



LAIA en Media Party Buenos Aires 2024

WORKSHOP ORGANIZADO POR LAIA:

"Herramientas IA de audio para producción de podcasts"

SPEAKERS: SERGIO PAOLETTI / DAVID CORONEL / GASTON LEMA

Viernes 30 de agosto

13.00 hs

Ciudad Cultural Konex,
Sala D

Sarmiento 3131
Abasto, CABA





QUIÉNES SOMOS.

Laboratorio Argentino de Inteligencia Artificial

@laialab_





Un grupo interdisciplinario de profesionales conscientes del cambio histórico que puede traer aparejado el avance de las herramientas de inteligencia artificial para toda la humanidad.

SOMOS

MISIÓN

Explorar de forma interdisciplinaria los avances tecnológicos en inteligencia artificial y sus impactos en la sociedad, participar activamente en el debate público de manera crítica y eficaz, compartir conocimiento y promover la adopción de tecnologías abiertas.

VISIÓN

Impactar positivamente en la sociedad a través de la divulgación y generación de herramientas de IA abiertas y transparentes que promuevan la soberanía tecnológica y la democratización de la IA.



QUÉ HACEMOS

Laboratorio Argentino de Inteligencia Artificial

@laialab_



LAB

CINE-DEBATE

REDES

IMPACTO

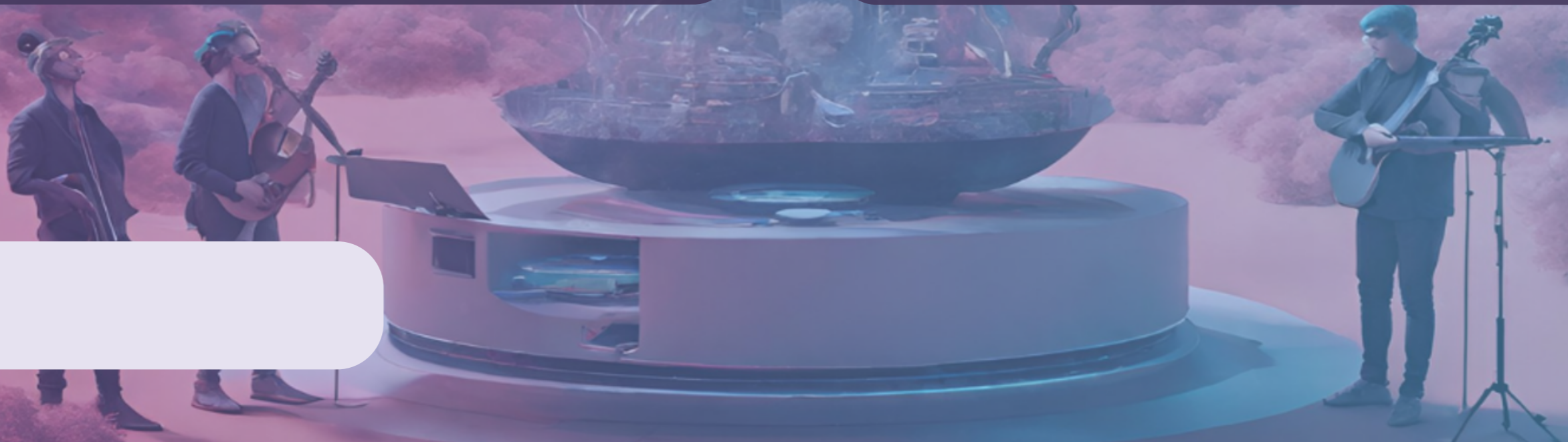


PLAYGROUND

MINI LAIA

BLOG

PODCAST





LAIA | entre IA | la señal y el ruido

PODCAST





MÚSICA & AUDIO

Herramientas IA de audio para producción de podcasts

DAW (Digital Audio Workstation)

Plugins para restauración

Plugins para ecualización

Plugins para dinámica / otros

Transcripción con Whisper

Texto a Audio

Bonus track: Música

MAPA DE RUTA



Nuestro bien #1:

La Voz

IMPORTANCIA DE LA CALIDAD DEL AUDIO

primera impresión / expectativa / credibilidad

adaptación a distintos entornos de escucha

accesibilidad

experiencia del oyente



Pro Tools

Every great session starts with an idea

Buy now

Try free

Pro Tools makes music creation fast and fluid, providing a complete set of tools to create, record, edit, and mix audio. Get inspired and start making music with a massive collection of plugins, instruments, and sounds. Tackle your most ambitious visions with proven features that power you through the most demanding projects.

Avid ProTools

CHOOSE A VERSION

show pricing for

Pro Tools Intro	Pro Tools Artist - Annual Subscription	Pro Tools Studio - Annual Subscription	Pro Tools Ultimate - Annual Subscription
Get started making music	Everything you need to start making music	For serious music artists, producers and engineers	The complete toolset for music and audio post
Free	\$99 USD/YEAR	\$299 USD/YEAR	\$599 USD/YEAR
Download Now	Buy Now	Buy Now	Buy Now
<p>Key features include</p> <ul style="list-style-type: none"> ✓ 8 audio and 8 MIDI tracks ✓ 39 instruments and effects ✓ 2GB of loops and samples 	<p>Key features include</p> <ul style="list-style-type: none"> ✓ 32 audio and 64 MIDI tracks ✓ 100+ instruments and effects ✓ Inner Circle and Sonic Drops 	<p>Key features include</p> <ul style="list-style-type: none"> ✓ 512 audio and 1 video track ✓ 120+ instruments and effects ✓ Immersive mixing support 	<p>Key features include</p> <ul style="list-style-type: none"> ✓ 2,048 audio/64 video tracks ✓ Full audio post workflows ✓ Immersive mixing support

Students and teachers get special pricing with [verified academic eligibility](#).

**Produce music. Produce podcasts.
Take total control of your sound.**

[More about our new release](#)

Audacity

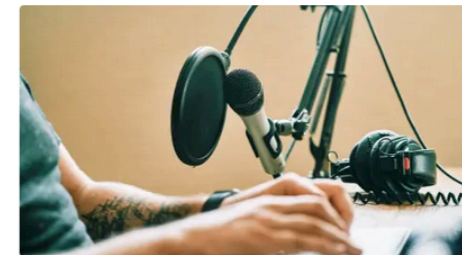
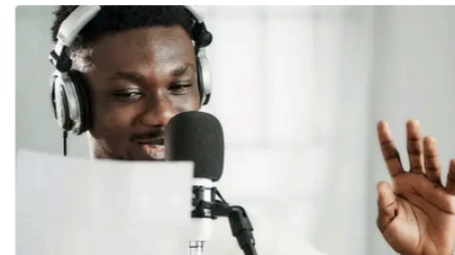
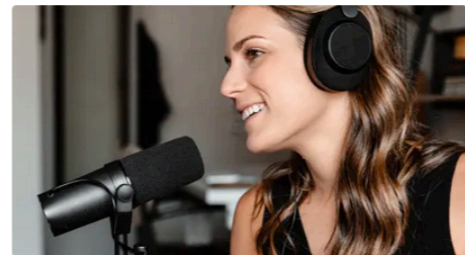
3.6 



NEW EFFECTS!

Record anything

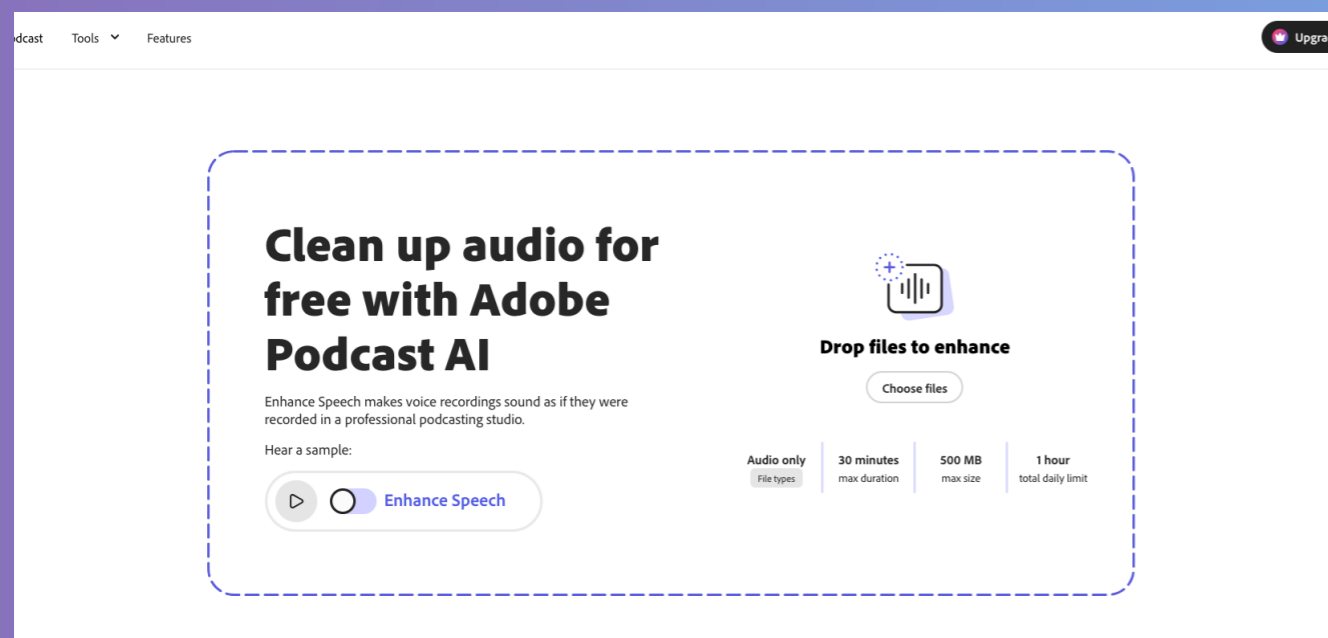
Excellent for podcasts, voice-over work or even just recording memos.



