



AIID + Sound

INPUT

Acapella singing

Birds chirping

Carnatic singing

Cello performing

Pots and pans clanging

Synthesizer riffing

+ Add your own



Concert

TRANSFORMATION

None

Flute

Saxophone

Trumpet

Violin

Wan Fang



Southern University of Science and Technology



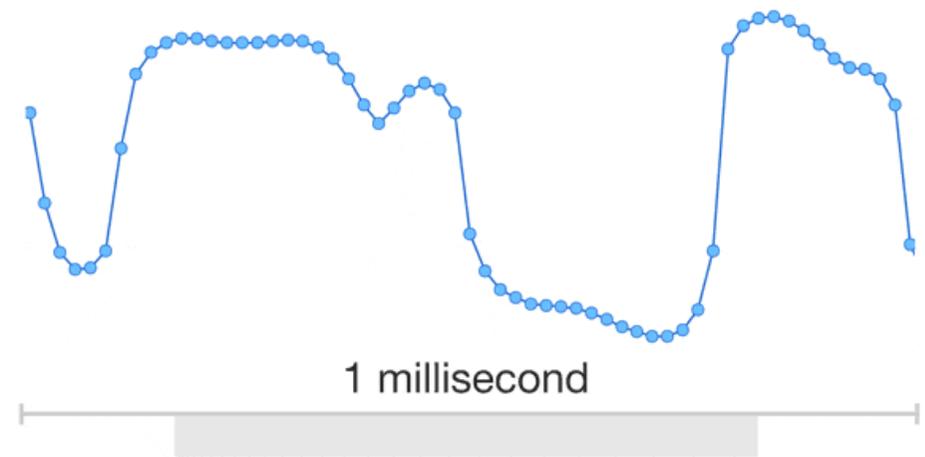
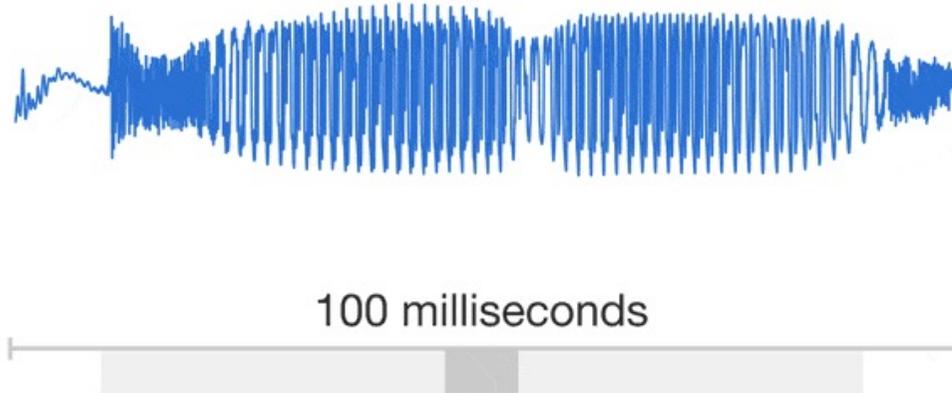
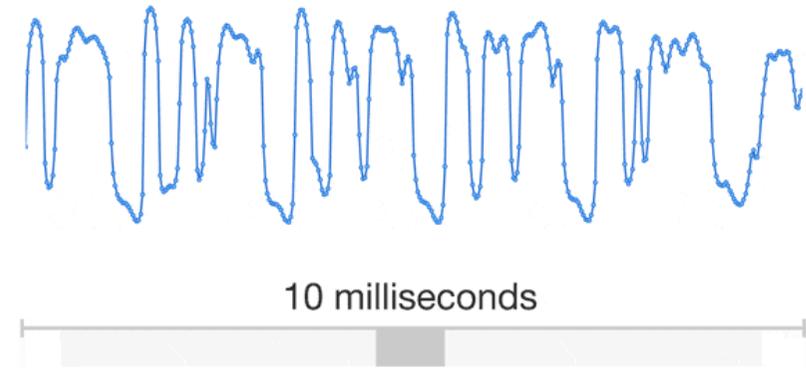
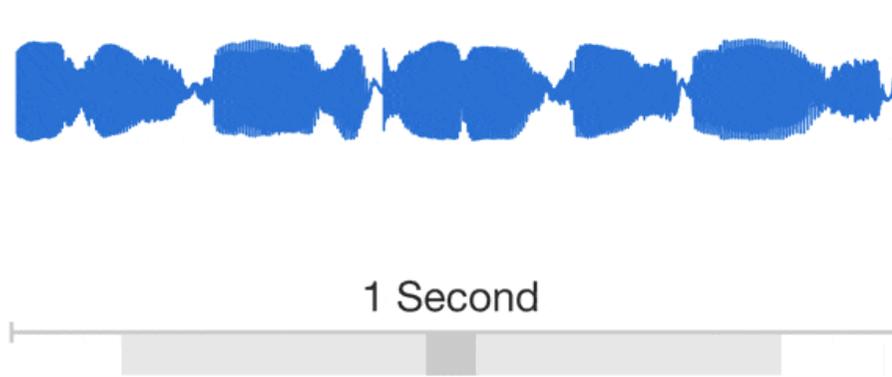
TONE
TRANSFER

