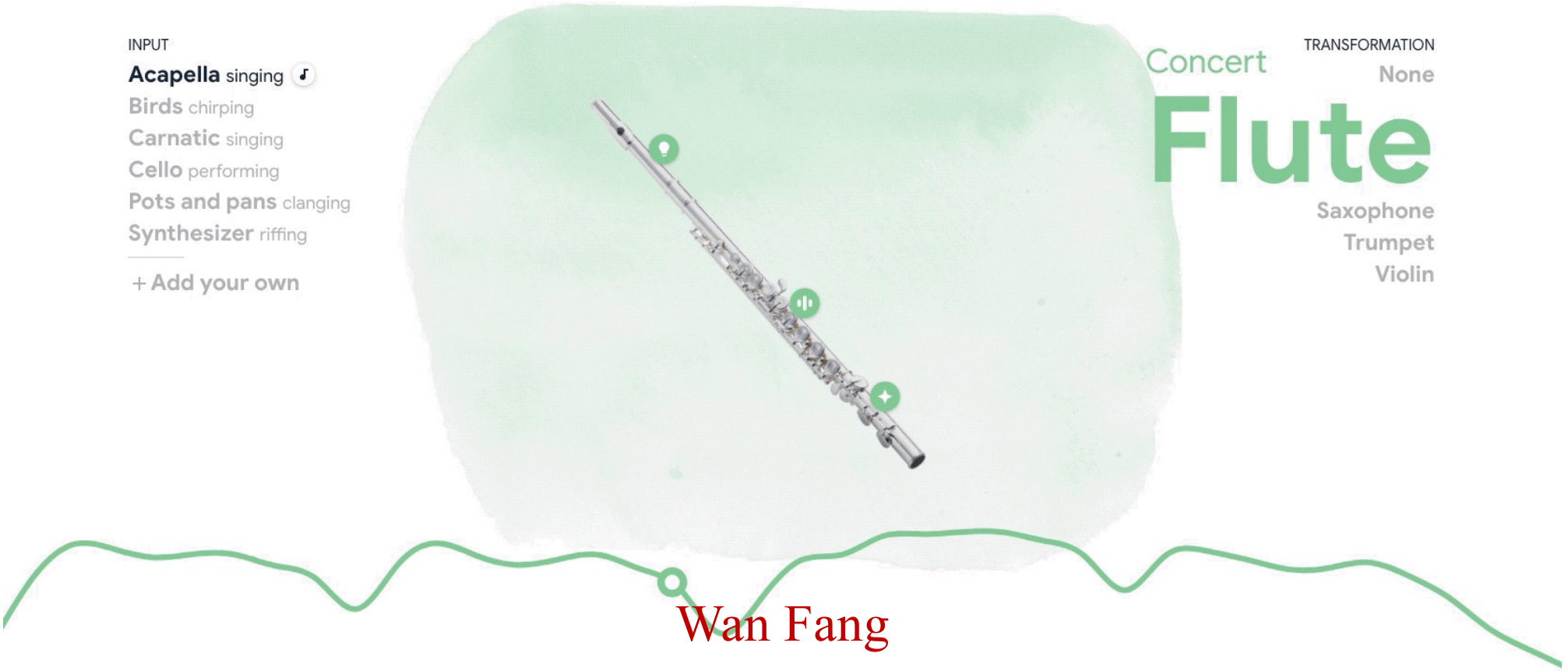
AIID + Sound



Wan Fang

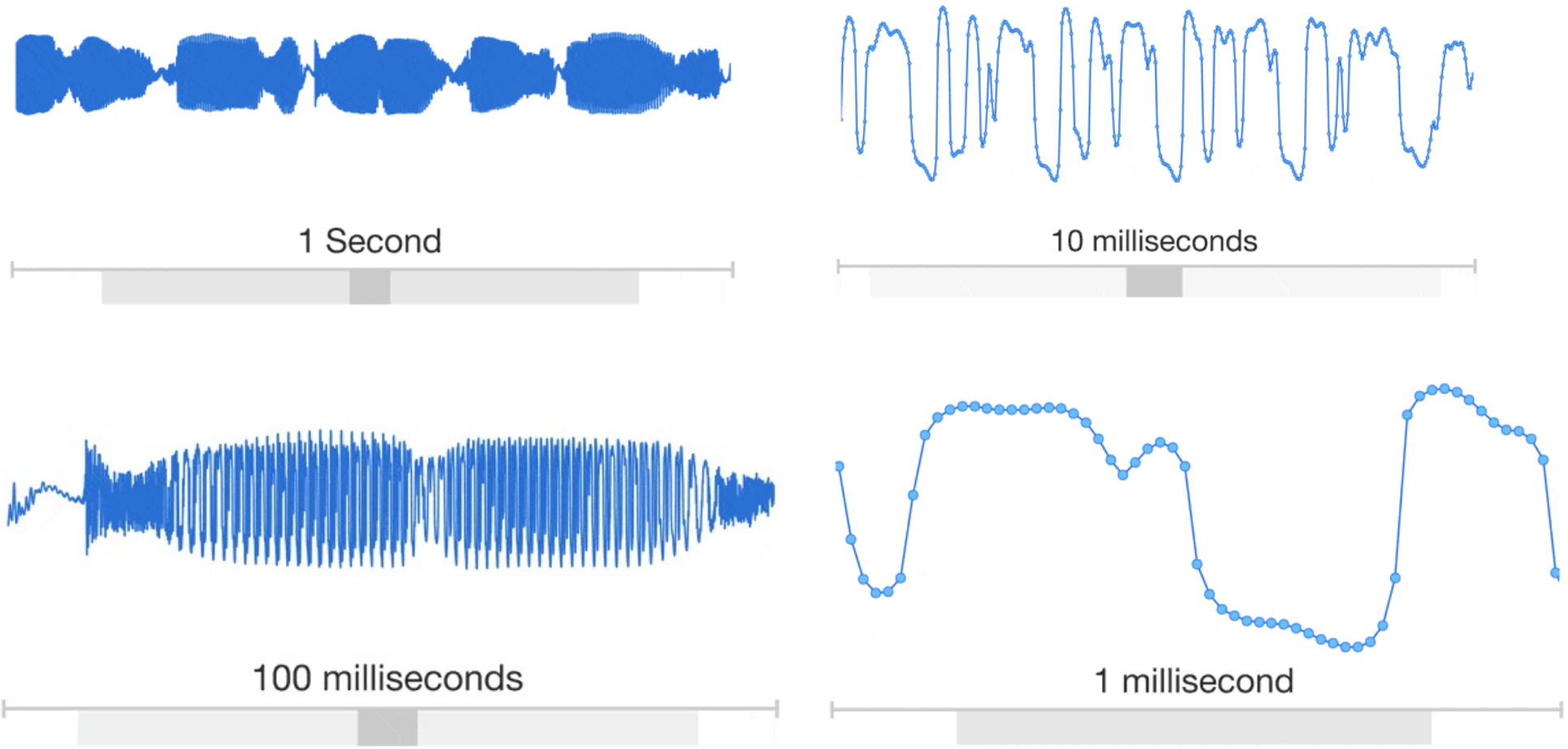
Southern University of Science and Technology

