

AI-POWERED VOICE-OVER

workflows june 2026



For AI-powered voice-over projects, InText offers three main approaches:

- 1. Using an existing AI voice**
- 2. Cloning a real speaker's voice**
- 3. Creating a custom AI voice**

The best option depends on the target language(s), pronunciation requirements, voice consistency needs, and whether a specific speaker's voice must be preserved.

1. Using an existing AI voice

In this approach, we select a suitable voice from a library of AI voices cloned from real speakers. Voices can be filtered by language, accent, gender, age, and use case.

This is usually the fastest, most cost-effective, and flexible option, as it does not require any recordings from the client.

1.1 Universal / multilingual voice

A universal voice can be used across several languages.

Best suited for: Multilingual e-learning, corporate training, product explainers, and projects where a consistent narrator persona should be maintained across languages

Main advantage: Voice consistency across language versions

Possible limitation: Pronunciation may be less natural in some languages compared with a language-specific voice

1.2 Language-specific voice

A language-specific voice is optimized for a particular language.

Best suited for: Marketing videos, customer-facing materials, public training content, and projects where natural pronunciation is the main priority

Main advantage: Better pronunciation, accent, and natural delivery in the selected language

Possible limitation: For multilingual projects, different languages may require different voices

Note: Library voices may become unavailable if removed by their owners. For long-term projects, we confirm voice availability in advance.

2. Cloning a real speaker's voice

In this approach, an AI voice is created based on a recording of a real speaker. This is used when the client wants the voice-over to sound like a specific person.

Please note that you must have the necessary rights and permission to use the speaker's voice.

2.1 Instant voice clone

An instant voice clone is created from a short audio sample. Around 10 seconds of audio is sufficient for a basic clone.

Best suited for: Quick tests, demos, internal materials, pilot projects, and cases where similarity is sufficient

Main advantage: Fast setup and minimal source audio required

Possible limitation: The result may be similar to the original voice, but not an exact replica. Stability and pronunciation may be less consistent than with a professional clone

2.2 Professional voice clone

A professional voice clone is created from a longer recording. It usually requires at least 30 minutes of suitable clean audio. The speaker will be asked to complete a short voice verification to confirm authorization.

Best suited for: Official corporate narration, branded voice-over, long-term e-learning programs, product tutorials, and recurring content updates where a specific human voice must be preserved consistently over time

Main advantage: Higher similarity to the original speaker and better stability for longer or repeated use

Possible limitation: Requires more preparation, including an extended speaker recording session

3. Creating a custom AI voice

In this approach, a completely new AI voice is generated from a detailed voice description, without using any existing voice library or real speaker recording as a base. The voice characteristics (tone/timbre, age, gender, pacing, accent) are defined collaboratively with the client.

Best suited for: Clients who need a unique branded voice with a specific personality — for example, a distinctive corporate narrator or a character voice for e-learning

Main advantage: A distinctive voice profile that can be used consistently across future content, with no dependency on third-party voice libraries or real speakers

Possible limitation: Output quality may vary depending on the prompt and use case. Requires a detailed voice brief and an iterative approval process; typically requires more time than selecting an existing voice

HOW WE WORK

1. PROJECT BRIEFING

We align on your content type, target language(s), voice requirements, and approval process to define the most suitable AI-powered voice-over workflow for your project.

2. DATA HANDLING AND CONFIDENTIALITY

For sensitive projects, we align with the client on approved tools and applicable confidentiality terms before production begins. Materials are processed only through tools and services that meet our data-protection standards, with EU-based processing and no use of client data for AI training where required.

3. SCRIPT REVIEW AND PRONUNCIATION GUIDE

We review the script; create a pronunciation guide covering brand names, product terms, abbreviations, and language-specific terms; and share it with you for approval.

4. VOICE SELECTION AND APPROVAL

We prepare voice samples for client approval and confirm the final voice before production.

5. PRODUCTION

We generate the voice-over, applying the pronunciation guide throughout. For video projects, we also handle audio-to-video alignment, with lip synchronization available on request.

6. QUALITY ASSURANCE

Every project goes through a structured QA review based on a defined scoring framework, including pronunciation, terminology, clarity, naturalness, tone, pacing, audio quality, and synchronization. Native language reviewer checks the target-language aspects, while our production operator verifies timing, technical quality, and final file readiness.

7. DELIVERY AND REVIEW

We deliver files for client review and incorporate feedback.

CONTACT



Valeriya Vezovyk

Head of Project Management

Tel: +380 676 34 29 77

Email: crm@intext.com

MS Teams: localize@intext.com