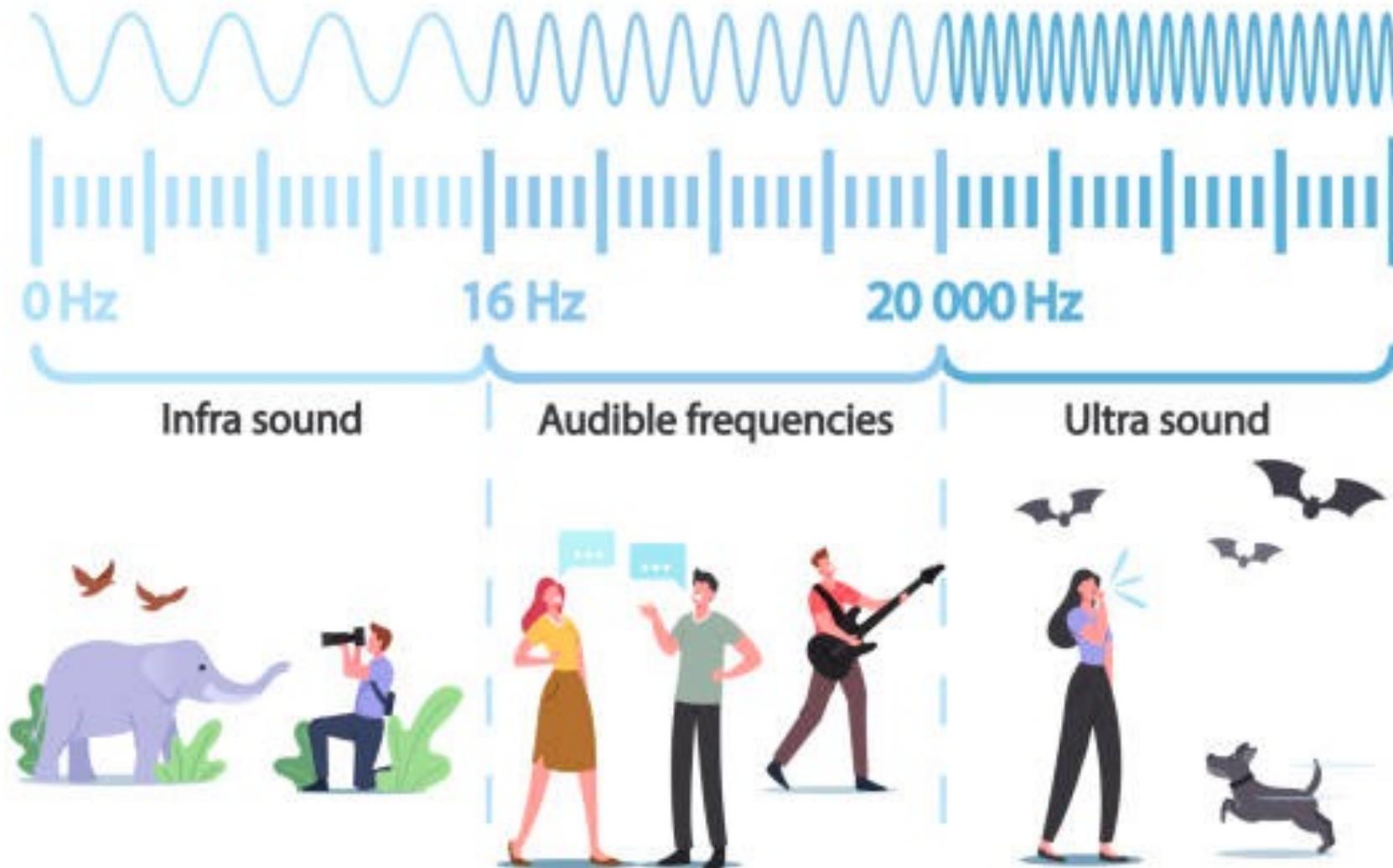


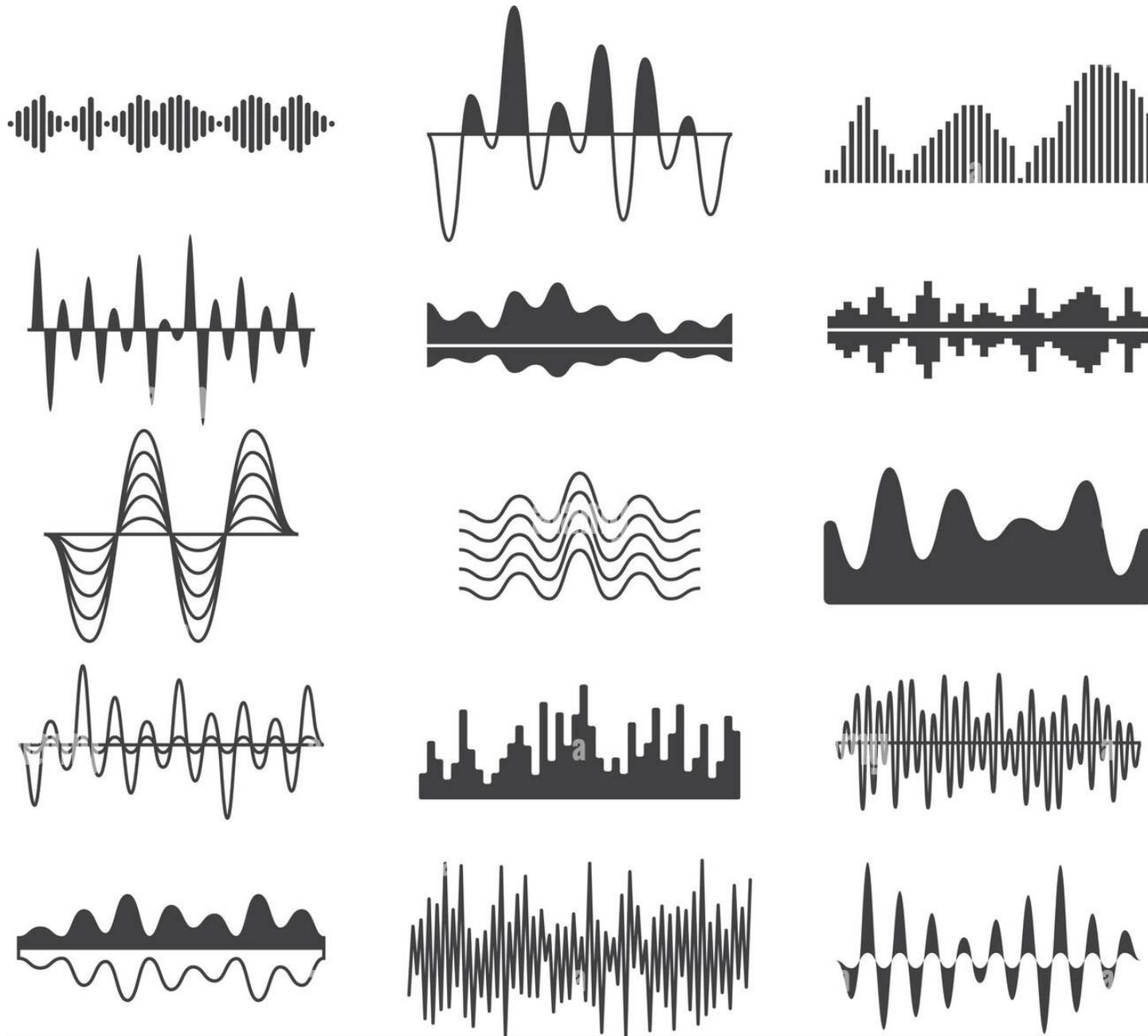
Discover more

Agenda

- Sound as Data
- Automatic Speech recognition (ASR)
- Voice Recognition
- Music Generation
 - Symbolic AI vs Audio AI systems
 - Tools to Make Your Own Generative Music
 - Concluding Thoughts and Further Questions