Source: https://magenta.tensorflow.org/tone-transfer

Agenda

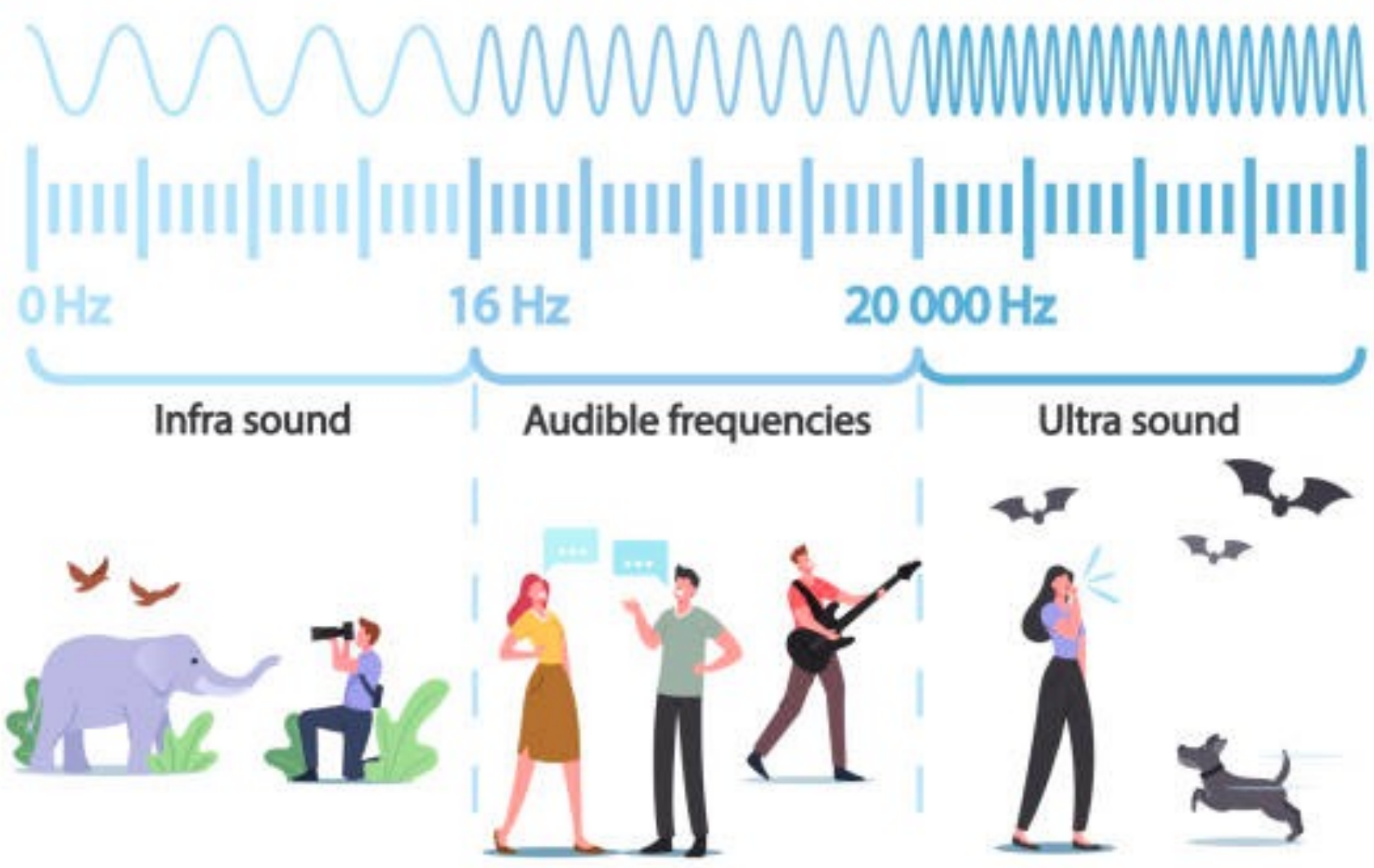
- Sound as Data
- Automatic Speech recognition (ASR)
- Voice Recognition
- Music Generation
  - Symbolic AI vs Audio AI systems
  - Tools to Make Your Own Generative Music
  - Concluding Thoughts and Further Questions