Audacity

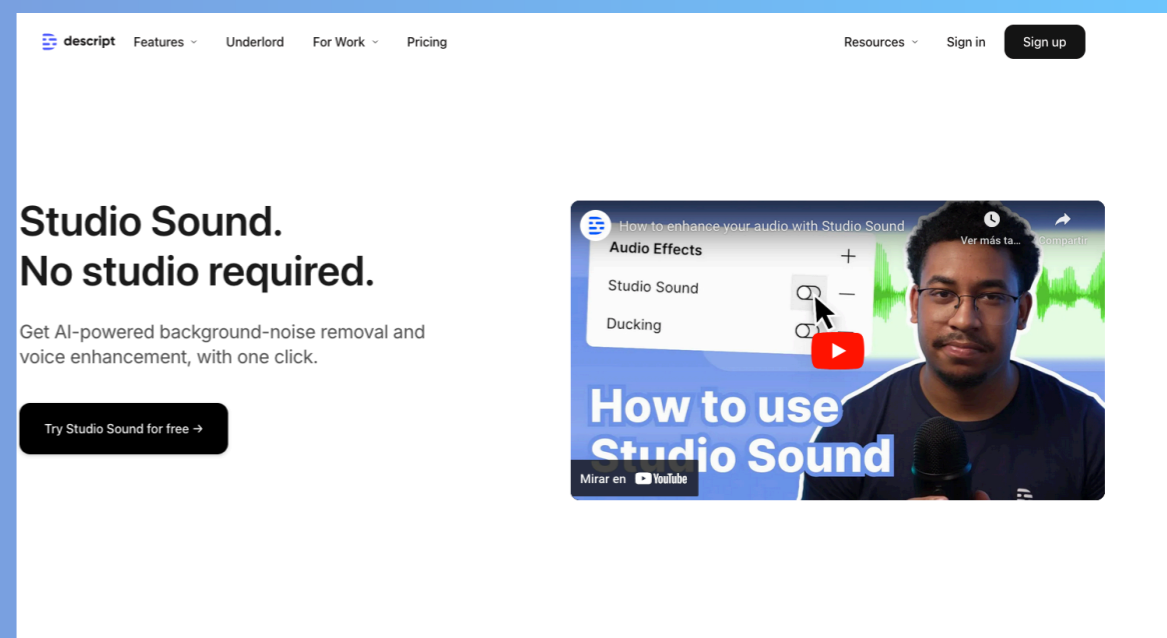
BASADOS EN LA WEB

Limpieza / Restauración



Adobe Speech Enhance

<https://podcast.adobe.com/enhance>

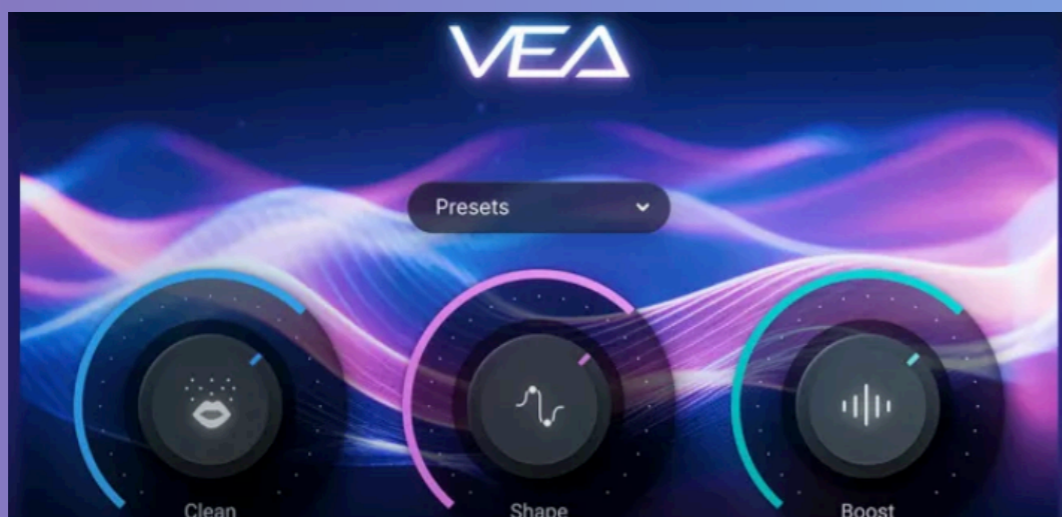


Descript Studio Sound

<https://www.descript.com/studio-sound>

Limpieza de ruido ambiente / Restauración

PLUGINS PARA DAWs



VEA, de iZotope

<https://www.izotope.com/en/products/vea.html>



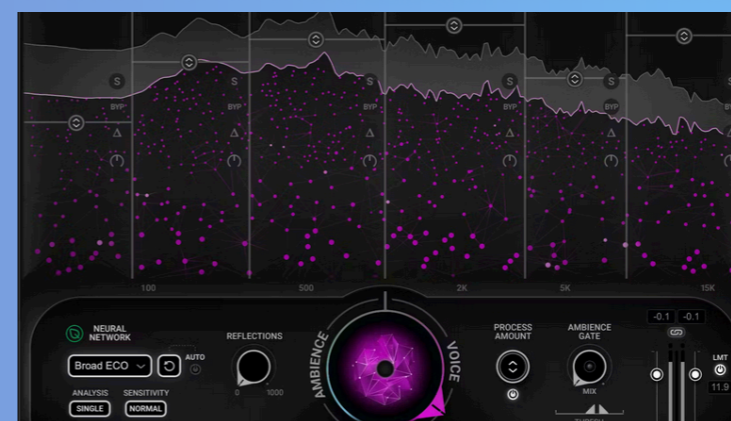
Clear, de Superpone

<https://product.supertone.ai/clear>



RX de iZotope

<https://www.izotope.com/en/products/rx.html>



Clarity VX Pro, de Waves

<https://www.waves.com/plugins/clarity-vx-pro>

Limpieza y acondicionamiento de voces

PLUGINS PARA DAWs



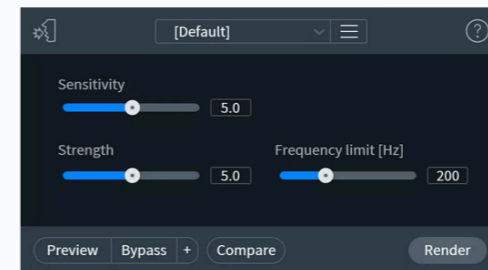
Included in Advanced
Dialogue Contour

Adjust the character and expressivity of a voice with new levels of control in Dialogue Contour, ensuring that dialogue edits flow seamlessly. Smooth out fluctuations in pitch or tone in voiceover recordings, and seamlessly piece together Frankenbites to create a cohesive performance.

De-plosive*

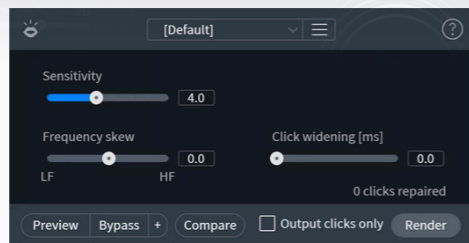
Quickly remove plosives with De-plosive. Improve the quality of your dialogue by getting rid of plosive pops and mic bumps that can occur during a recording.

Intelligent analysis detects the presence of a plosive between the desirable harmonics of a dialogue or vocal signal. It then transparently reduces the plosive far more effectively than simply using an EQ to roll off the low end.



Mouth De-click*

Mouth De-click is a de-clicker that is finely tuned to detect and reduce mouth noises such as clicks and lip smacks. It's designed for use on longer audio selections, but it can also be used to remove individual clicks.



Dialogue Isolate

We've overhauled Dialogue Isolate, adding in Dialogue De-reverb and state-of-the-art machine learning that enables lightning-fast (real time) processing. Now you can de-noise dialogue or vocals and control reverb in one easy-to-use module, also available as part of RX 11 Standard for the first time. The Advanced version includes a brand new, highest-quality offline processing mode as well as multi-band processing.



Repair Assistant

Clean up tracks faster than ever with enhanced precision. Polish dialogue and sung vocals with ease thanks to new machine learning tech, and make final tweaks with deeper controls. The next generation Repair Assistant goes from trusty helper to lifesaver.