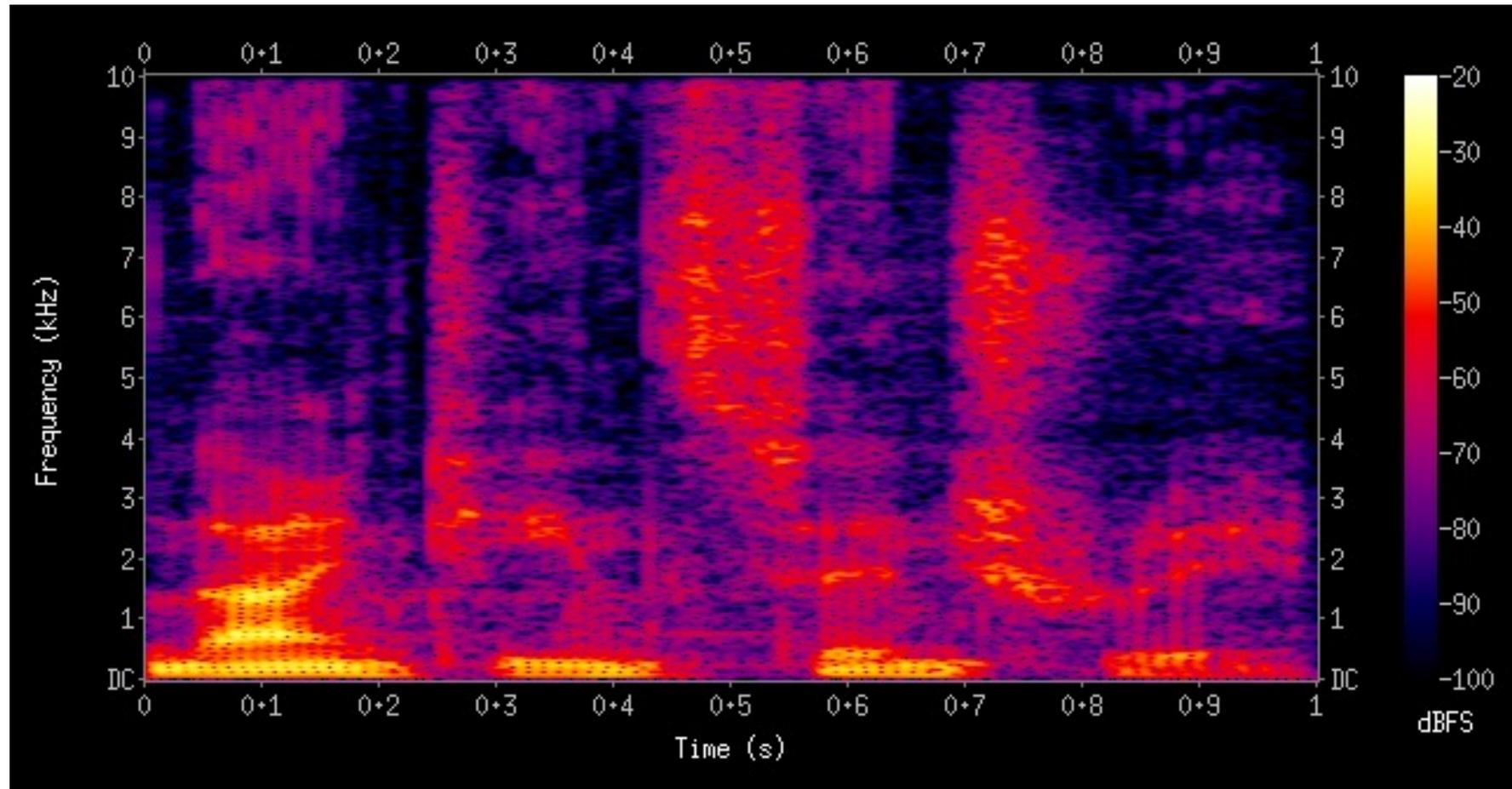




4 Properties of Sound

- Frequency
pitch, 音调
- Amplitude
音强
- Duration
音长
- Timbre
音色

Digitally produced spectrogram of a male voice saying 'nineteenth century'



<https://en.wikipedia.org/wiki/Spectrogram>

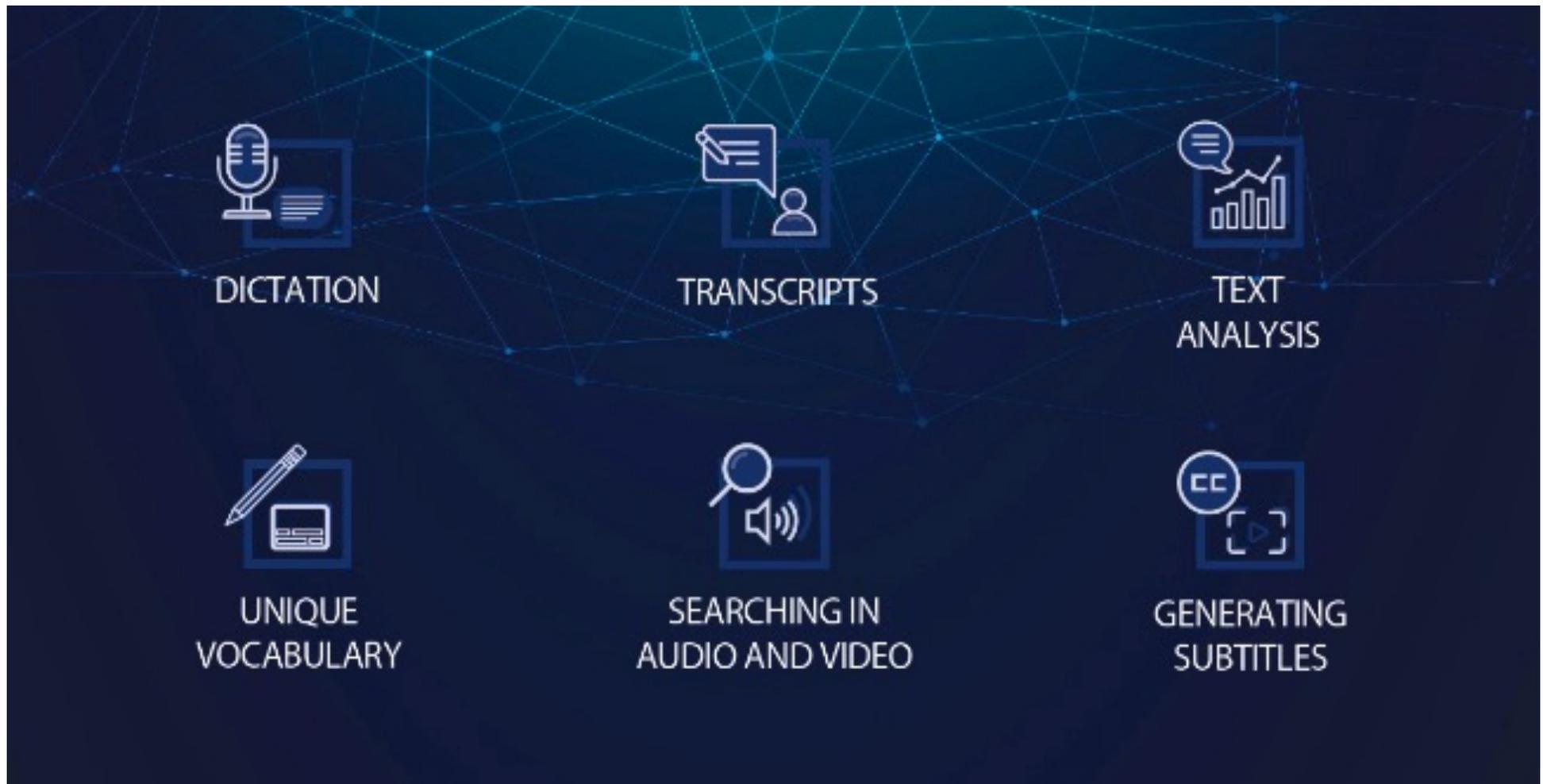
A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time.