AIID + Sound

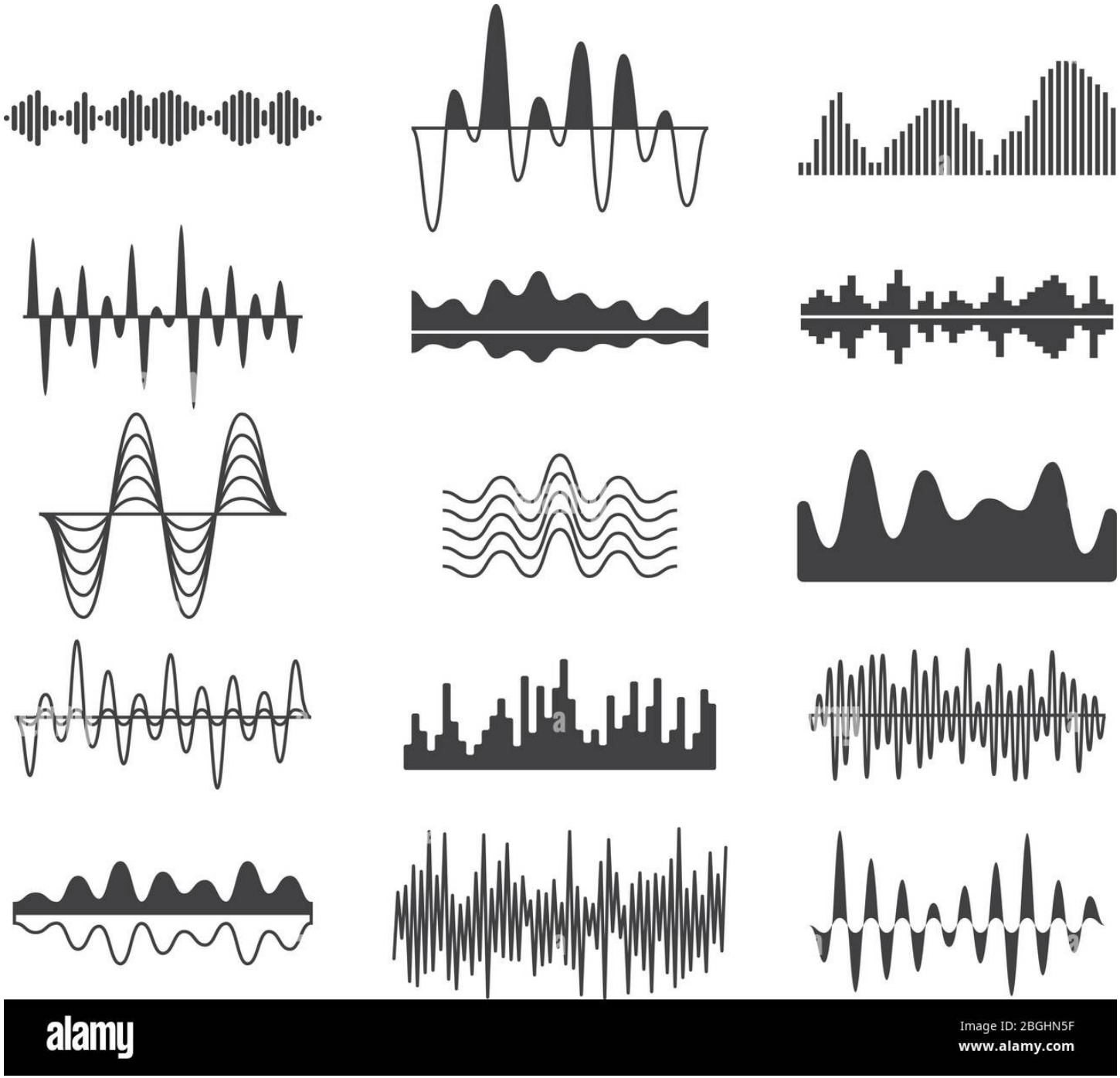


Audio samples from WaveNet paper (2018)

