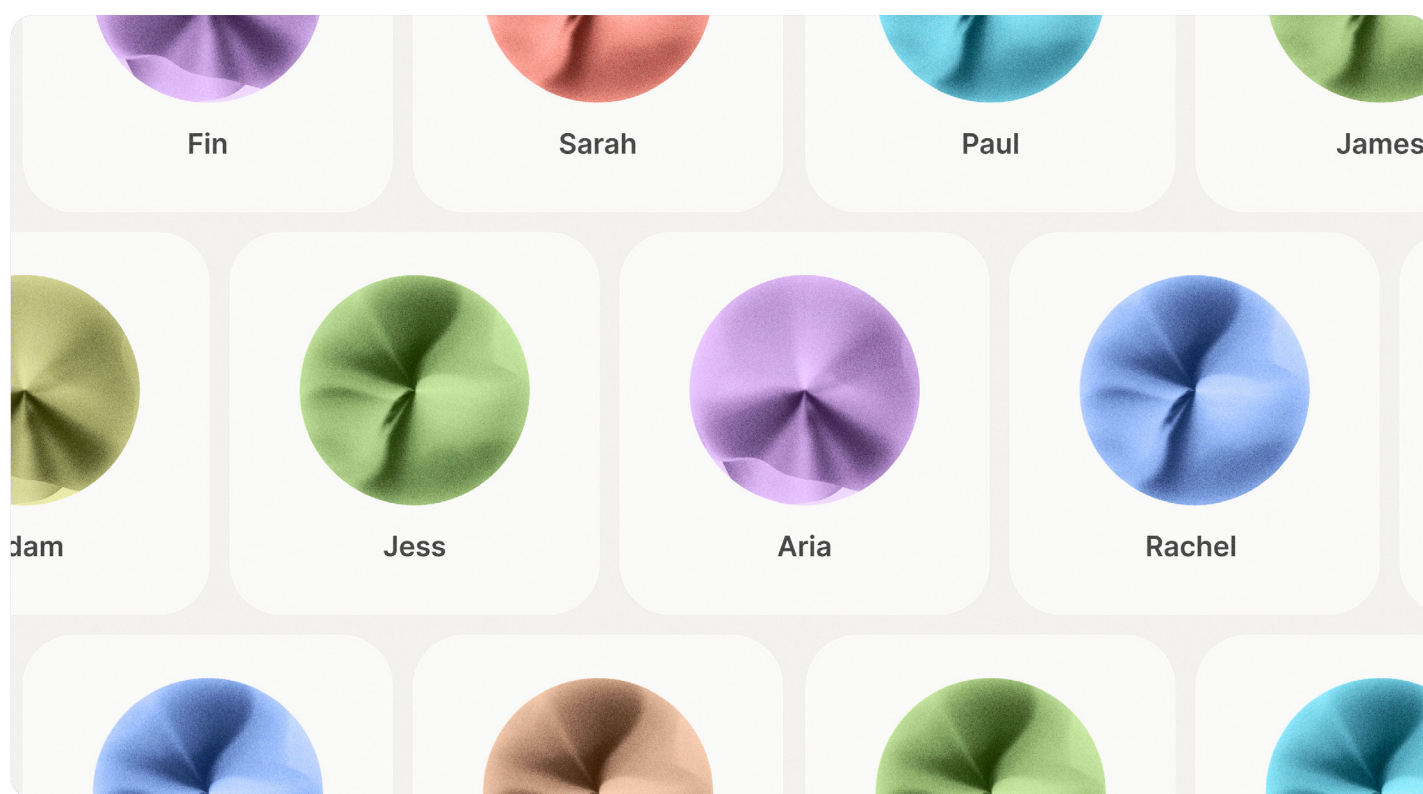
Prompt the exact voice needed

Voice Remixing

Alter existing voices you own

Voice Library

ElevenLabs' Voice Library offers nearly 10,000 Professional Voice Clones, with new voices added daily - shared by both professional voice actors and enthusiasts alike across more than 70 languages. Our marketplace model, combined with leadership in voice cloning and audio AI, gives us orders of magnitude greater voice diversity than any other provider.