中文是频谱图,它是一种可视化技术,通过颜色深浅来表示信号在不同时间、不同频率下的能量强度。

Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR)

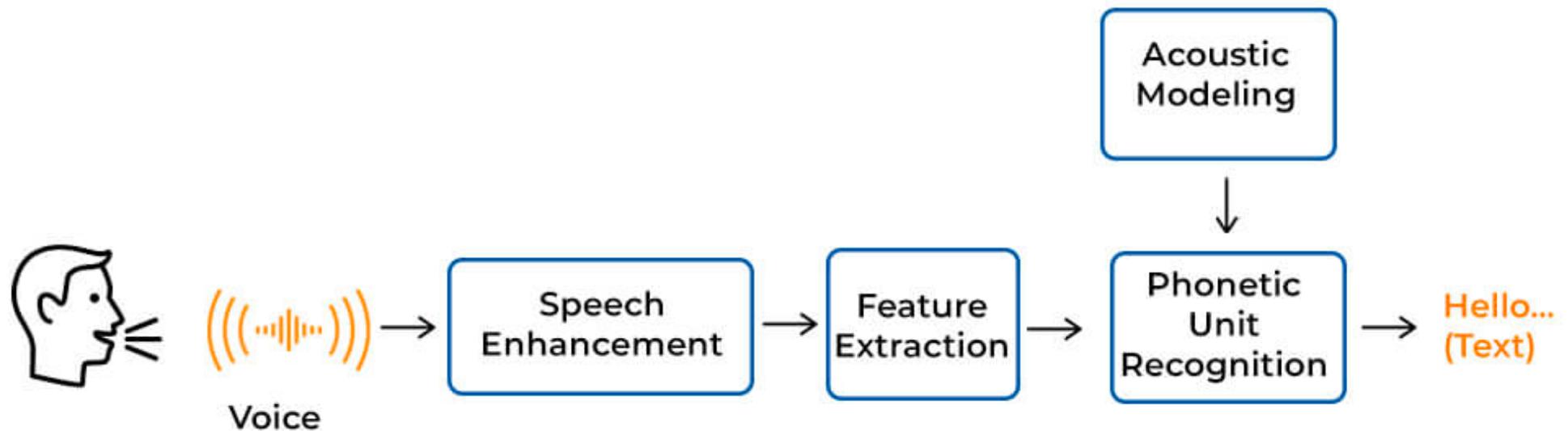
- Speech recognition is the ability of AI systems to identify spoken words and convert them into text.