AIID + Sound


0 Hz
16 Hz
20 000 Hz
Infra sound
Audible frequencies
Ultra sound

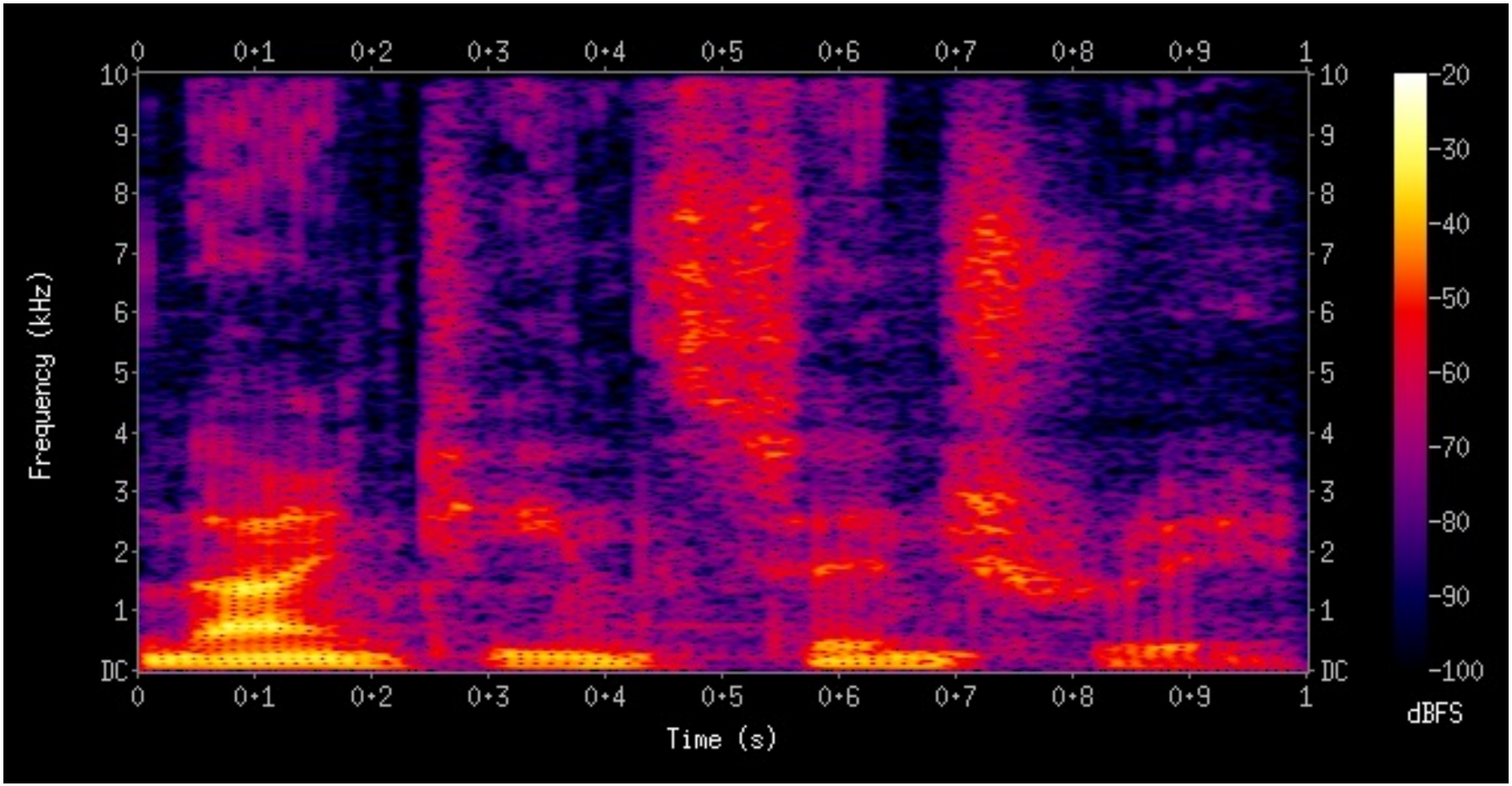
AIID + Sound



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

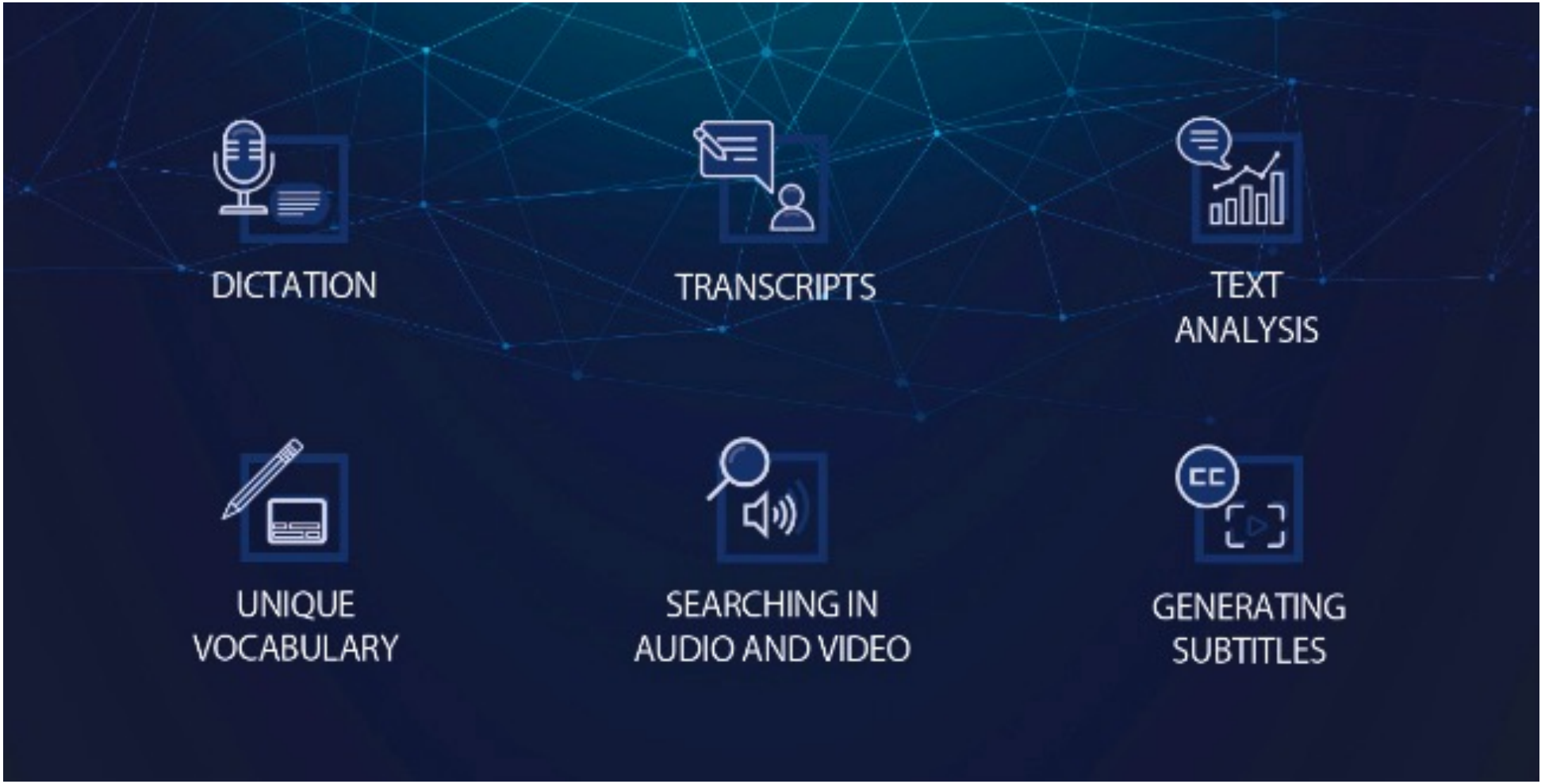
A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time.