Within the library, over 1,000 voices carry the "High-Quality" designation. Each has been reviewed by audio experts to meet professional standards for recording quality, mixing, and technical accuracy. Voices in this category are verified to be free from distortion, reverb, and other artifacts.

This curated selection provides organizations with reliable voice options for a wide range of business applications. Additionally, the library includes comprehensive filtering capabilities to streamline voice selection:

Language and accent

Filter by specific languages and regional accents to match target audiences.

Use case categories

Narrow results by application type, including advertisements, conversational, narration, and educational content.

Demographics

Sort by gender and age range to align with brand requirements.

In addition, you can sort voices by trending, date added and most used.

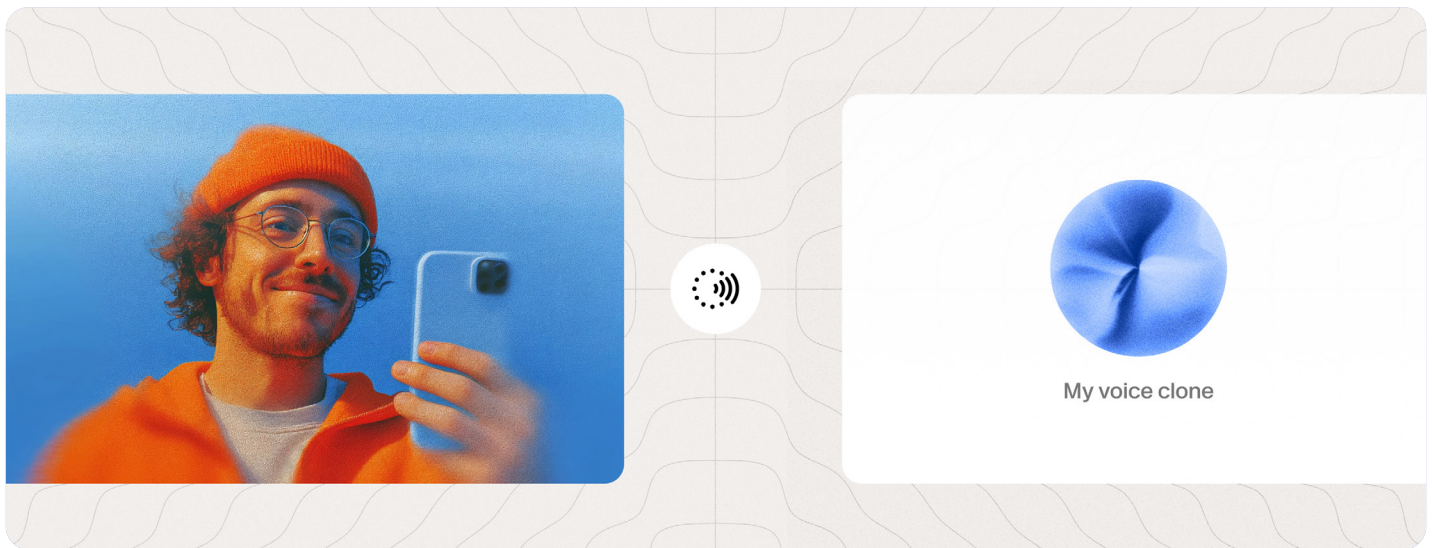
With this range of languages, accents, and styles, enterprises can identify voices that meet brand requirements across all use cases and regions - accelerating deployment timelines. For retailers, this means instantly launching localized voice experiences. For financial institutions, it ensures the same trusted tone carries consistently across markets.

A brand's voice often needs to be unique - something never heard before. Sorting by "newest" helps discover emerging voices that may not yet appear in trending results but can provide fresh alternatives for upcoming projects.

The Voice Library continues to expand rapidly, with new voices added every day.

Voice Cloning

Voice Cloning enables organizations to create custom voice models using recordings from professional voice actors, employees, or any authorized speaker. This allows brands to maintain a consistent voice across applications, scaling voice deployment efficiently, and reduce ongoing voice talent costs while preserving quality and authenticity.