Used to be like



SPEECH RECOGNITION PROCESS

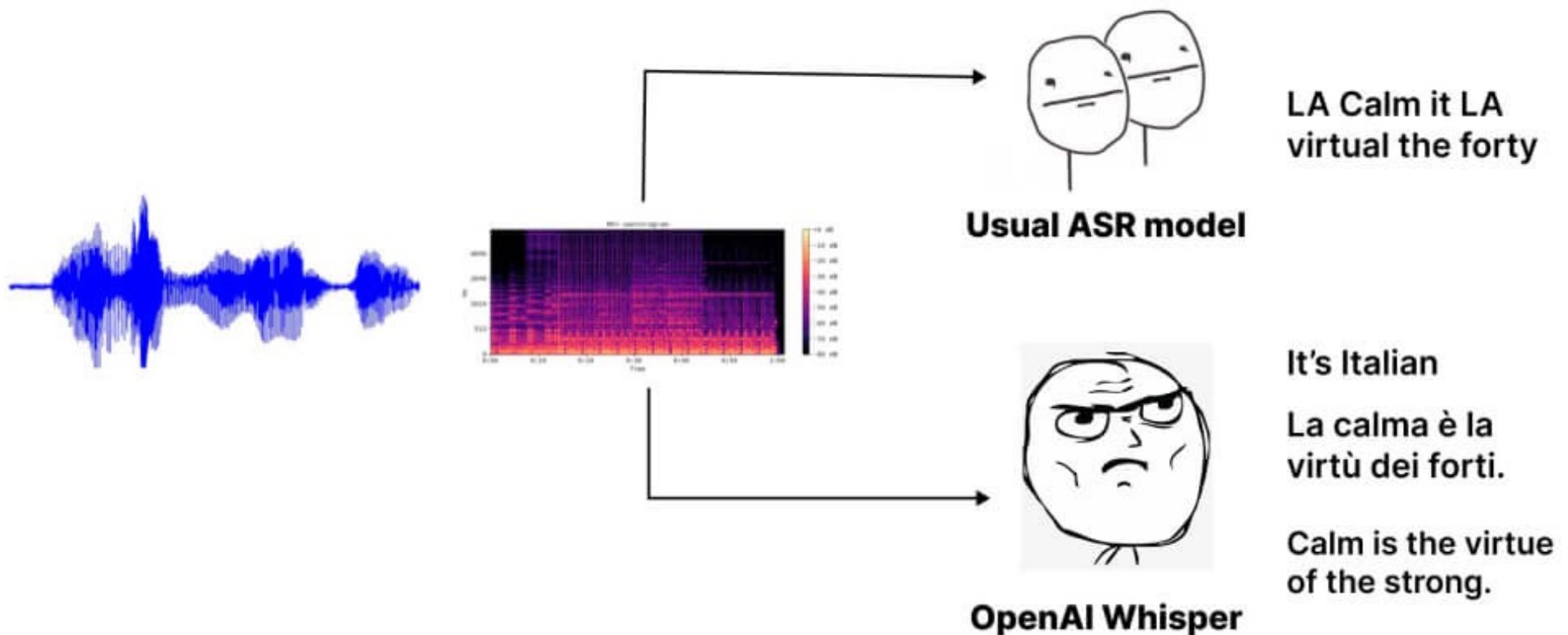


**WHEN YOU FACE A UNIVERSE
OF LABELED DATA**

**EMBRACE
END TO END DEEP LEARNING**

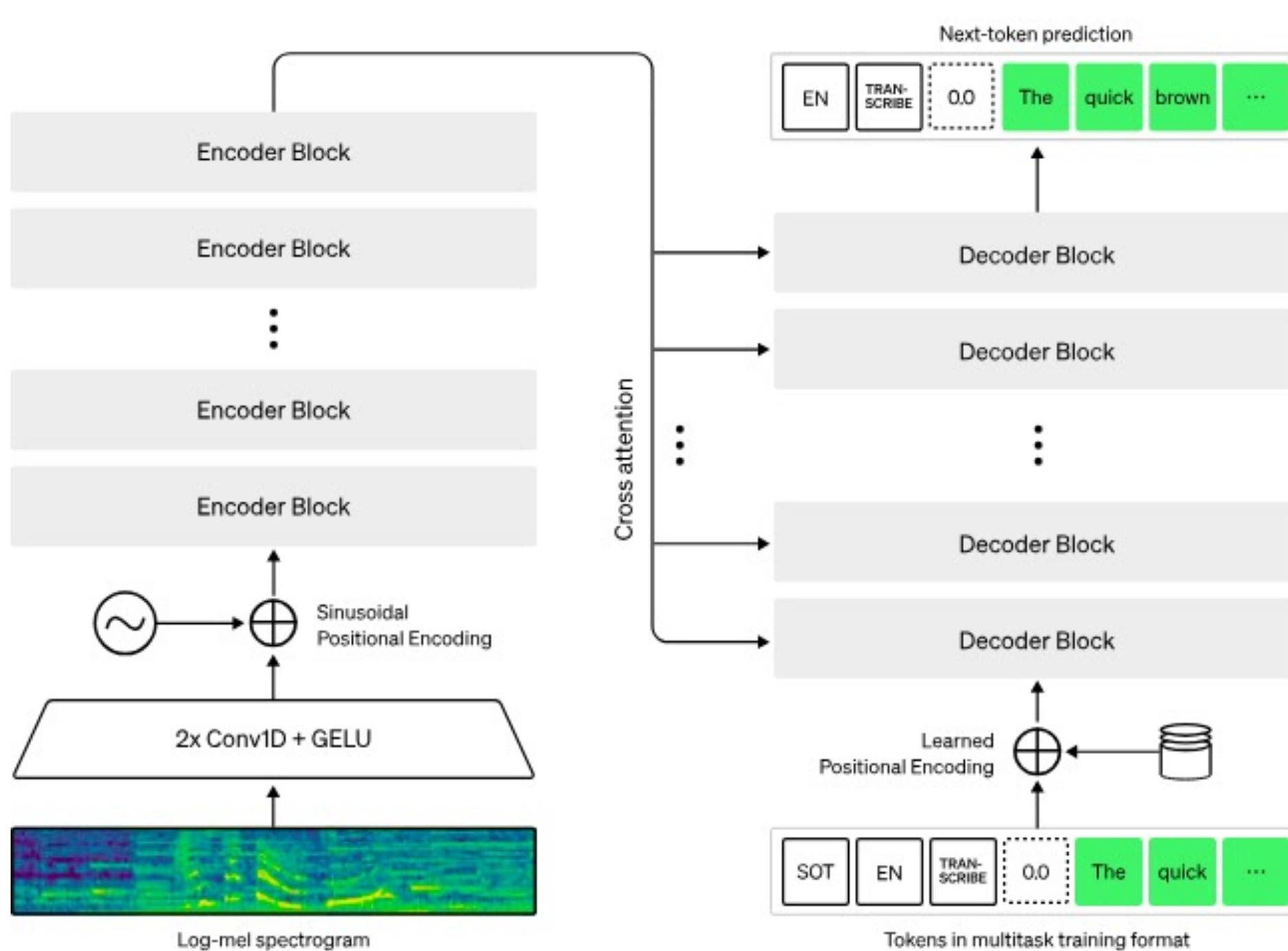
Whisper from OpenAI

- In 2022, this idea of training on large data to achieve cross-domain performance arrived in the world of speech recognition with OpenAI's launch of [Whisper](#).



Whisper from OpenAI

- 680,000 hours of multilingual and multitask supervised data
- A third of Whisper's audio dataset is non-English.



Whisper from OpenAI

- <https://huggingface.co/openai/whisper-large-v2>

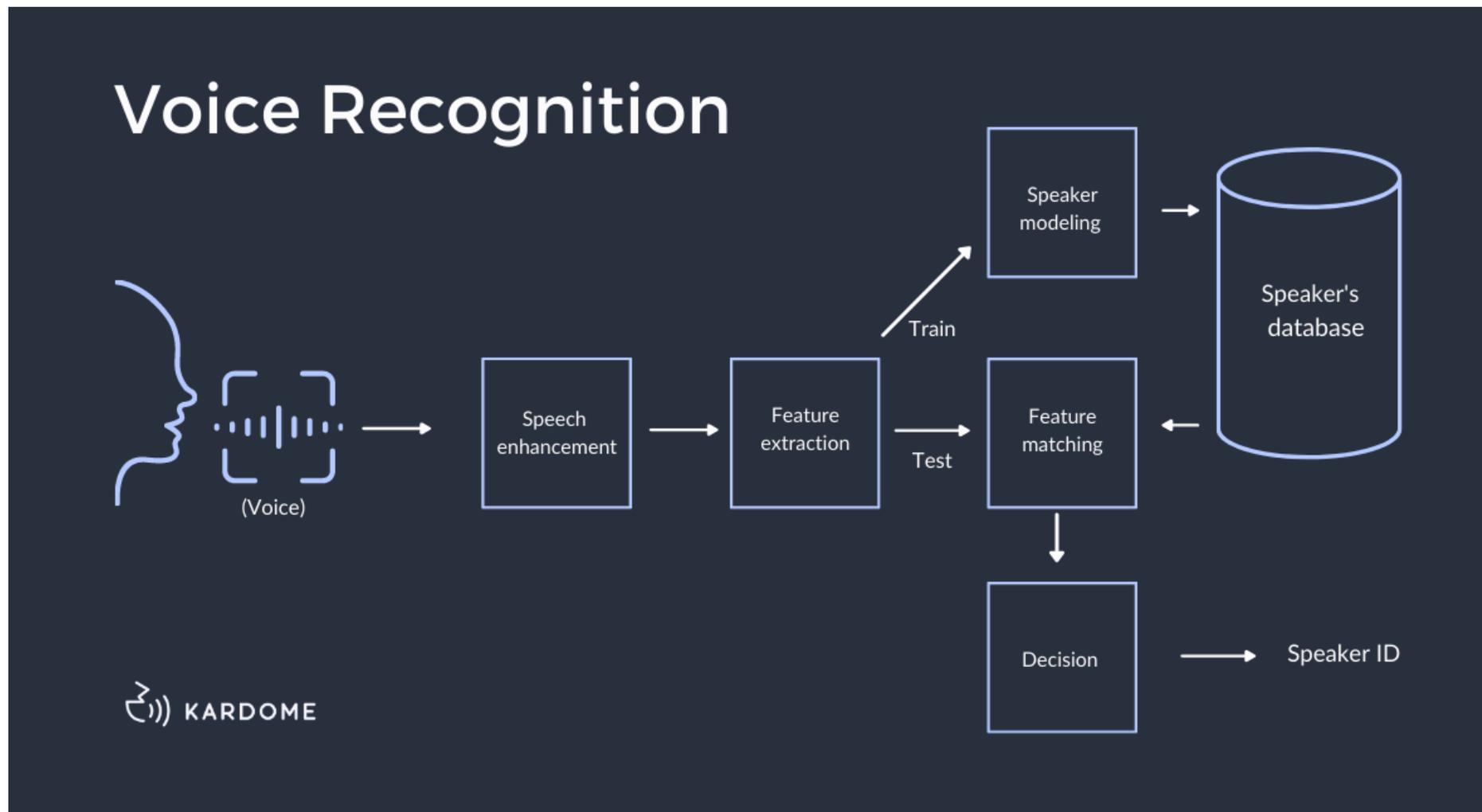
The screenshot displays the Hugging Face interface for the OpenAI Whisper Large V2 model. At the top, it shows 'Safetensors' and 'Model size 1.54B params' with a 'Tensor type F32' dropdown. Below this, the 'Hosted inference API' section is active, showing 'Automatic Speech Recognition' with an 'Examples' dropdown. There are three input options: 'Browse for file', 'Record from browser', and 'Realtime speech recognition'. A 'Librispeech sample 2' audio player is shown with a 'Compute' button. Below the player, it states 'Computation time on Intel Xeon 3rd Gen Scalable cpu: cached'. At the bottom, a green box contains the transcribed text: 'Before he had time to answer, a much-encumbered Vera burst into the room with the question,—'I say, can I leave these here?' These were a small black pig and a lusty specimen of black-red game-cock.'