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。

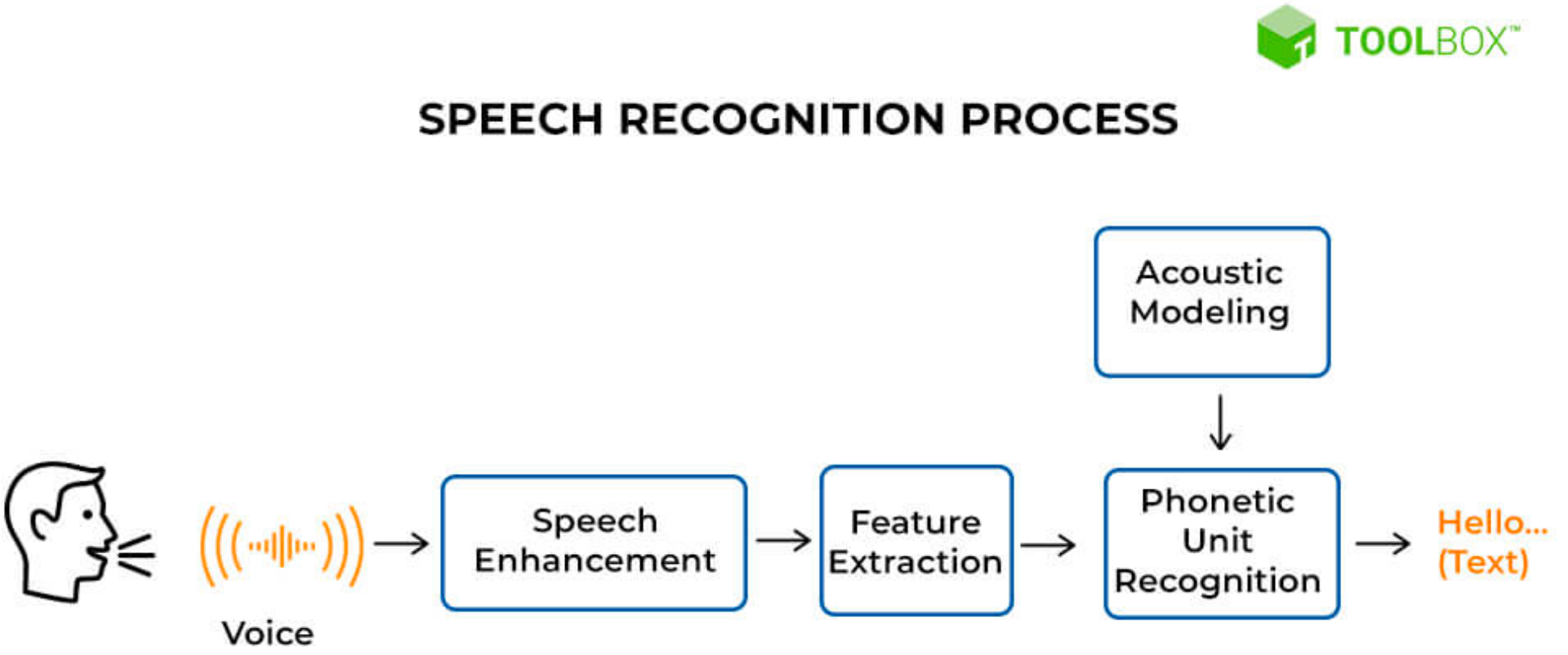
Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR)