En RX, de iZotope

<https://www.izotope.com/en/products/rx.html>

Clarity VX Pro, de Waves

<https://www.waves.com/plugins/clarity-vx-pro>

IA PARA PLUGINS EN DAWs

Ecuallizadores

Presets vs Profiles

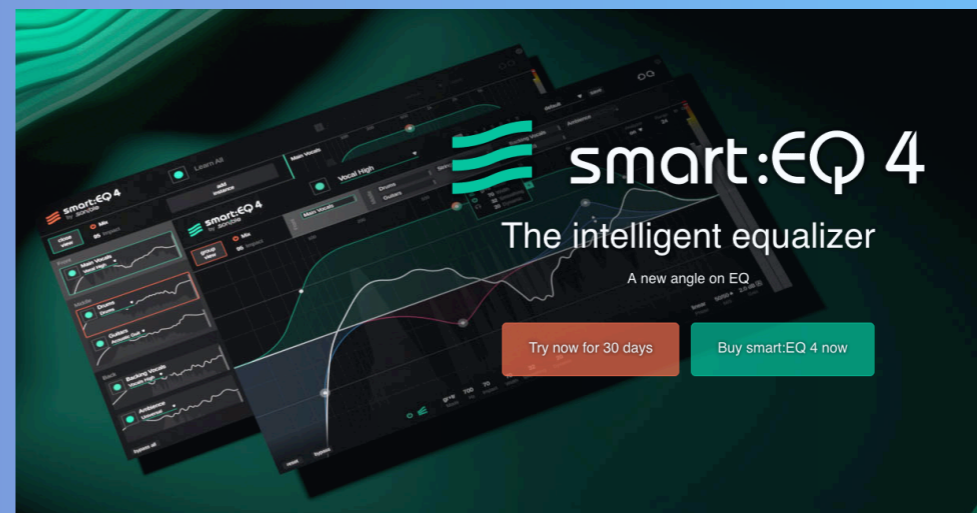


PLUGINS PARA DAWs

Ecuualizadores



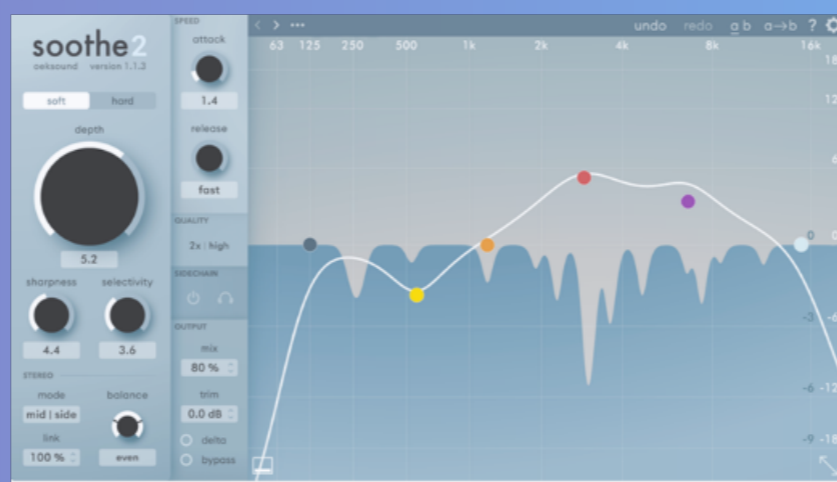
FAST Equalizer, de Focusrite



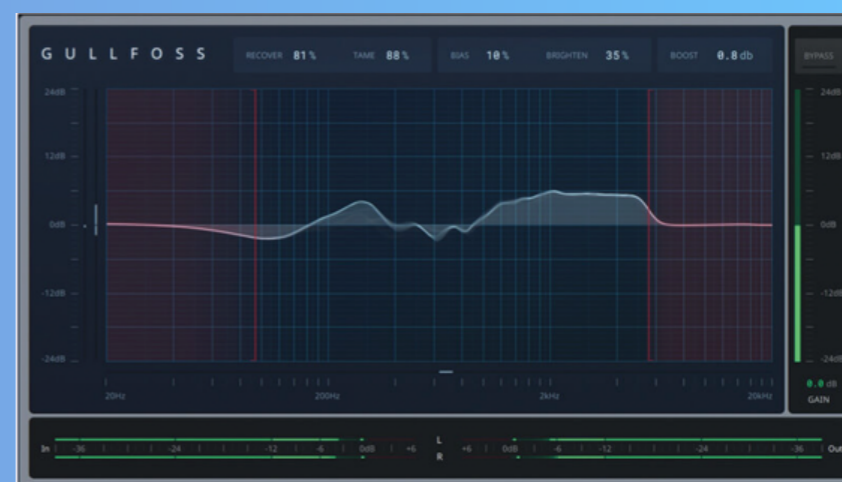
Smart EQ 4, de Sonible
<https://www.sonible.com/>



Neutron de iZotope



Soothe 2 de Oeksound



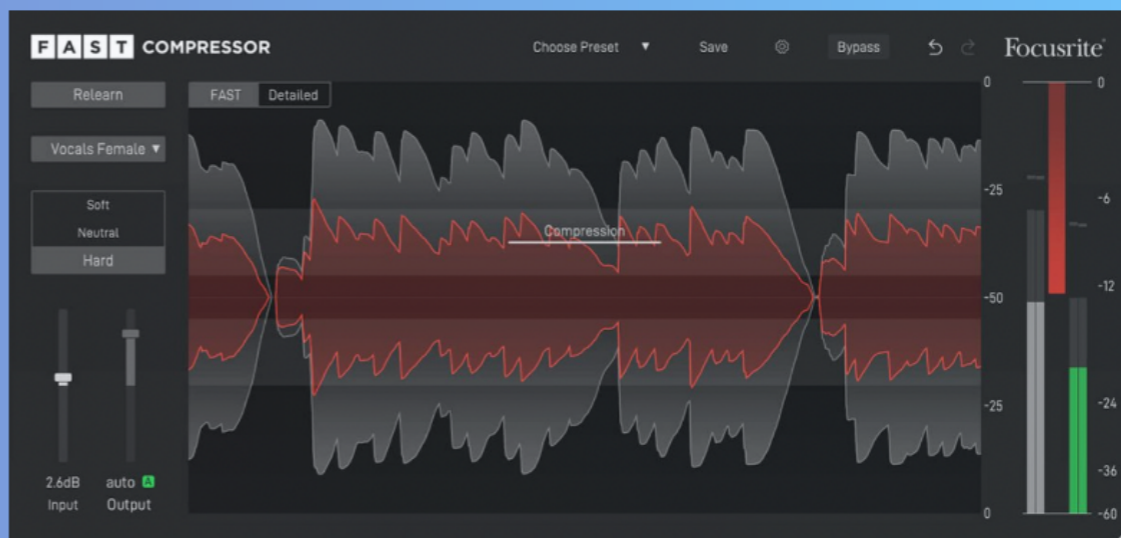
Gullfoss de Soundtheory

IA PARA PLUGINS EN DAWs

Compresores



LVL-01B de Tone Empire



FAST Compressor, de Focusrite



Smart Comp2 de Sonible



THE_COMPRESSOR de Phil Speiser

PLUGINS PARA DAWs

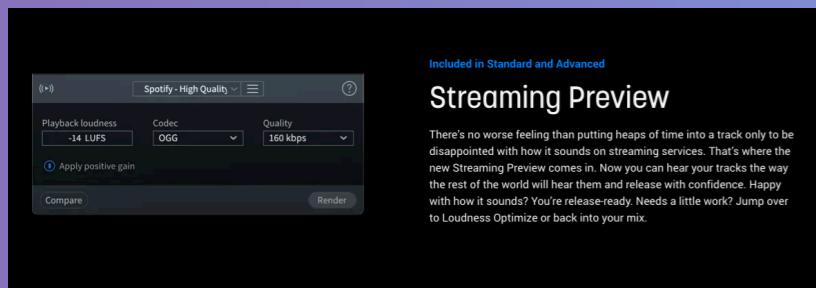
... más!



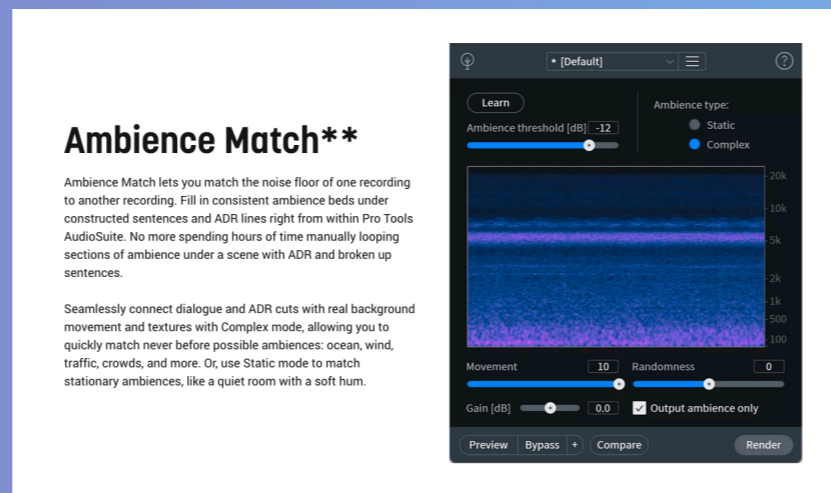
ADPTR AUDIO Streamliner

Optimize your music for streaming services

Streamliner de Adaptr Audio



Streaming preview de iZotope

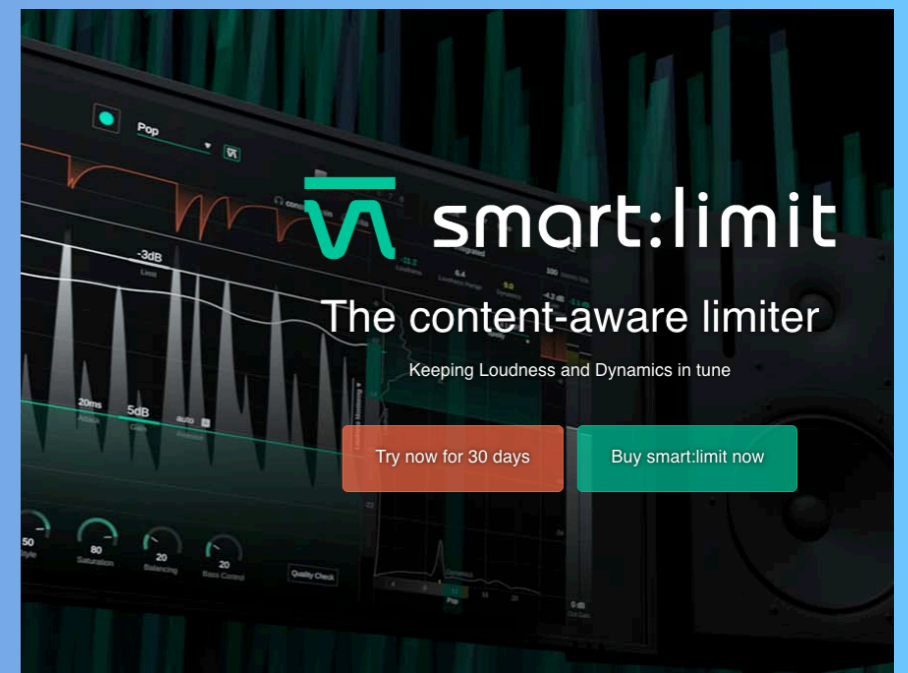


Ambience Match**

Ambience Match lets you match the noise floor of one recording to another recording. Fill in consistent ambience beds under constructed sentences and ADR lines right from within Pro Tools AudioSuite. No more spending hours of time manually looping sections of ambience under a scene with ADR and broken up sentences.