Voice Recognition

Pay attention to the difference from speech recognition

Voice Recognition

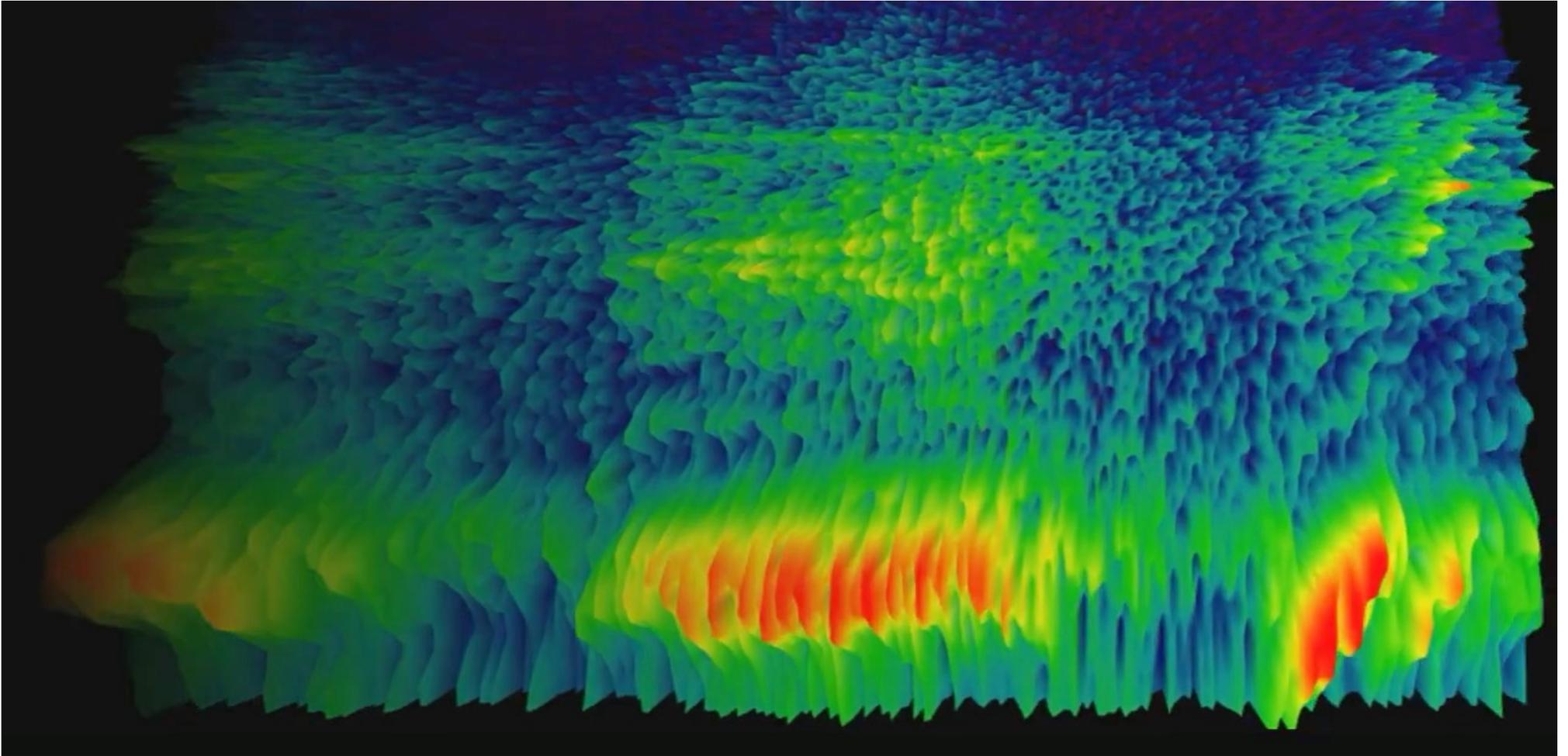
- Identify an individual user's voice (Biometric)



Recognize sounds in circumstances



Using AI to listen to all of Earth's Species



[More resources](#)