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

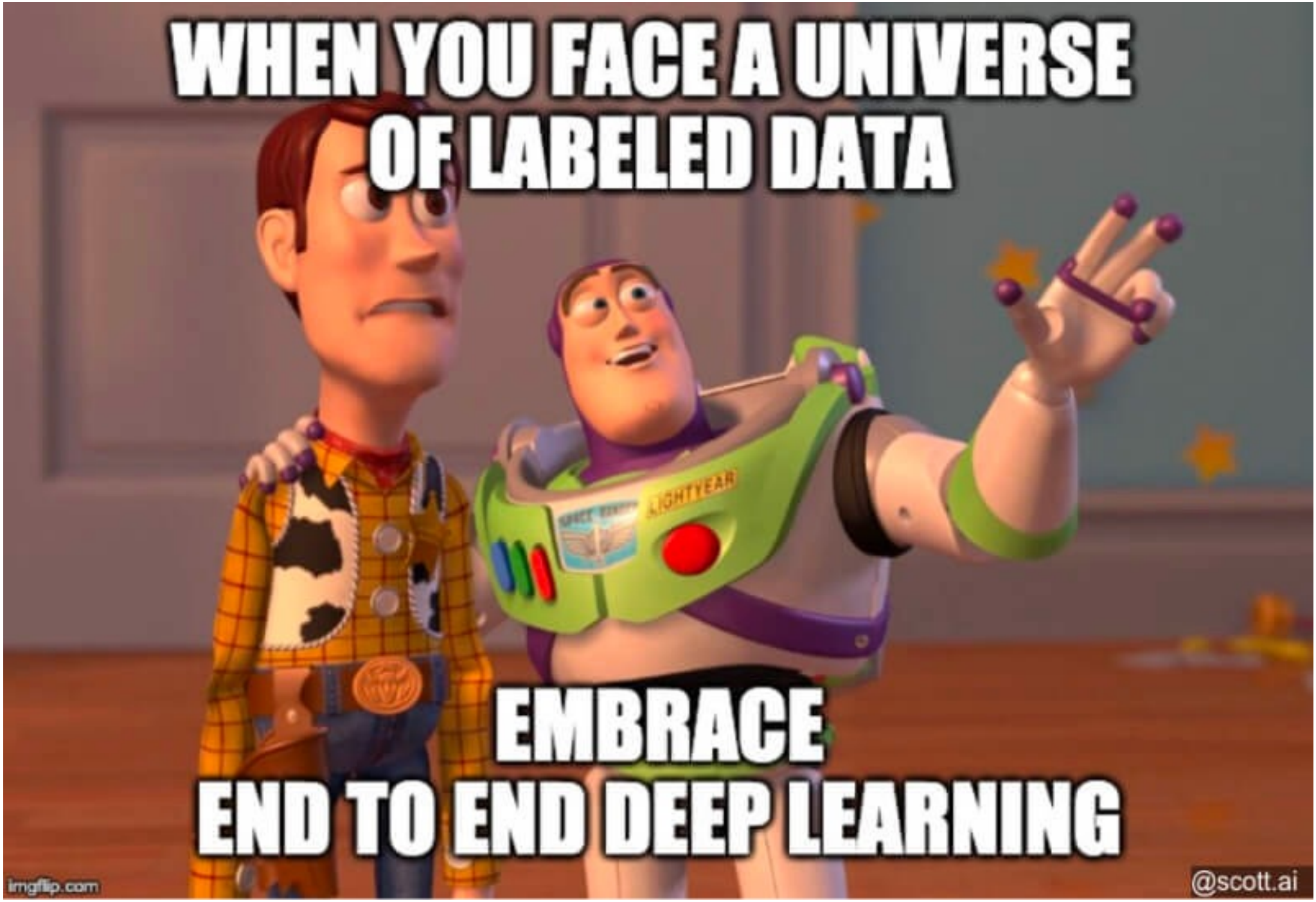


Used to be like



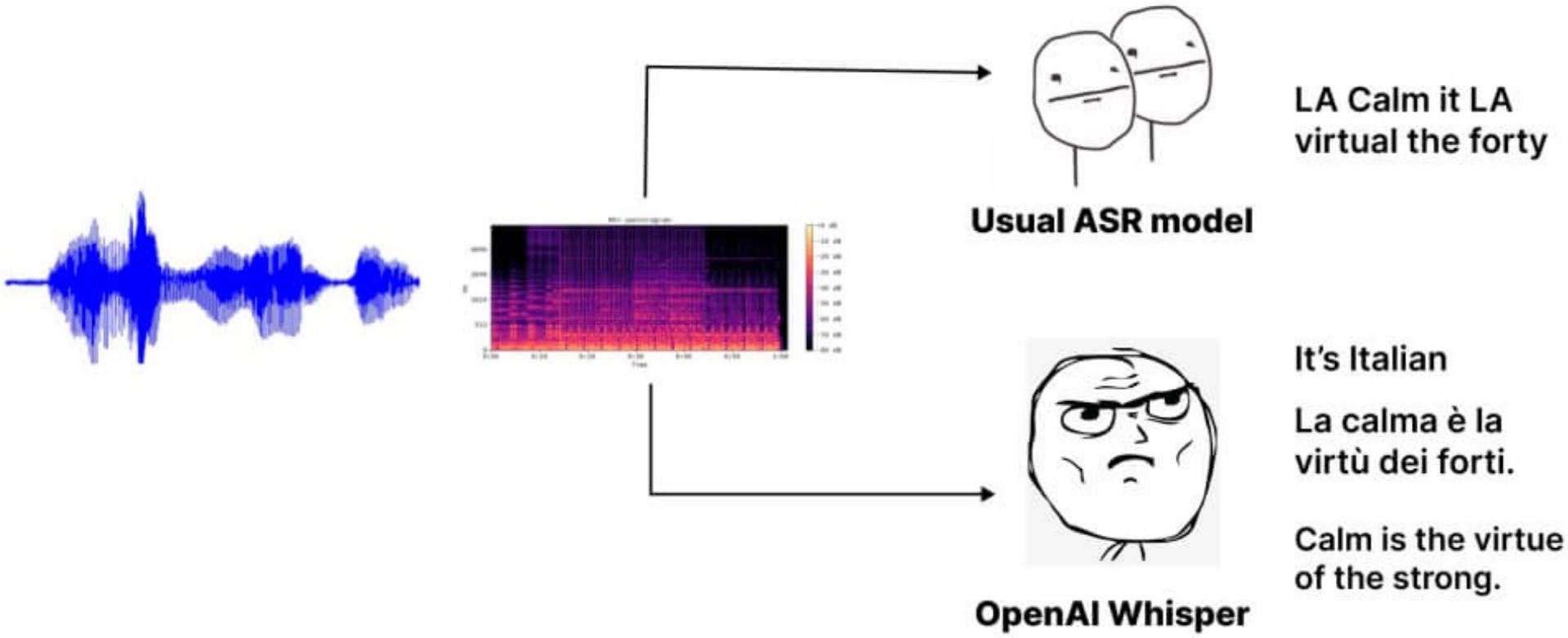
https://www.superannotate.com/blog/openai-whisper-automatic-speech-recognition-system

AIID + Sound


WHEN YOU FACE A UNIVERSE
OF LABELED DATA
EMBRACE
END TO END DEEP LEARNING
imgflip.com
@scott.ai