Two cloning methods are available to accommodate different requirements:

Instant Voice Cloning (IVC):

Requires 10 seconds to 2 minutes of recorded audio. Ready immediately, it's ideal for scaling multiple agents or voices simultaneously with minimal recording overhead.

Professional Voice Cloning (PVC):

Requires 30 minutes or more of recorded audio to produce a high-fidelity clone that preserves tone, prosody, and mannerisms indistinguishable from the original

speaker. Trains a custom model in 1-2 hours. Best suited for organizations working that need an exact replica of the original voice.

For enterprises deploying voice cloning in customer support or conversational agent use cases, ElevenLabs' cloning tools capture tone and prosody indistinguishable from the source helping brands deliver natural, lifelike voices that enhance realism and consistency across every interaction.

At ElevenLabs, we've had the privilege of working alongside some of the most recognized voices in history. Voice Cloning enables creators, estates, and enterprises to preserve and extend a speaker's legacy while scaling their presence across new formats, languages, and audiences. We work to capture the emotion, personality, and nuance that make every voice distinctly human. With iconic voices such as Sir Laurence Olivier, it's about honoring the depth of their craft and the layers of character that define them. We're not just reproducing a sound, but reflecting the gravitas, artistry, and presence that made it extraordinary. Voice Cloning, in this sense, is as much about preserving legacy as it is about capturing likeness.

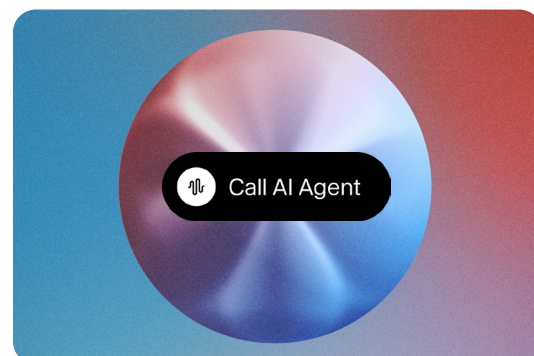
01 Acting Audition with AI Sir Laurence Olivier

Audition for the role of a lifetime with Sir Laurence Olivier, the legendary thespian who brings every word to life with gravitas and drama.



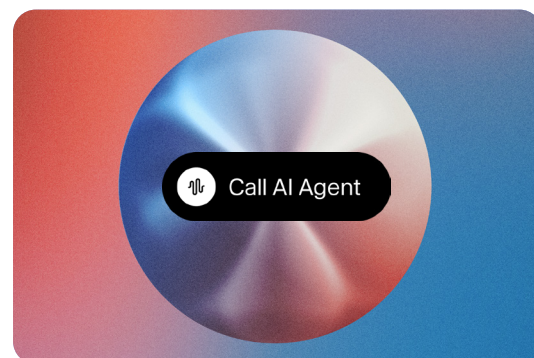
02 Guided Meditation with AI Deepak Chopra

Find your center with Deepak Chopra, the world-renowned spiritual teacher who guides you toward calm, clarity, and inner peace through mindful conversation.



03 Physics Puzzle Challenge with AI Richard Feynman

Step into a lively demo game of ideas with Richard Feynman, the brilliant and curious physicist who turns complex science into simple, playful puzzles you can solve together.



Voice Design

ElevenLabs Voice Design is a text-to-voice generation tool that enables organizations to create custom voices through descriptive prompts. This feature addresses scenarios where specific voice characteristics - particularly niche accents or specialized performance styles - are unavailable in the Voice Library or impractical to clone. Voice Design eliminates the need for audio recordings or voice talent sourcing when precise specifications can be articulated through text.


The system utilizes our v3 model architecture with the following capabilities:

- **Prompt-based generation:** Create voices by describing attributes such as age, gender,

accent, tone, pacing, emotion, and character type. Detailed prompts produce more specific results, while concise descriptions generate broadly applicable voices.