Using AI to listen to all of Earth's Species



Music Generation

Symbolic AI vs Audio AI systems

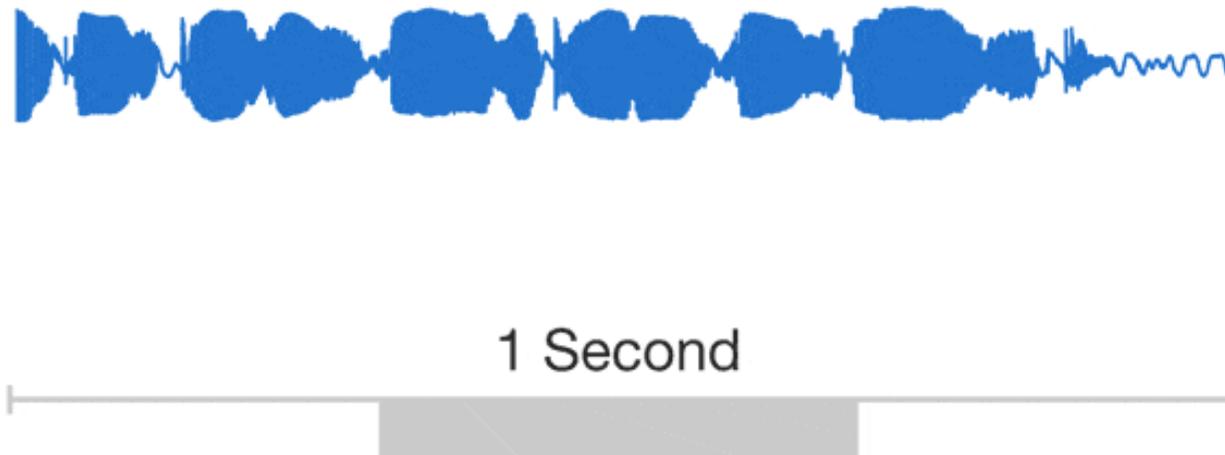
- Music AI generation into two broad camps: symbolic generation and audio generation
- A **symbolic** AI system generates the notes making up music
 - Exactly like a text generation model!
 - It requires a human to play the music notes, or additional music software to transform the notes into actual sound.

Computation time on Intel Xeon 3rd Gen Scalable cpu: 12.294 s

```
L1/8 Q:1/4=60 M:4/4 K:C "^Slowly and with feeling" z4 z2 z G | A2 B2  
c2 BA | G2 A2 G2 E2 | D4 z4 | z8 | A2 AB c2 Bc | d2 e2 d2 cB | A6 z2 | z6  
AB | c2 de d2 cd | e2 dc B2 AG | A8 |]
```

Symbolic AI vs Audio AI systems

- An **audio** generation model synthesizes the waveform of the music directly!
 - *This is a very challenging for machine learning task!*
 - *A full 3-minute song in stereo puts us at over a billion samples. Keeping musical coherency across the first sample to the millionth sample is a difficult task.*



Tone Transfer

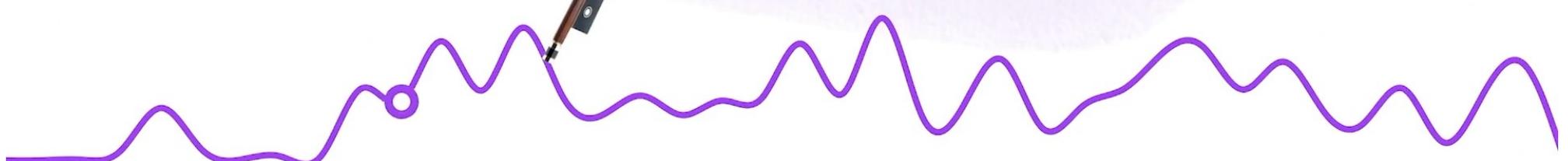
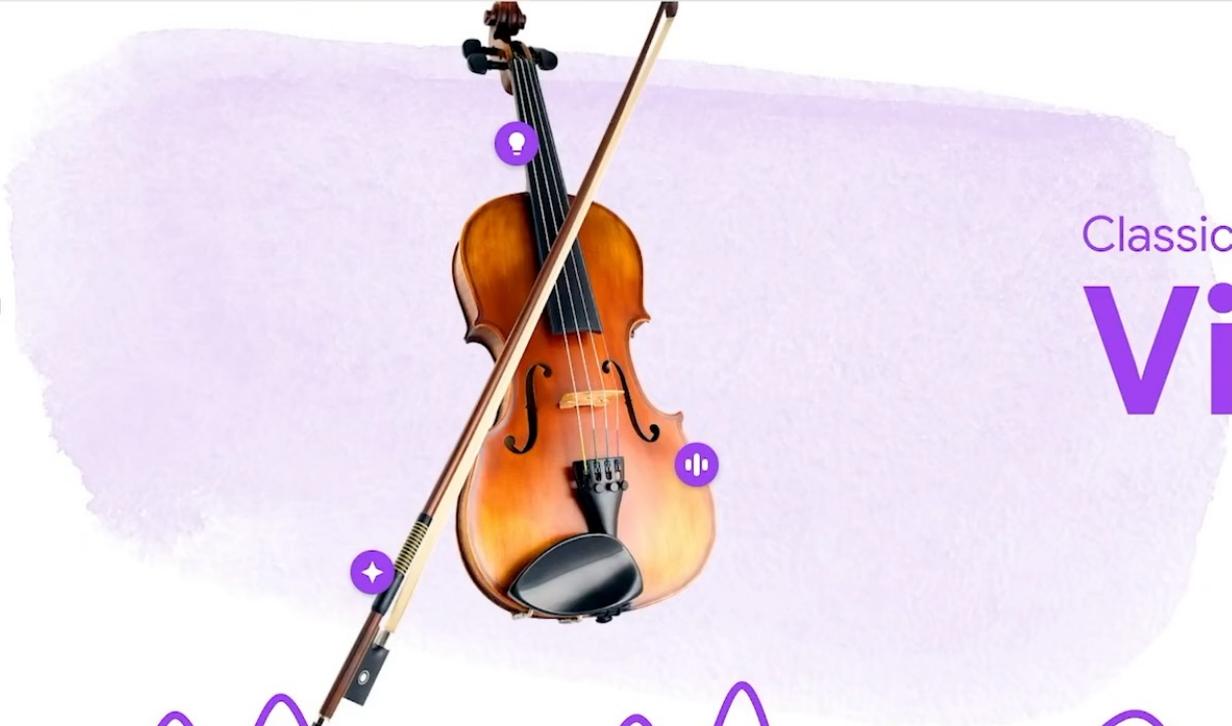
INPUT

- Acapella singing
- Birds chirping
- Carnatic singing
- Cello performing
- Pots and pans clanging
- Synthesizer riffing
- Your recording 

TRANSFORMATION

- None
- Flute
- Saxophone
- Trumpet

Classical
Violin



TONE
TRANSFER



Discover more

MusicLM: Generating Music From Text

- Not open source, published 2023
- **Dataset:** The [MusicCaps dataset](#) contains 5,521 music examples, each of which is labeled with an English aspect list and a free text caption written by musicians.

About MusicFX

MusicFX is an experimental technology that allows you to generate your own music. Certain queries that mention specific artists or include vocals will not be generated.

MusicFX is powered by Google's [MusicLM](#) and uses Google DeepMind's novel watermarking technology, [SynthID](#) to embed a digital watermark in the outputs.

We need your help to improve AI for everybody. Generated audio and prompt suggestions are experimental. You can [report](#) content under our policies or applicable laws, or give feedback by clicking the flag icon so we can improve AI responsibly together.

What is DJ mode

Generate a real-time stream of music by adding and adjusting musical prompts to evolve the music live.