Whisper from OpenAI

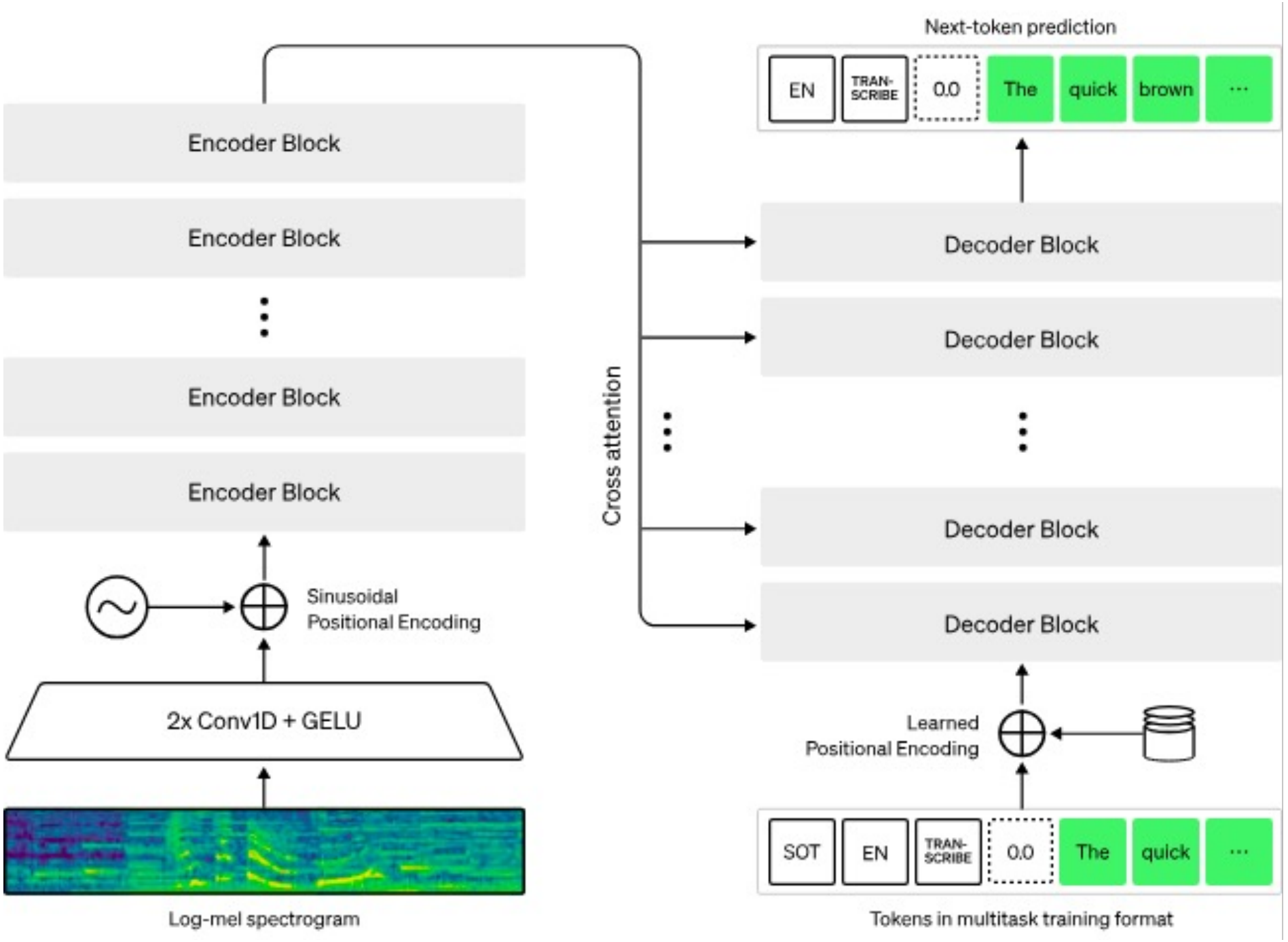
- In 2022, this idea of training on **large data** to achieve cross-domain performance arrived in the world of speech recognition with OpenAI's launch of Whisper.



Whisper from OpenAI

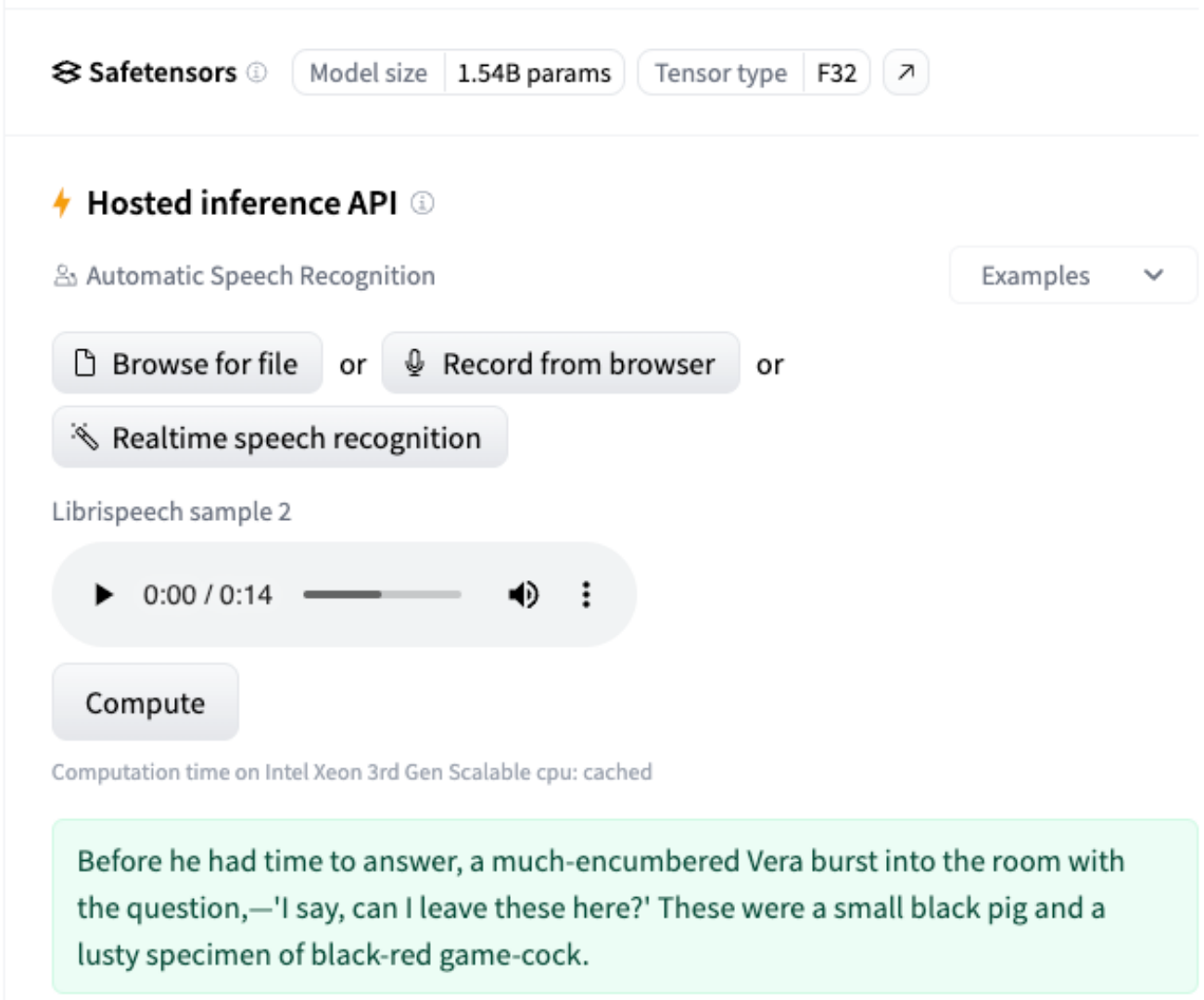
- 680,000 hours of multilingual and multitask supervised data
- A third of Whisper's audio dataset is non-English.

