



DATASHEET · CREATE · VOICE TTS

Recording every announcement by hand?

From now on, your text does the talking.

Voice TTS turns text into natural-sounding speech: more than 800 voices in 37 languages, dialogues with several speakers and your own voice from a voice sample. For announcements, instructions and customer communication, in the app and via API.

AT A GLANCE

More than 800 voices with audio samples, three quality tiers, dialogues with speaker tags, voice cloning from Premium and a pronunciation lexicon for names and brands.

01 · WHAT IT IS ABOUT

Re-record after every change? Edit the text, regenerate the audio.

The phone greeting gets new opening hours, the manual gets a new step, and once again someone has to step up to the microphone. With Voice TTS you only edit the text and regenerate the audio, in the same voice and without booking a recording session.

INPUT

Text up to 50,000 characters, for dialogues with one speaker tag per section

VOICES

More than 800 voices in 37 languages, female and male, or your cloned voice

OUTPUT

Audio file as MP3, from Premium also as WAV or OGG

ACCESS

App under

Media AI › Voice Studio

, REST API and MCP


Large voice catalog

More than 800 voices in 37 languages, with two variants each for Spanish and Portuguese. Basic offers 5 voices per language, Premium and up all of them.

Dialogues with several speakers

In **Multi Speaker** mode, each section starts with a language and voice, e.g. [de|Anneliese]. The studio merges all sections into one audio file.

Your own voice

Under **Voice Cloning**, upload a voice sample of at least 10 seconds, and your text is spoken in that voice. Available from the Premium tier.

Pronunciation lexicon

In the **TTS lexicon**, you define how names, brands and abbreviations are pronounced. Entries are saved to your account, from the Premium tier.

Three quality tiers

Basic for the fastest generation, Premium with all voices, Ultra for the best quality. You choose the tier under **Model**.

Speaking speed

Fine-tune the pace, e.g. 0.9 for slightly slower or 1.1 for slightly faster. The default is 1.0.

02 · HOW IT WORKS

Enter text, pick a voice. Listen and download.

In Voice Studio it takes four steps. The REST API and MCP offer the same voices, tiers and formats.

1

Open Voice Studio

Under **Media AI > Voice Studio**, choose the **Text to speech (TTS)** tab and type or paste your text.

2

Pick language, voice and tier

Select language and voice, then set Basic, Premium or Ultra under **Model**. For dialogues, switch to **Multi Speaker**.

3

Generate the audio

The studio creates the file as a job and shows it for playback as soon as it is ready.

4

Download and reuse

Download saves the file. In the **History**, you can restore earlier settings and regenerate the audio.

TECHNICAL DATA

FEATURE	DETAILS
Voices and languages	more than 800 voices, 37 languages, list with audio samples via GET /voice/voices
Text length	up to 50,000 characters per job
Output formats	MP3, WAV, OGG; in the app WAV and OGG from the Premium tier
Voice sample for cloning	WAV, MP3, M4A, FLAC, OGG, WebM, AAC; at least 10 seconds, up to 10 MB
Quality tiers	Basic, Premium (default), Ultra
REST API	POST /job/add/paperoffice_voice___tts, dialogues with voice=Multi
MCP	po_tts_generate, po_tts_multivoice_generate, po_tts_voice_clone, po_tts_voices_list

03 · IN PRACTICE

Hotline, training, product video. Each with the right voice.

A new greeting for the hotline

Support writes the greeting, picks a voice and downloads the MP3. When opening hours change, they just edit the text and regenerate the greeting.

A training dialogue with two voices

HR writes a dialogue with the tags [de|Anneliese] and [de|Florian] and gets one audio file in which both speak in turn.

A product video in several languages

Marketing translates the script in Translate Studio and voices each version with a voice in that language. The lexicon makes sure the product name is pronounced correctly.

FREQUENTLY ASKED QUESTIONS

How many voices and languages are there?

More than 800 voices in 37 languages, with two variants each for Spanish and Portuguese. GET /voice/voices returns the list with name, gender and audio sample.

How do dialogues with several speakers work?

Each section starts with a tag for language and voice, e.g. [en|Sarah]. Via API you send the text with voice=Multi, and every section then needs a tag.

Can I use my own voice?

Yes, from the Premium tier. Upload a voice sample of at least 10 seconds; 15 to 30 are recommended. The studio detects the spoken text of the sample, and you can correct it.

How is it billed?

By usage, depending on the quality tier. Current prices are on the pricing page.

GET STARTED

Create your first announcement in Voice Studio.

Sign in to your PaperOffice account, open **Media AI › Voice Studio** and choose **Text to speech (TTS)**. Enter text, pick a voice, generate the audio.

paperoffice.ai/en/ai-voice-tts-text-to-speech-voice-synthesis-tts



MORE ABOUT THE MODULE

paperoffice.ai/en/ai-voice-tts-text-to-speech-voice-synthesis-tts

PRICING

paperoffice.ai/en/pricing

CONTACT

paperoffice.ai/en/contact