

Getting started with the transcription

Studios receiving an original version with no script

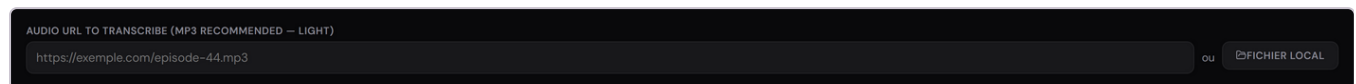
Transcribe an episode with no script

A programme arrives with no text, or with a script that no longer matches the cut.

5 minutes of setup, then the wait

01 Point at the source

Transcription works on the **audio**, not the picture. Paste the media address, or pick it under **Médias**. An audio-only file is enough, and faster than a whole video.

A dark-themed interface element showing a text input field with the placeholder "AUDIO URL TO TRANSCRIBE (MP3 RECOMMENDED — LIGHT)". The field contains the URL "https://example.com/episode-44.mp3". To the right of the field is a button labeled "ou" followed by "FICHIER LOCAL".

The address field for the audio to transcribe, and the local file button.

02 Pick the language, or don't

Auto-detect is the recommended setting, and genuinely so: on audio mixing two languages (Wolof and French in the same scene) it often beats forcing one.

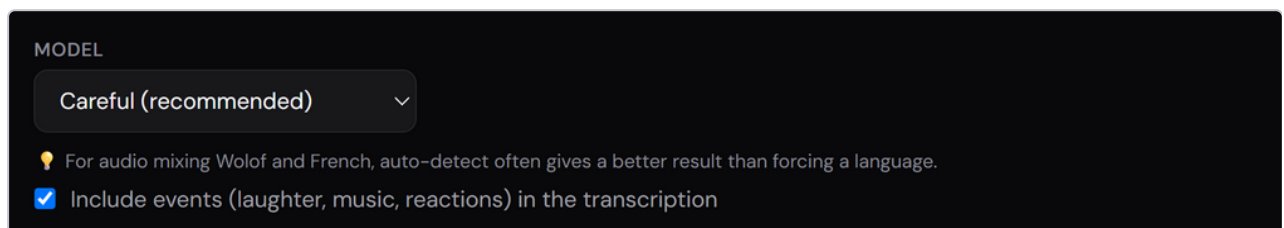
Forcing a language only helps when detection gets it wrong, which happens on very short programmes.

A dark-themed interface element with the title "LANGUAGE". Below the title is a dropdown menu showing "Auto-detect (recommended)" with a small blue icon to the left and a downward arrow to the right.

The Language menu, on automatic detection.

03 Set the level and the events

Careful is slower and more accurate; **Standard** is enough for a first pass. The **Include events** box adds laughter, music and reactions, useful in dubbing, where they take real space on the band.

A dark-themed interface element with the title "MODEL". Below the title is a dropdown menu showing "Careful (recommended)" with a downward arrow. Below the dropdown is a lightbulb icon followed by the text "For audio mixing Wolof and French, auto-detect often gives a better result than forcing a language." Below that is a checkbox with a blue checkmark and the text "Include events (laughter, music, reactions) in the transcription".

The model choice, Careful or Standard, and the events checkbox.

04 Work with no network

An **offline engine** installs from **Moteurs locaux**, in one click. The content no longer leaves the machine and nothing is billed by the minute: that is what an embargoed programme, a confidentiality clause, or a booth with no connection calls for.

In exchange, **no speaker separation**: every line arrives on a single character, to be split afterwards. And it is the machine's processor doing the work — expect longer than the programme's own duration.

Local engines

Two features run with no network once installed: **offline transcription** and **voice / M&E separation**. They are not in the installer — these are hundreds of megabytes that not everyone needs.

Offline transcription ✓ ready

The content never leaves the machine and nothing is billed by the minute. In exchange, **no speaker separation** — every line arrives on a single character.

Large fast model (recommended) · 547 MB

The best quality/time ratio. Holds up on African languages. This is the one to take if you hesitate.

installed

Remove

Large full model · 1030 MB

The best recognition, markedly slower. For a difficult episode, not for a season.

Install

Small model · 181 MB

⚠ Weak on African languages and on difficult audio. For a modest machine only, and knowingly.

Install

Voice / M&E separation not installed

Separates the original dialogue from the music and effects. The M&E obtained is a real stem, deliverable.

About 1 GB to download, several minutes. Python 3.14 detected.

Installed from its authors' official repository; its weights are downloaded by Demucs itself on the first separation.

Install

Close

The Moteurs locaux window: offline transcription and voice / M&E separation, to install on the machine.

05 Start it, and let it run

The screen names the current phase and the elapsed time, and estimates what remains from your previous runs. You can interrupt at any point.

Transcription

Transcribe from a URL, a hosted file, or a **fichier local** (audio of any kind — mp3, wav, m4a, aac, ogg, flac, opus... — or video). Watch the video (MP4) to make corrections.

Where does the text come from?

I have no text
Have it written from the audio or video

I have a script
SRT, VTT, DetX, TTAL, TXT, Excel

I already have a transcription
Paste a result in JSON format

AUDIO URL TO TRANSCRIBE (MP3 RECOMMENDED — LIGHT)

ou **FICHIER LOCAL**

VIDEO URL FOR THE PLAYER (MP4 — OPTIONAL)

Charger **Local video** **Waves**

⚡ Slow connection? Download the video once, then open it as a **Local video**: instant playback and scrubbing, with no network.

⚡ Put the MP3 in the Audio field (fast transcription) and the MP4 in the Video field (to see who is speaking). The two are synchronised by timestamp. If you only have a video, put it in the Audio field.

MODEL

Careful (recommended)

LANGUAGE

Auto-detect (recommended)

⚡ For audio mixing Wolof and French, auto-detect often gives a better result than forcing a language.

☒ Include events (laughter, music, reactions) in the transcription

Transcribe **Diagnostic**

• Diagnostics and raw response

The transcription screen: source, language and fidelity before starting the job.

06 Collect the text

The lines arrive **timecoded** in the script, ready to adapt. That is the starting point for everything else: detection, the band, the voices.

TC	CHARACTER	TEXTE	ACTIONS
▶ 0:09.40 0:05.40 5.00s	SAMARA	The night before my mating ceremony, my twin sister took my partner's bed.	↑ ↓ ↺ ⌂ 🗑
+ Insérer ici			
▶ 0:06.10 0:08.20 2.10s	FINN	You knew. You always knew.	↑ ↓ ↺ ⌂ 🗑
+ Insérer ici			
▶ 0:08.90 0:12.40 3.50s	SAMARA	Now she can marry that monster in my place.	↑ ↓ ↺ ⌂ 🗑
+ Insérer ici			
▶ 0:13.20 0:16.00 2.80s	ALARIC	Nobody asked you anything.	↑ ↓ ↺ ⌂ 🗑
+ Insérer ici			
▶ 0:17.00 0:20.60 3.60s	SIENNA	He looked broken, as if he had not slept.	↑ ↓ ↺ ⌂ 🗑
+ Insérer ici			
▶ 0:21.40 0:24.00 2.60s	SAMARA	Maybe. I do not know any more.	↑ ↓ ↺ ⌂ 🗑
+ Insérer ici			

The transcribed lines, timecoded and assigned to their character, in the script editor.

This tutorial online : <https://axeho.studio/en/tutorials-transcription.html>