- **Niche accent and performance optimization:** Particularly effective for generating uncommon regional accents, specialized acting styles, or character voices that may be difficult to source or clone through traditional methods.
- **Audio tag integration:** Supports performance tags including [laughing], [shouting], [whispering], [sighs], and [exhales] to incorporate natural speech behaviors and emotional expressions into generated voices.

Conversational



Voice Design

✕


Prompt

A sarcastic and relaxed woman from California. She is natural and conversational and has a husky but attractive tone. She speaks at a normal quick pace. The audio quality is good.

Generate voice

⊙ Settings

Advertisement



Voice Design

✕


Prompt

A professional man in his 50s with an American accent. He has a resonant and engaging tone. He bright and natural voice is perfect for advertisements, as he speaks with a clear voice at a quick pace. The audio quality is perfect.

Generate voice

⊙ Settings

Narrative & Story



Voice Design

✕


Prompt

A warm, gentle grandfather with a thick British accent. His speech has a slight gravelly texture from age, but maintains a soothing, mellow quality. He speaks at a conversational, quick and engaging pace. The audio quality is good.

Generate voice

⊙ Settings

Characters & Animation



Voice Design

✕

Prompt

A high-pitched, raspy gnome with snarky comedic timing and true emotional depth. Switches between grumpy humor and warm wisdom. This is a twisted creature voice balancing feral madness with childlike vulnerability. Authentic British accent with unique gnomish personality. Clean, expressive vocal performance.