Seamlessly connect dialogue and ADR cuts with real background movement and textures with Complex mode, allowing you to quickly match never before complex ambiances: ocean, wind, traffic, crowds, and more. Or, use Static mode to match stationary ambiances, like a quiet room with a soft hum.

Ambience Match de iZotope



smart:limit

The content-aware limiter

Keeping Loudness and Dynamics in tune

Try now for 30 days

Buy smart:limit now

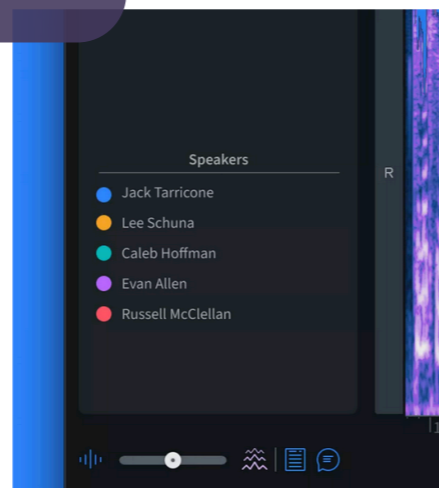
Smart Limit de Sonible

PLUGINS PARA DAWs

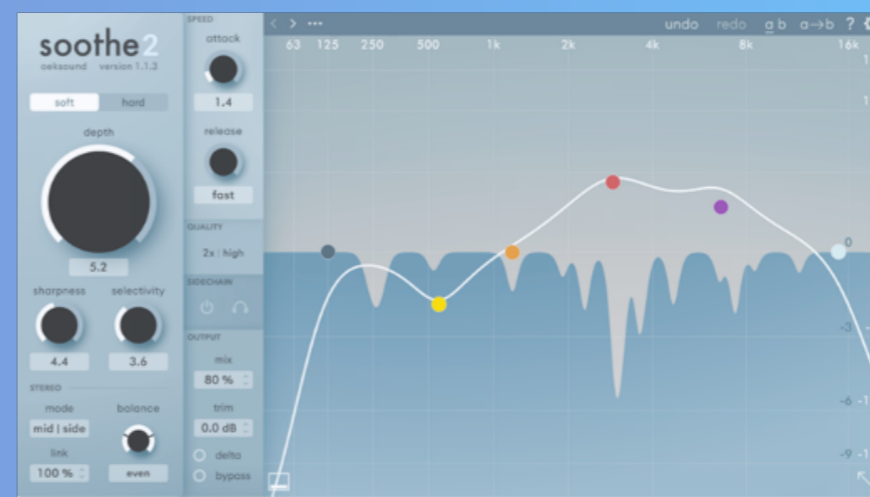
... más!

Multiple Speaker Detection

Are you working on a single audio track with more than one speaker? Save time navigating through the file by using Multiple Speaker Detection to help find and tag the sections of speech associated with each individual voice. This is incredibly helpful, especially when different speakers require different processing.



Multiple speaker detection, de iZotope



Soothe 2, de Oeksound

Audiorake

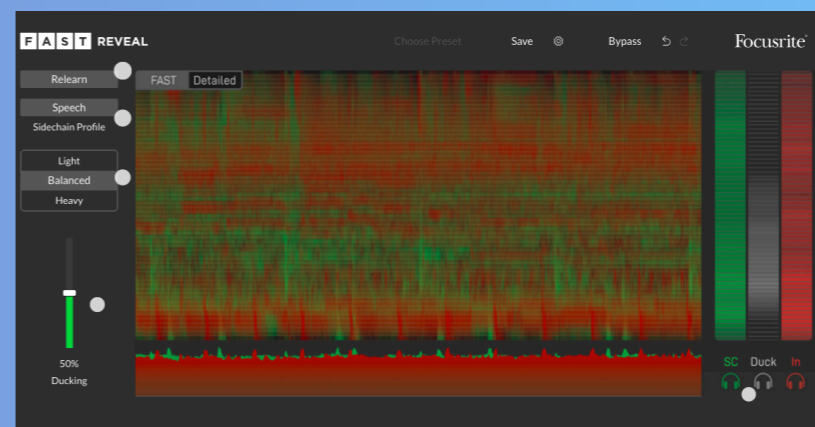
Audiorake Preps Your Vocals So You Don't Have To.

Audiorake cleans your vocals with 0 artefacts guaranteed. Perfect for live recordings and breath control.

[View Pricing](#) [Try Free](#)



Audiorake, de Parsonos



Fast Reveal, de Focusrite

EJEMPLO DE WORKFLOW

DAW

LIMPIEZA: CLEAR

DE-PLOSIVE / DE-MOUTH.
CLICK

ECUALIZADOR

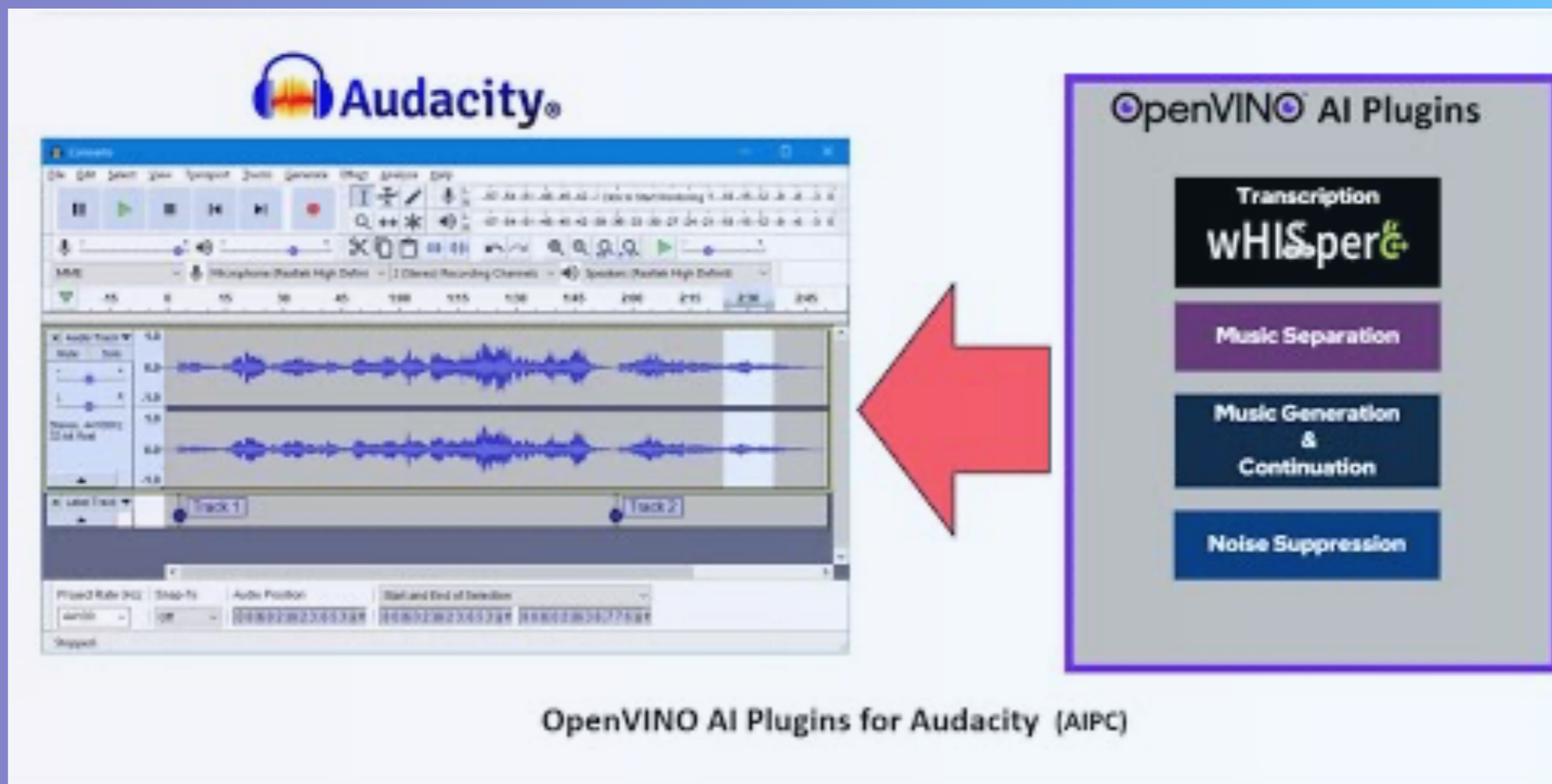
COMPRESOR

CONTROL RESONANCIAS

SIDECCHAIN

LIMITADOR

STREAM CONTROL



Audacity



Audacity

Introducing OpenVINO AI effects for Audacity

By Leo Wattenberg

TUESDAY, JANUARY 2, 2024

Intel has built a suite of AI tools for Audacity, useful for spoken word audio and music alike. These AI features run 100% locally on your PC.

AI tools for podcasts

For spoken word content, the OpenVINO effects contain a noise suppression and a transcription plugin.

The **Noise Suppression** does what it says on the tin - it suppresses noise. As such it behaves similar to Audacity's built-in Noise Removal effect.