- Steer the music by adjusting the musical prompts with the sliders. It may take a few seconds to adjust.
- Add up to 10 musical prompts. This can include instruments, genres, emotions, etc.
- If you get stuck, press reset to get fresh music from the same prompts.

Since MusicFX DJ is in high-demand, sessions are limited to 60 minutes and will automatically end if you're inactive for more than 10 minutes.

MusicFX: Generating Music From Text

The screenshot displays the MusicFX web application interface. At the top left, the "MusicFX" logo is visible. A "DJ mode" toggle switch is located at the top center. On the top right, there are icons for help, a menu, and a user profile labeled "F".

The main content area is split into two sections. On the left, a text input field contains the prompt: "Optimistic melody about the arrival of spring, full of joy and hope, tranquil flute in the background, upbeat with a gentle guitar riff". Below the text is a "tab" button. At the bottom of this section are a "Start over" button and a highlighted "I'm feeling lucky" button.

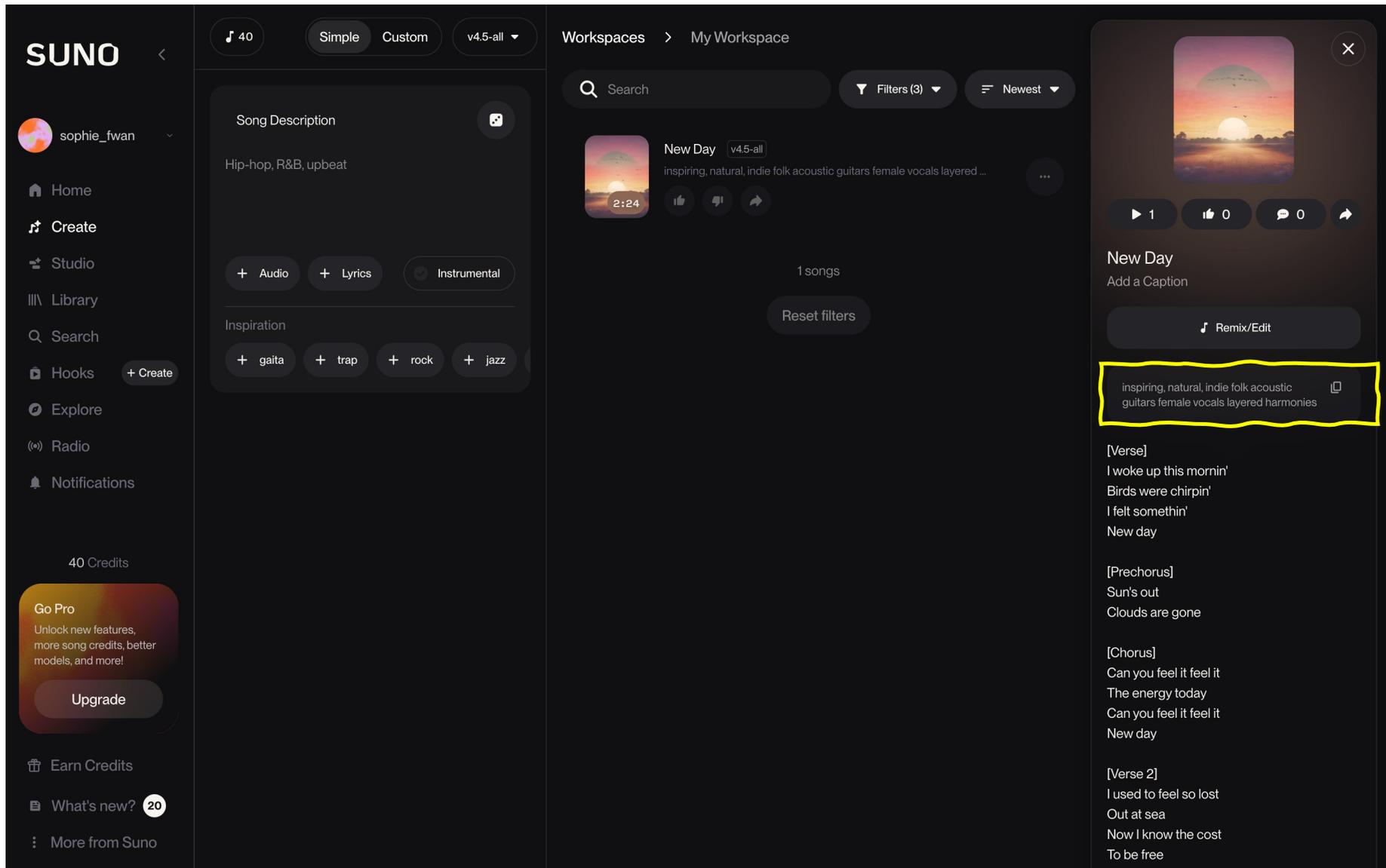
On the right, a "NO TRACKS YET" message is displayed between two navigation arrows. Below this is a placeholder for "YOUR BEAUTIFUL SONG" with a play button, a progress bar showing "0:00 / 0:30", and a settings menu icon.

At the bottom of the right section are three buttons: "Settings", "Download", and "Share".

At the very bottom of the interface, there is a "More" button followed by genre tags: "soothing", "jazz", "funk", "soul", and "saxophone".

At the bottom left, a disclaimer reads: "Disclaimer: AI outputs may sometimes be offensive or inaccurate". At the bottom right, there are links for "Privacy" and "Terms of Service".

SUNO from OPENAI



Further questions

- What about music AI Copyright?
- Can a machine claim copyright if it is not a human?

Ownership & Copyright

Who owns the songs I generate using Suno?

If you are a paying subscriber to Suno, then you own the songs you generate while subscribed to [Pro](#) or [Premier](#), subject to your compliance with Suno's [Terms of Service](#).

If you are using a free version of Suno, we retain ownership of the songs you generate, but you are allowed to use those songs for non-commercial purposes, subject to your compliance with Suno's [Terms of Service](#).

Suno is best suited for making new music with new lyrics, and you must obtain permission for any and all lyrics and other content that you upload to Suno or otherwise incorporate into your songs. See Suno's [Terms of Service](#) for a more detailed discussion about the ownership and usage rules for the content generated using Suno.

If I write my own lyrics, do I still own them after submitting them to Suno?

Yes. Regardless of which version of Suno you use, you retain all ownership and rights to any original content you create and input into Suno.

Activity: Play with models

- 阿里云Paraformer语音识别:
<https://help.aliyun.com/zh/dashscope/developer-reference/quick-start-7>
- <https://jointoucan.com/bark>
- MusicLM: Generating Music From Text:
<https://aitestkitchen.withgoogle.com/tools/music-fx>
- SUNO: <https://suno.com/>



<https://ds323.ancorasir.com/>

Thank you~

Wan Fang
Southern University of Science and Technology