Generate voice

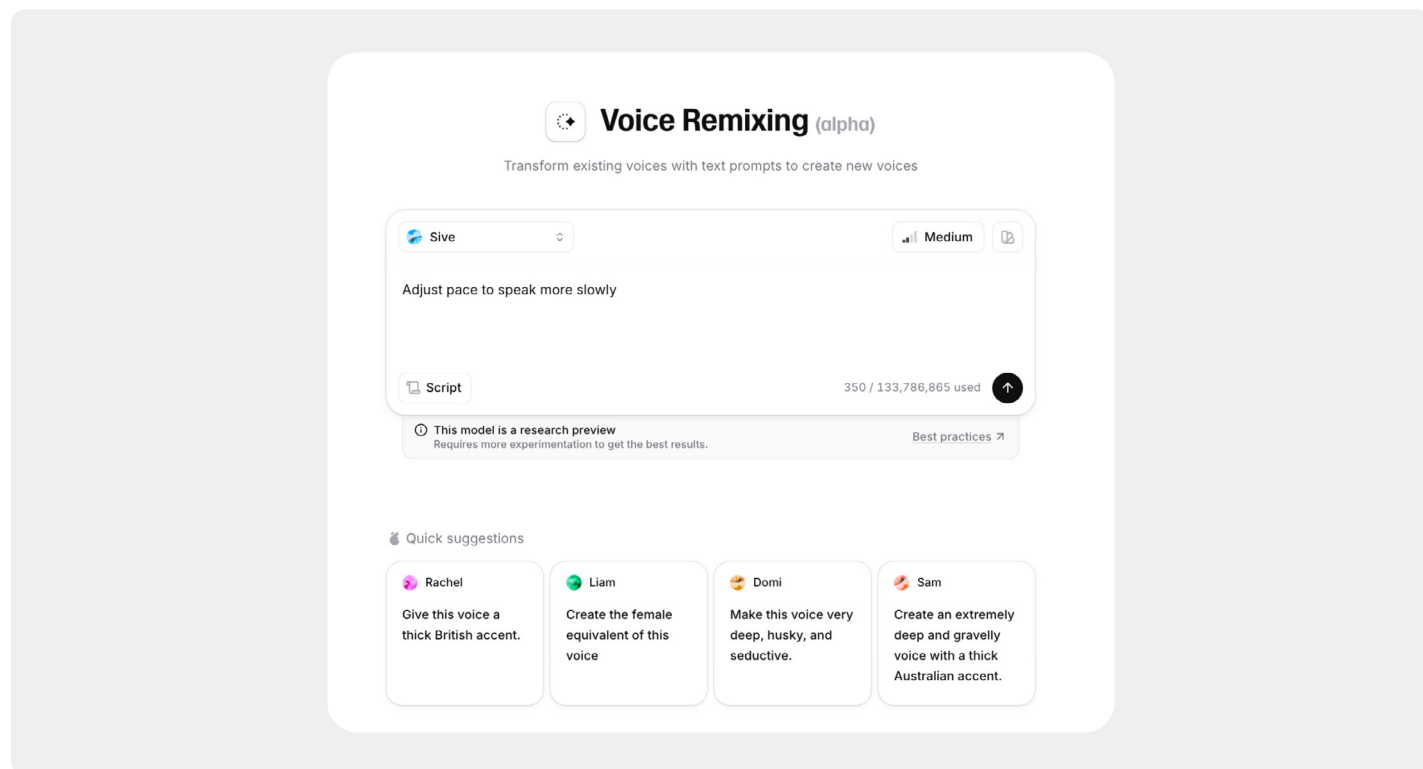
⊙ Settings

Voice Remixing

Voice Remixing gives the ability to modify your existing voices by adjusting specific attributes while preserving core voice characteristics. You can adjust pacing, accent, tone, or audio quality independently and apply refinements over time without re-recording voice clones or going back to square one on voice design.

This allows enterprises to maintain tone while slowing speaking speed for elderly demographics in a financial services call center, or adapt accents for different regional telecom audiences.

Voice Remixing offers adjustable prompt strength levels to control transformation intensity, from subtle refinements to substantial changes. The feature is compatible with all owned voice types including Instant Voice Clones, Professional Voice Clones, and Voice Design outputs. Original voices remain unchanged, with each remix creating a new variant in the library.



How leading companies found the right AI voice for their use case

Voice isn't just functional - it's a brand's personality made audible. Every interaction - from customer support to storytelling - communicates identity, quality, and intent.

Here are a few examples how selecting the right voice can elevate customer engagement:

Chess.com - bringing AI tutoring to life



With more than 225 million members worldwide, Chess.com hosts one of the largest gaming communities.

Last year, a user of one of Chess.com's top learning apps, Learn Chess with Dr. Wolf, left a review: "Is there a way to get Dr. Wolf to talk through the lessons? My 7-year-old is struggling to read his commentary." Dr. Wolf, an interactive virtual chess teacher, adapts to each player's ability and explains moves step by step.

Finding the right voice for Dr. Wolf was essential to connect with players of all ages and skill levels. It needed to sound authoritative yet warm, like a coach who teaches with both confidence and empathy. After sampling several voices from the ElevenLabs Voice Library, the team found one that perfectly matched the character.

[Listen to Dr. Wolf's voice](#)

This month, Chess.com expanded its voice-powered learning tools by launching Play Coach, a feature that delivers feedback on players' moves in the voices of iconic chess grandmasters such as Hikaru Nakamura, Levy Rozman, and Magnus Carlsen.

"Learning works best when it feels personal. With Play Coach, Chess.com shows how familiar voices can make expert guidance easier to follow and more enjoyable to learn from." - Mati Staniszewski, Co-founder at ElevenLabs

Learn more about our work helping Chess.com design the ideal voice for Dr Wolf [here](#) and launch their iconic voice coaching offering, Play Coach, [here](#).

NVIDIA - scaling multilingual keynote delivery through voice AI



NVIDIA, a global AI infrastructure company valued at over \$4 trillion, partnered with ElevenLabs to redefine how multilingual presentations are delivered.

Traditionally, live keynote broadcasts required human translators or on-screen subtitles to reach global audiences. Using ElevenLabs' proprietary voice cloning technology, the NVIDIA team cloned Jensen's voice to deliver a real-time, multilingual keynote presentation, preserving his unique voice traits and cadence, at Computex in 2024.

The presentation reached millions worldwide and showed how voice AI makes global communication more authentic, scalable, and on-brand.

You can read more about NVIDIA's voice experience built on ElevenLabs [here](#).