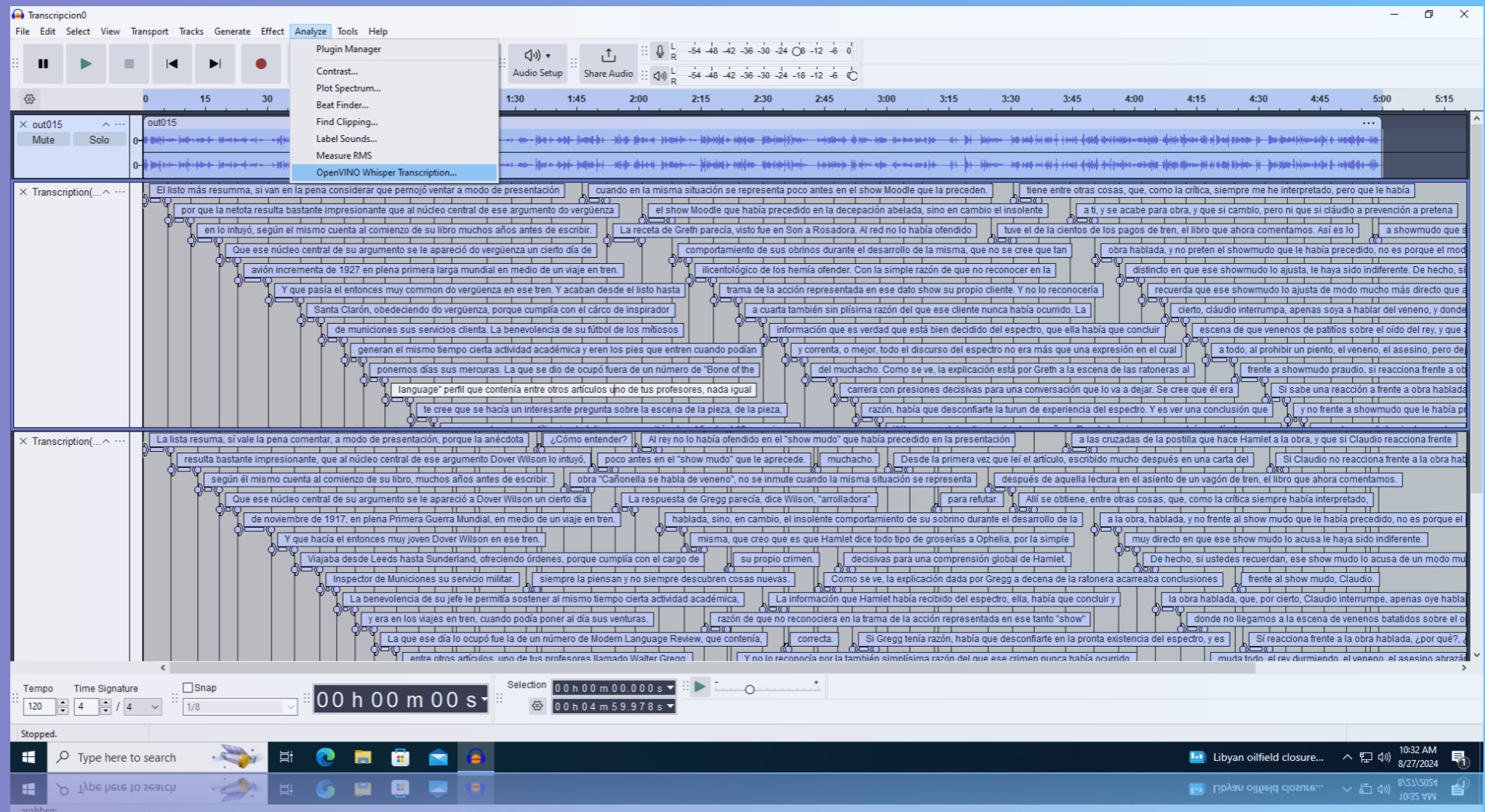
- [More information on noise suppression](#)

The **Transcription** powered by [Whisper.cpp](#) can both transcribe and translate words and outputs to a label track. If you want to export these transcriptions, you can do so via File → Export Other → Export Labels.

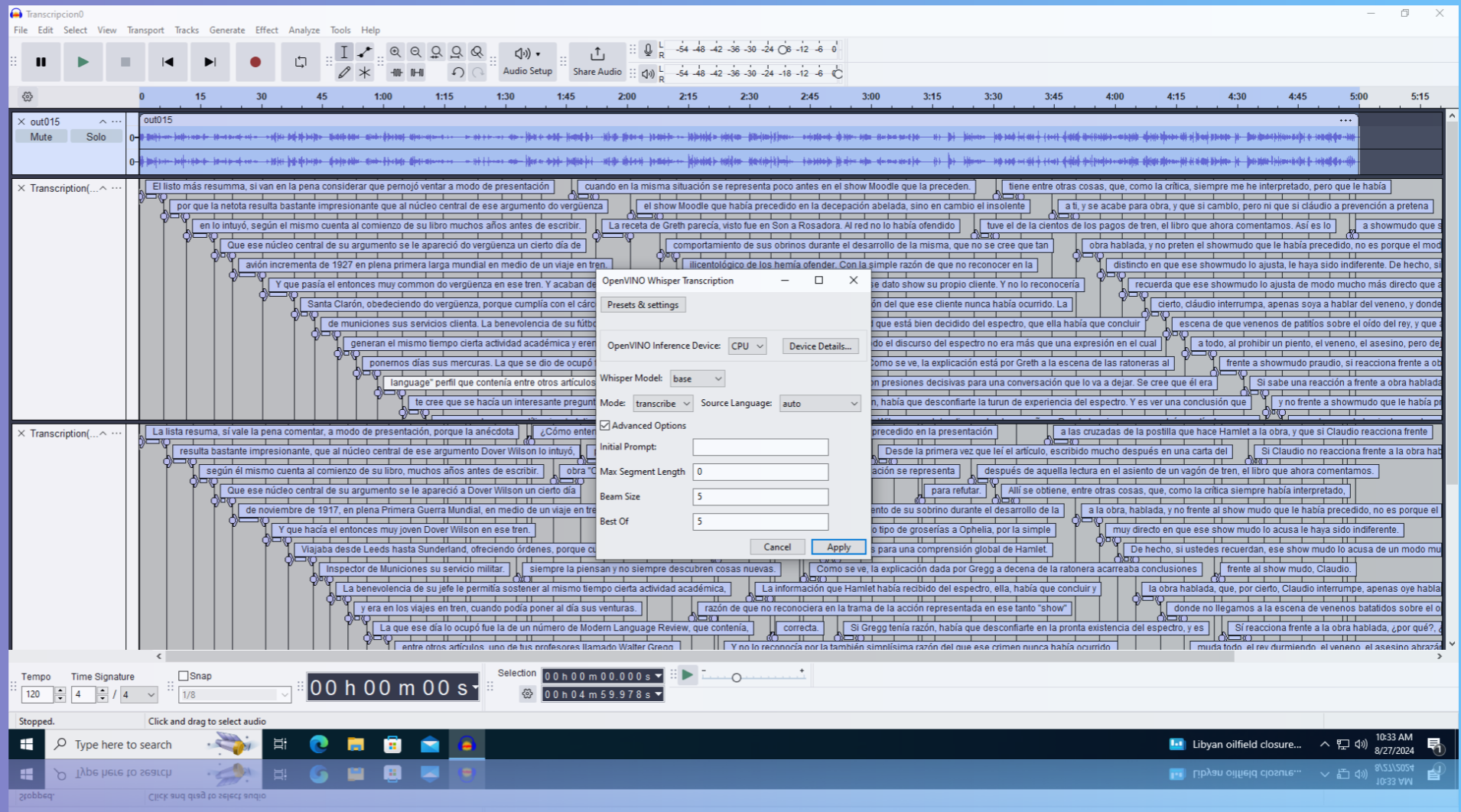
- [More information on transcription](#)

- [More information on transcription](#)

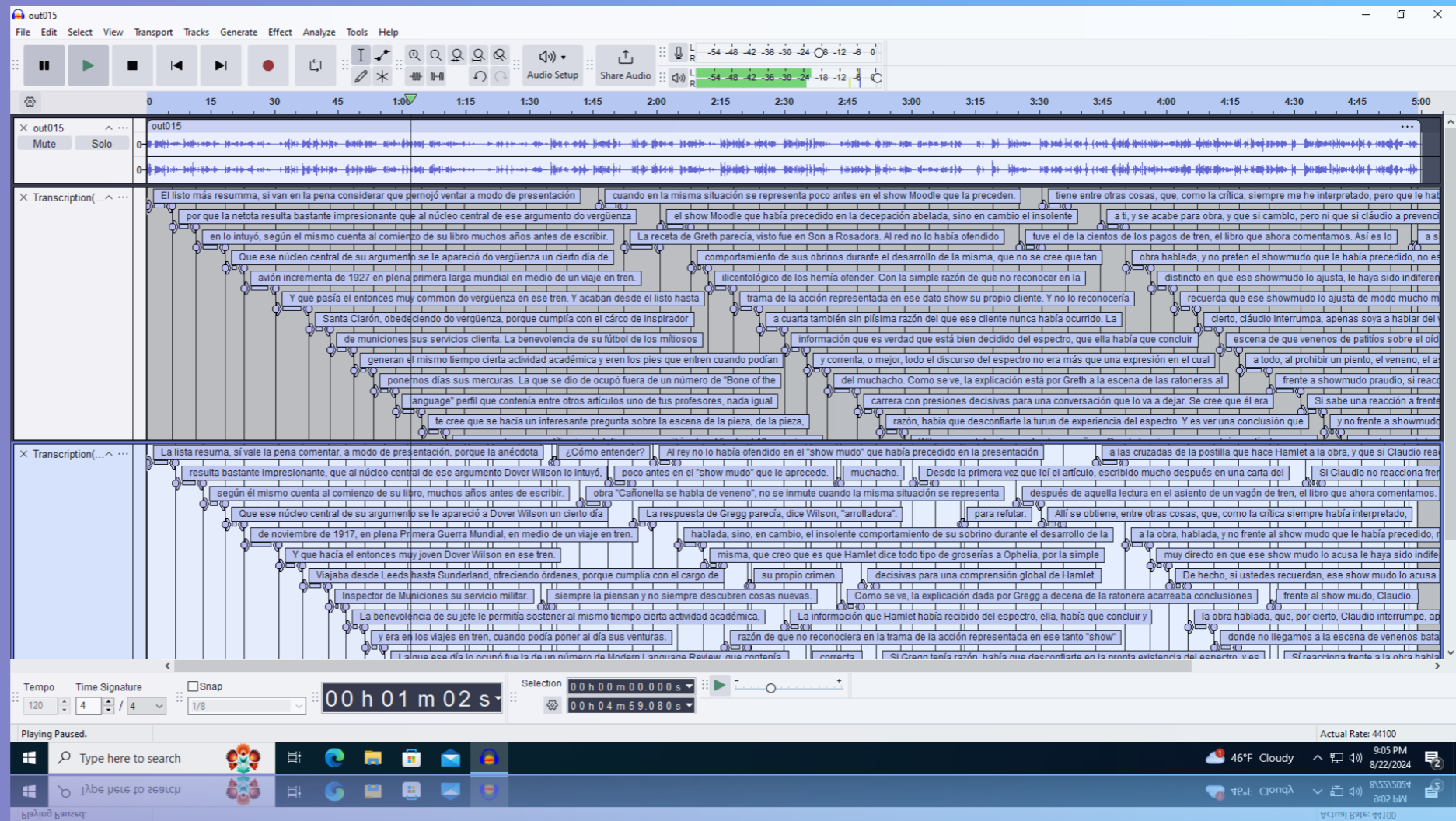
you can do so via File → Export Other → Export Labels.
words and outputs to a label track. If you want to export these transcriptions?



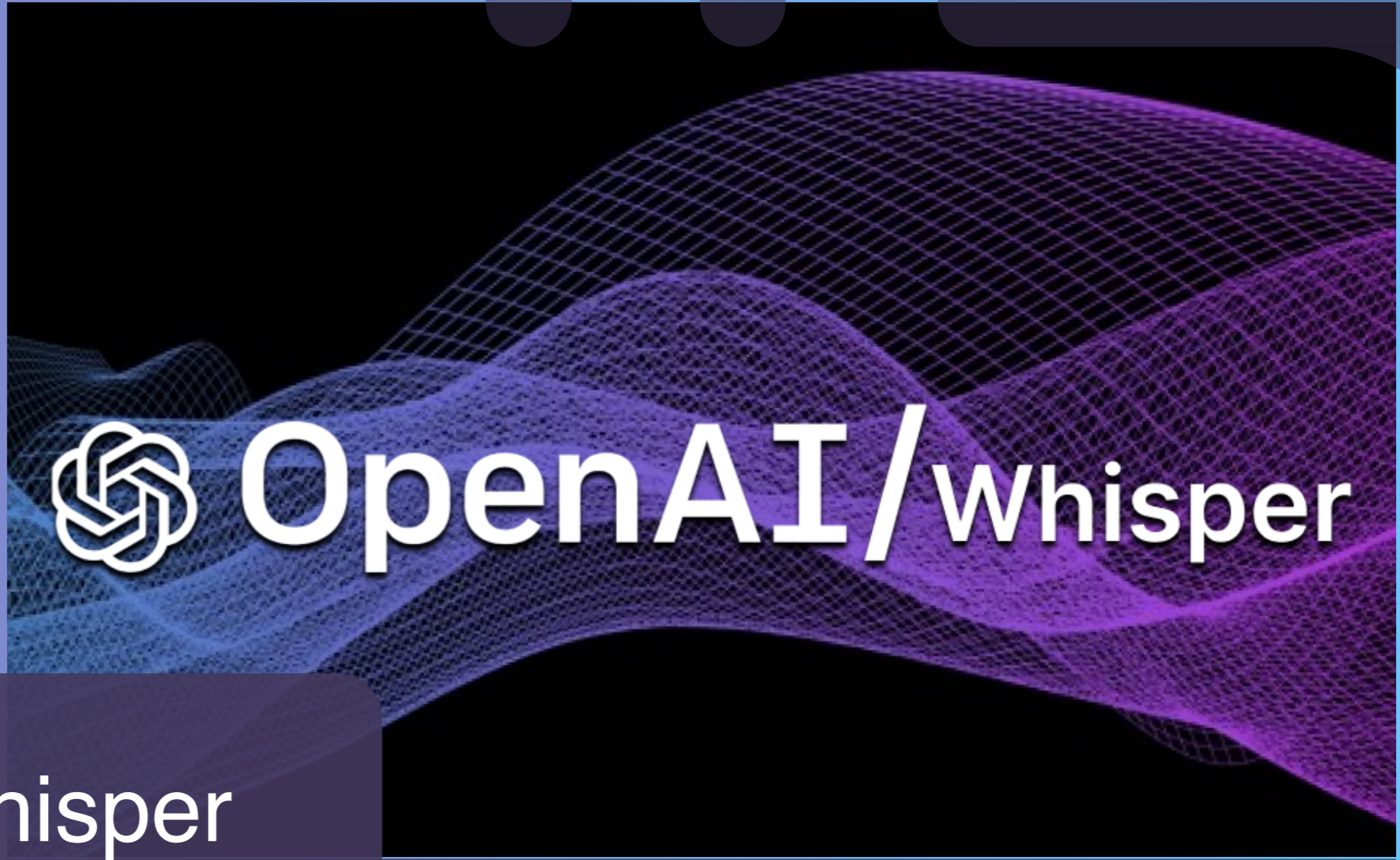
Audacity



Audacity



Audacity



OpenAI Whisper





Speech Studio > Voice Gallery

Voice catalog Examples by use case

Search Language: Spanish (Argentina) Sort by: Recommendation Show only pinned 2 voices

Filter

Open AI voices are available in two regions: Sweden Central and North Central US. Switch to a supported region to view the voices.

Elena
Female
Spanish (Argentina)
A clear-sounding voice with great versatility that can adapt to any use case...

Tomas
Male
Spanish (Argentina)

Voice details Try it out Sample code

Elena
Female - Adult
Spanish (Argentina)

Speaking styles: 1 style
Default

Description: A clear-sounding voice with great versatility that can adapt to any use case and speak in a way that's easy to understand.

Go to Audio Content Creation

Speech Studio > Audio Content Creation > My files > Entrevista a ChatGPT > 01-15

File Save Export Template Auto predict SSML 00:00 24kHz

1. [Elena] [lexicon] ChatGPT es un modelo de IA conversacional desarrollado por Open AI. Forma parte de la familia de modelos lingüísticos GPT (transformador generativo pre-entrenado), que utilizan técnicas de aprendizaje profundo para generar texto similar al humano. ChatGPT funciona utilizando una red neuronal profunda con un gran número de parámetros que han sido entrenados en una cantidad masiva de datos de texto. La red se ha entrenado para predecir la siguiente palabra de una frase, dadas las palabras precedentes. Durante el proceso de entrenamiento, el modelo aprende patrones en los datos de texto y construye una representación del lenguaje subyacente y su estructura. Cuando un usuario introduce una pregunta, ChatGPT la utiliza para generar una continuación del texto, que puede utilizarse como respuesta. El modelo utiliza la información de la pregunta para decidir qué palabras es más probable que vengan a continuación, teniendo en cuenta el contexto y el estilo del lenguaje utilizado. Una de las principales ventajas de este método es que ChatGPT puede generar textos muy pertinentes y apropiados para cada pregunta, lo que le permite mantener conversaciones y responder a preguntas de forma similar a un ser humano. En general, ChatGPT es un modelo lingüístico muy avanzado capaz de generar texto fluido e informativo, lo que lo convierte en una herramienta útil para una amplia gama de aplicaciones, como la IA conversacional, la traducción de idiomas y la creación de contenidos.

2. [Elena] [lexicon] El aprendizaje profundo es un subcampo del aprendizaje automático (machine learning) que utiliza redes neuronales con muchas capas (de ahí el nombre "profundo") para aprender patrones en grandes cantidades de datos. Se basa en la idea de que un modelo puede aprender a realizar una tarea si se expone a muchos ejemplos, igual que un niño puede aprender un concepto viendo muchos ejemplos. Una red neuronal es un tipo de modelo de aprendizaje automático inspirado en la estructura y el funcionamiento del cerebro humano. Está formada por nodos interconectados, o "neuronas", que procesan y transmiten información. En una red neuronal, las entradas se procesan a través de varias capas de neuronas, cada una de las cuales realiza un cálculo simple y pasa el resultado a la siguiente capa. La última capa produce la salida, que puede ser una predicción, una clasificación u otra forma de cálculo. Los parámetros de una red neuronal, como los pesos y los sesgos de las neuronas, se aprenden mediante entrenamiento. Durante el entrenamiento, el modelo se expone a un gran número de ejemplos y sus parámetros se ajustan para que las predicciones del modelo sean lo más precisas posible. En el caso de ChatGPT, las técnicas de

Total characters: 19,897 / 20,000

Tuning Comment

Recent

Voice

Language: Spanish (Argentina)

Elena Tomas

Break

Standard Advanced Custom

Silence

+ Add a silence type

Pronunciation

Alias Phoneme Reading rules

Click to edit phonemes

Lexicon

Default

Intonation


Click to modify intonation

Azure Speech Studio

HERRAMIENTAS TEXTO A AUDIO

Spaces | coqui/x tts | like 1.84k | Running on T4

Hugging Face is way more fun with friends and colleagues! [Join an organization](#)



This demo is currently running XTTS v2.0.3 XTTS is a multilingual text-to-speech and voice-cloning model. This demo features zero-shot voice cloning, however, you can fine-tune XTTS for better results. Leave a star on Github [XTTS](#), where our open-source inference and training code lives.

Supported languages: Arabic: ar, Brazilian Portuguese: pt, Mandarin Chinese: zh-cn, Czech: cs, Dutch: nl, English: en, French: fr, German: de, Italian: it, Polish: pl, Russian: ru, Spanish: es, Turkish: tr, Japanese: ja, Korean: ko, Hungarian: hu, Hindi: hi

CoquiTTS | Stars 33k

Documentation | [ReadTheDocs](#)

Questions | [GitHub Discussions](#)

Community | chat on discord 1.1k online

Text Prompt
One or two sentences at a time is better. Up to 200 text characters.