Hear the voice we helped NVIDIA create below:

Jensen's voice clone in mandarin

Delivering multilingual voice agents for a leading enterprise SaaS

A leading enterprise SaaS company serving more than 150,000 business customers worldwide partnered with ElevenLabs to expand its global voice capabilities.

The company needed synthetic voice options in multiple languages and accents that could be deployed quickly across its cloud infrastructure and customer platforms. Using a combination of Voice Design and Voice Mixing, ElevenLabs created high-quality, customizable voices in nine different languages - each tailored for clarity, tone, and accent precision.

These voices were optimized for rapid deployment through integration with Amazon SageMaker, allowing the client to scale voice experiences across products and regions in a fraction of the usual development time.

The project demonstrated how enterprise organizations can use ElevenLabs' voice technology to accelerate localization, maintain brand coherence across languages, and integrate AI voice directly into existing ML workflows.

Each of these projects reflects the same principle: every brand's voice is unique, and ElevenLabs tools provide the optionality to design, refine, and scale that voice with precision. Whether drawing from our library, cloning an existing voice, or designing one from scratch, brands gain full control over how they sound - across every language, product, and interaction.

Powering your voice

Once you've selected the ideal voice, the model determines how it performs - its realism, latency, and expressiveness. Model selection is as critical as voice selection, since every interaction - whether a marketing voiceover or a real-time support call - relies on the model powering it.

ElevenLabs offers several Text to Speech models designed to meet different performance and quality requirements.

Multilingual v2

Our Multilingual v2 model delivers exceptional quality across three core areas: audio fidelity, naturalness, and language fluency. It allows a single voice to perform seamlessly across more than 30 languages, accurately reproducing the desired accent in each.

This robust and stable model is best suited for offline applications where ultra-low latency isn't the primary concern.

Best suited for:

Voice agents

Ideal for real-time customer support and conversational applications where quality is the top priority.

Character voiceovers

Ideal for gaming and animation due to its emotional range.

Audiobooks and long-form content

Delivers consistent, stable performance across extended narration.

Professional content

Well-suited for marketing ads, corporate videos, and e-learning materials.

Multilingual projects

Maintains consistent voice quality and accent accuracy across languages.

Turbo and Flash models

Our Turbo and Flash models deliver unmatched low latency for real-time use cases, with a slight trade-off in audio quality compared to Multilingual v2.

Best suited for:

Voice agents

Ideal for real-time customer support and conversational applications where ultra-low latency is the top priority.

Interactive experiences

Perfect for games and live applications requiring immediate response.

Large-scale processing

Efficient for bulk Text to Speech conversion at reduced cost.

V3 (alpha)

V3 is our latest and most advanced speech synthesis model. It is a state-of-the-art model that produces natural, life-like speech with high emotional range and contextual understanding across multiple languages.

Best suited for:

Character dialogue

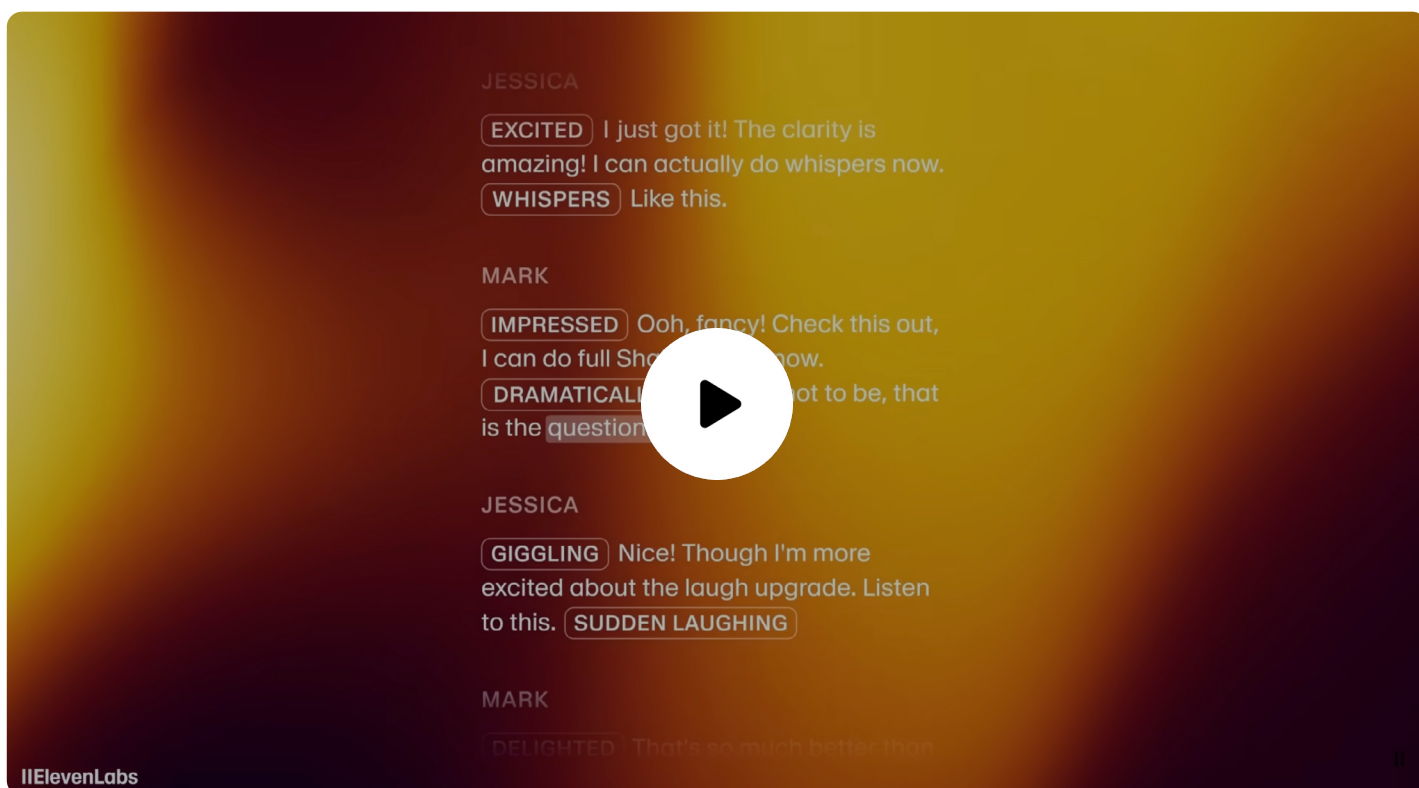
Ideal for audio experiences with multiple interacting characters.

Audiobook production

Perfect for long-form narration with nuanced emotional delivery.

V3 brings an unprecedented level of emotion and expressiveness to voice AI. With a wide selection of audio tags, it allows precise control over delivery, speed, accent, and pitch for specific parts of speech output.

For example, a voice can add the audio tags such as [whispering], [laughing] or [in excitement] to a phrase to direct emotion or tone precisely. This flexibility across expressive and dynamic use cases makes V3 a breakthrough model for truly lifelike voice performance.



Model selection guide for use case

V3 is our latest and most advanced speech synthesis model. It is a state-of-the-art model that produces natural, life-like speech with high emotional range and contextual understanding across multiple languages.

Model selection guide for use case

Creative projects

Use **V3 (alpha)**.

Best for high-fidelity audio with the most emotionally rich expression.

Conversational agents

Use **Multilingual v2** for the highest quality, **Flash v2.5** for the lowest latency, and **Turbo v2.5** for a balance of quality and latency.

Delivering the next generation of voice experiences

Voice has become more than an interface - it's the foundation of how people experience brands. As automation expands across every industry, the ability to design, deploy, and scale voices that reflect a brand's identity is now a competitive advantage.

At ElevenLabs, we're building the technology that makes this possible. Whether it's crafting a new voice through text, cloning an existing one, or powering real-time conversational agents across global markets, our tools give enterprises full control over how they sound - consistently, authentically, and at scale.

The future of brand experience isn't just powered by voice. It's defined by it.

[Speak with a voice design expert today](#)