Hi there, I'm your new voice clone. Try your best to upload quality audio.

Language
Select an output language for the synthesised speech

es

Reference Audio

0:00 / 0:11

Use Microphone for Reference

Record from microphone

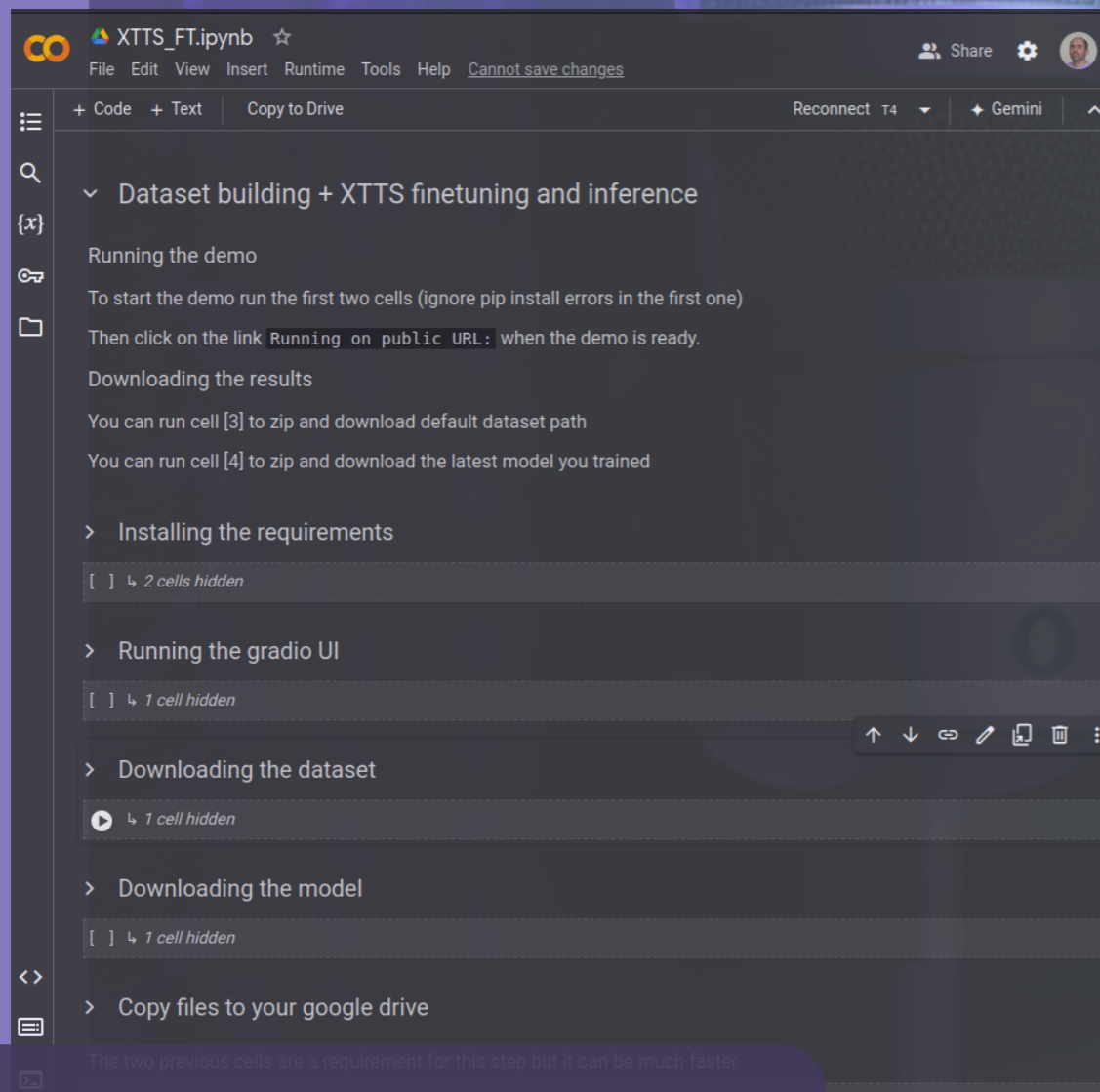
Waveform Visual

Synthesised Audio

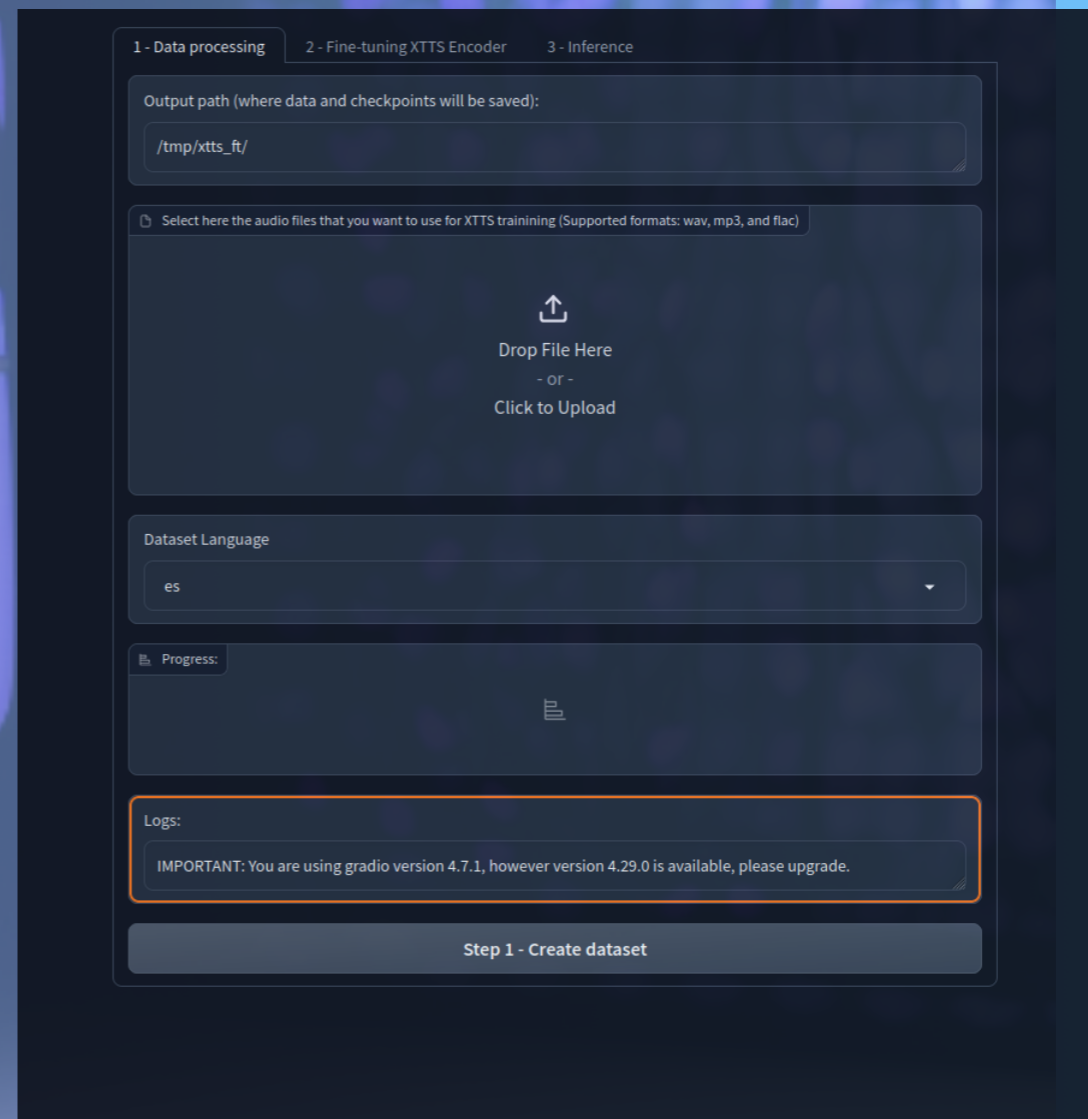
Metrics

COQUI TTS

HERRAMIENTAS TEXTO A AUDIO



The screenshot shows a Jupyter Notebook titled "XTTS_FT.ipynb" with a dark theme. The interface includes a top menu bar with options like File, Edit, View, Insert, Runtime, Tools, and Help. Below the menu, there are tabs for "+ Code", "+ Text", and "Copy to Drive". The main content area is divided into sections: "Dataset building + XTTS finetuning and inference", "Running the demo", "Installing the requirements", "Running the gradio UI", "Downloading the dataset", "Downloading the model", and "Copy files to your google drive". Each section contains code cells, some of which are collapsed and labeled "1 cell hidden". A toolbar at the bottom of the notebook provides navigation and editing tools.



The screenshot displays a Gradio web interface for XTTS. It features three tabs: "1 - Data processing", "2 - Fine-tuning XTTS Encoder", and "3 - Inference". The "1 - Data processing" tab is active, showing an "Output path" field set to "/tmp/xtts_ft/". Below this is a file selection area with the instruction "Select here the audio files that you want to use for XTTS training (Supported formats: wav, mp3, and flac)". A large "Drop File Here" area with an upload icon and "Click to Upload" text is provided. A "Dataset Language" dropdown menu is set to "es". A "Progress" section shows a progress bar and a document icon. A "Logs" section contains a warning message: "IMPORTANT: You are using gradio version 4.7.1, however version 4.29.0 is available, please upgrade." At the bottom, a button labeled "Step 1 - Create dataset" is visible.

COQUI TTS

HERRAMIENTAS TEXTO A AUDIO

1 - Data processing 2 - Fine-tuning XTTS Encoder 3 - Inference

Train CSV:
/tmp/xtts_ft/dataset/metadata_train.csv

Eval CSV:
/tmp/xtts_ft/dataset/metadata_eval.csv

Number of epochs: 6

Batch size: 2

Grad accumulation steps: 1

Max permitted audio size in seconds: 11

Progress: **Model training done!**

Logs:
IMPORTANT: You are using gradio version 4.7.1, however version 4.29.0 is available, please upgrade.

Step 2 - Run the training

1 - Data processing 2 - Fine-tuning XTTS Encoder 3 - Inference

XTTS checkpoint path:
August-21-2024_02+24AM-0000000/best_model.pth

XTTS config path:
/tmp/xtts_ft/run/training/XTTS_v2.0_original_model_files/config.json

XTTS vocab path:
/tmp/xtts_ft/run/training/XTTS_v2.0_original_model_files/vocab.json

Speaker reference audio:
/tmp/xtts_ft/dataset/wavs/output_00000039.wav

Language:
es

Input Text:
Sinceramente me voy porque me está echando la pandemia. Ser senador significa hablar con gente y andar por todos lados, el partido no se juega en los

Progress: **Model Loaded!**

Step 3 - Load Fine-tuned XTTS model

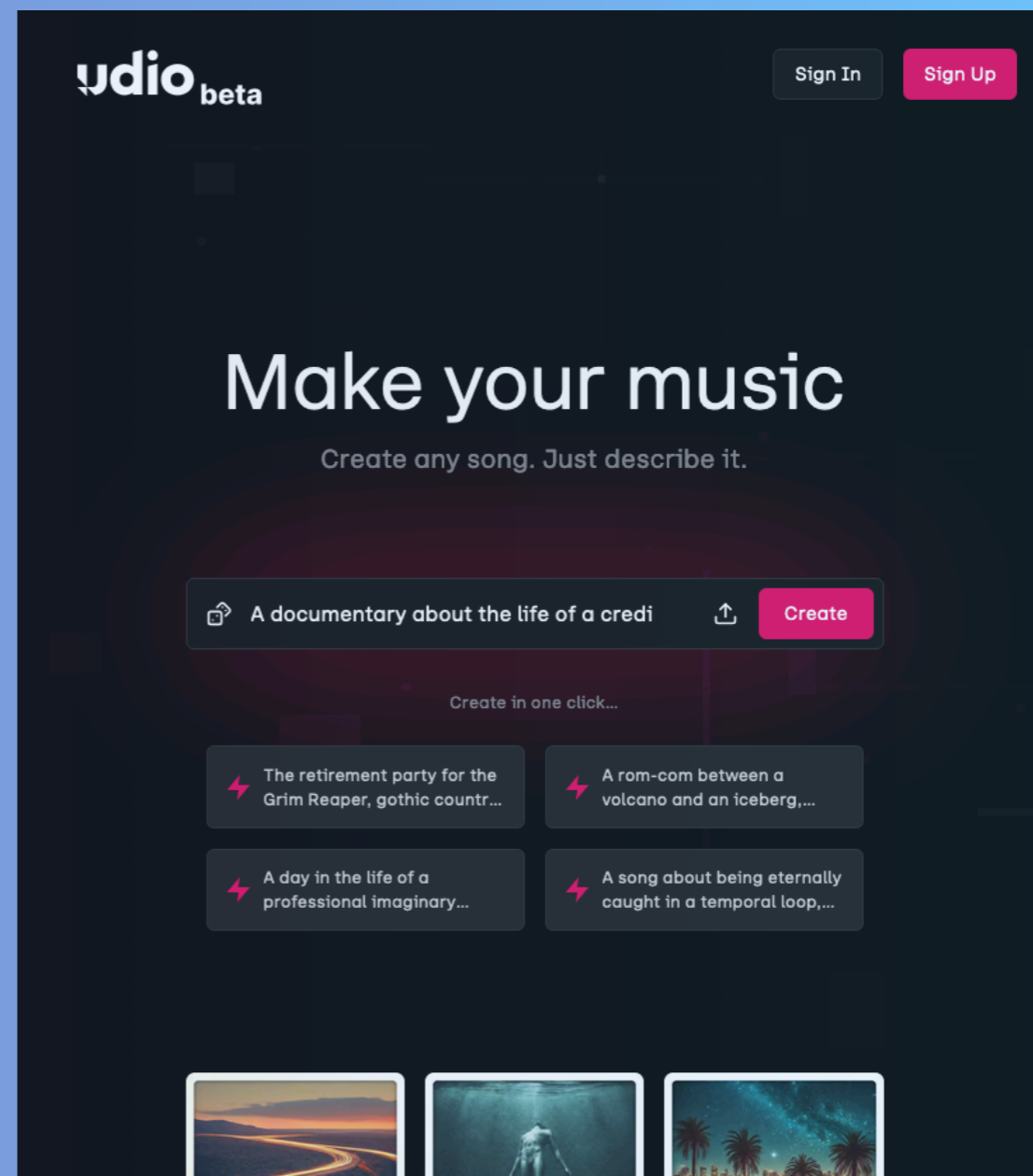
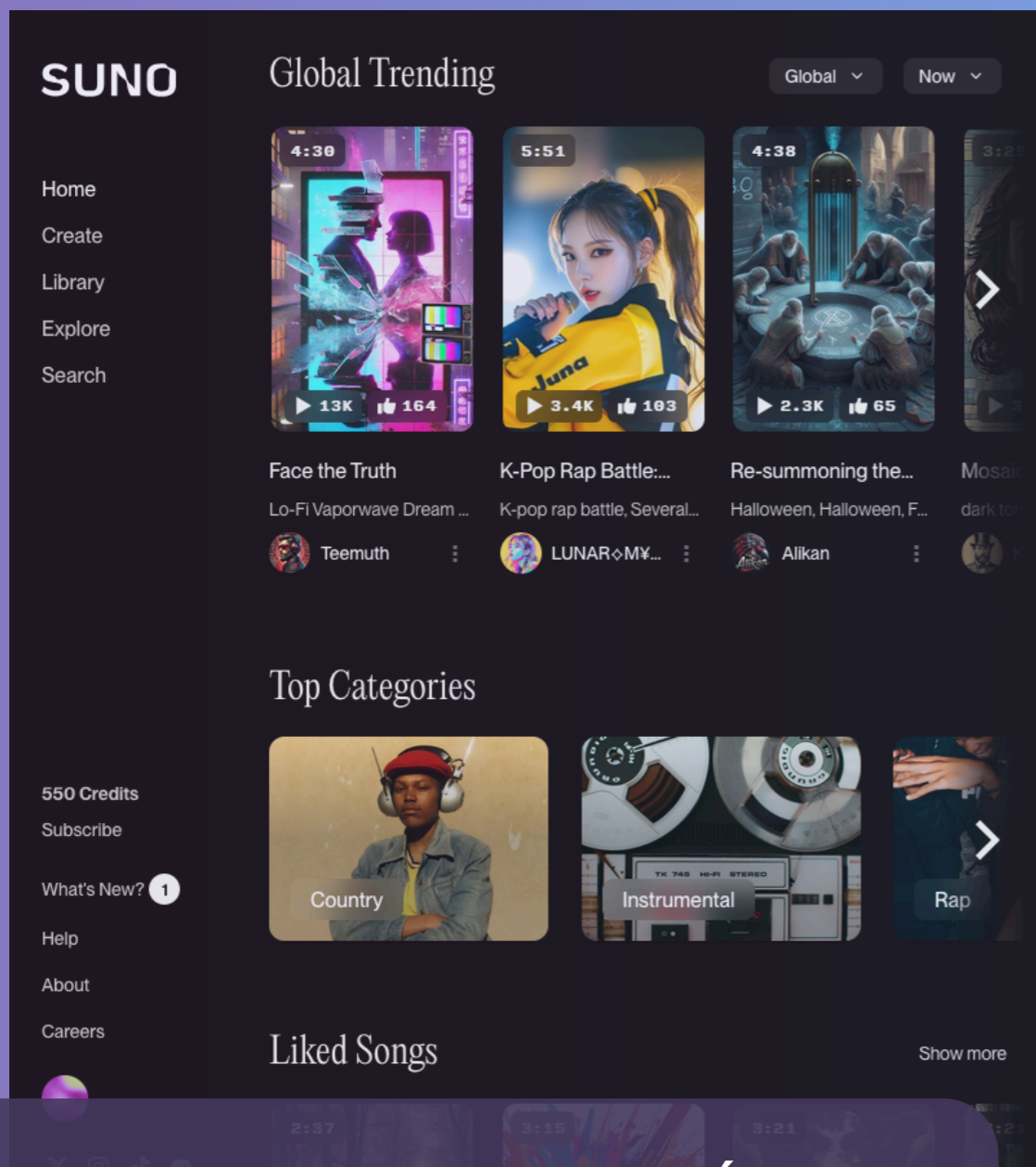
Progress: **Speech generated !**

Generated Audio:
0:08 0:12

Reference audio used:
0:36 0:36

COQUI TTS

BONUS TRACK



IA GENERATIVA DE MÚSICA:
SUNO / UDIO



David Coronel
Gastón Lema
Sergio Paoletti

Estefania Alvarez
Alejandro Biscione
Marcela Basch
Mariano Blejman
Diego Mascialino

Gracias por
ser parte!

LAIA en Media Party Buenos Aires 2024

WORKSHOP ORGANIZADO POR LAIA:

"Herramientas IA de audio para producción de podcasts"

SPEAKERS: SERGIO PAOLETTI / DAVID CORONEL / GASTON LEMA

Viernes 30 de agosto

13.00 hs

Ciudad Cultural Konex,
Sala D

Sarmiento 3131
Abasto, CABA



SUMATE!

Queremos estar,
queremos que estés.



www.laia.ar



@LaiaLab_



@laialab_



LAIA Entre la señal y el ruido



LAIA laboratorio



LAIA

Laboratorio Abierto de
Inteligencia Artificial