AIID + Sound


Next-token prediction
EN
TRAN-SCRIBE
0.0
The
quick
brown
...
Encoder Block
Encoder Block
Encoder Block
Encoder Block
Cross attention
Decoder Block
Decoder Block
Decoder Block
Decoder Block
Sinusoidal Positional Encoding
2x Conv1D + GELU
Learned Positional Encoding
SOT
EN
TRAN-SCRIBE
0.0
The
quick
...
Log-mel spectrogram
Tokens in multitask training format

Whisper from OpenAI

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

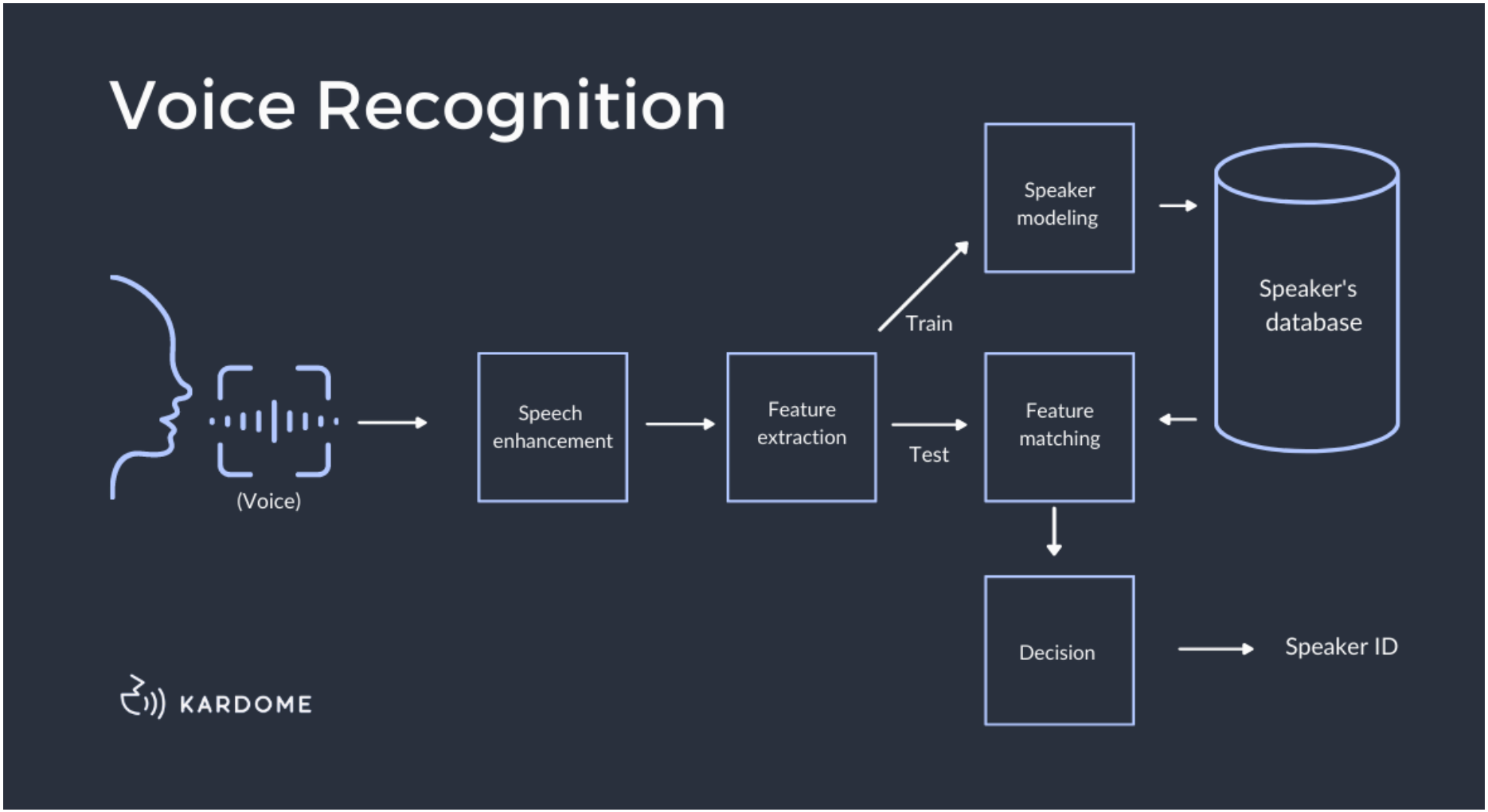


Voice Recognition

Pay attention to the difference from speech recognition

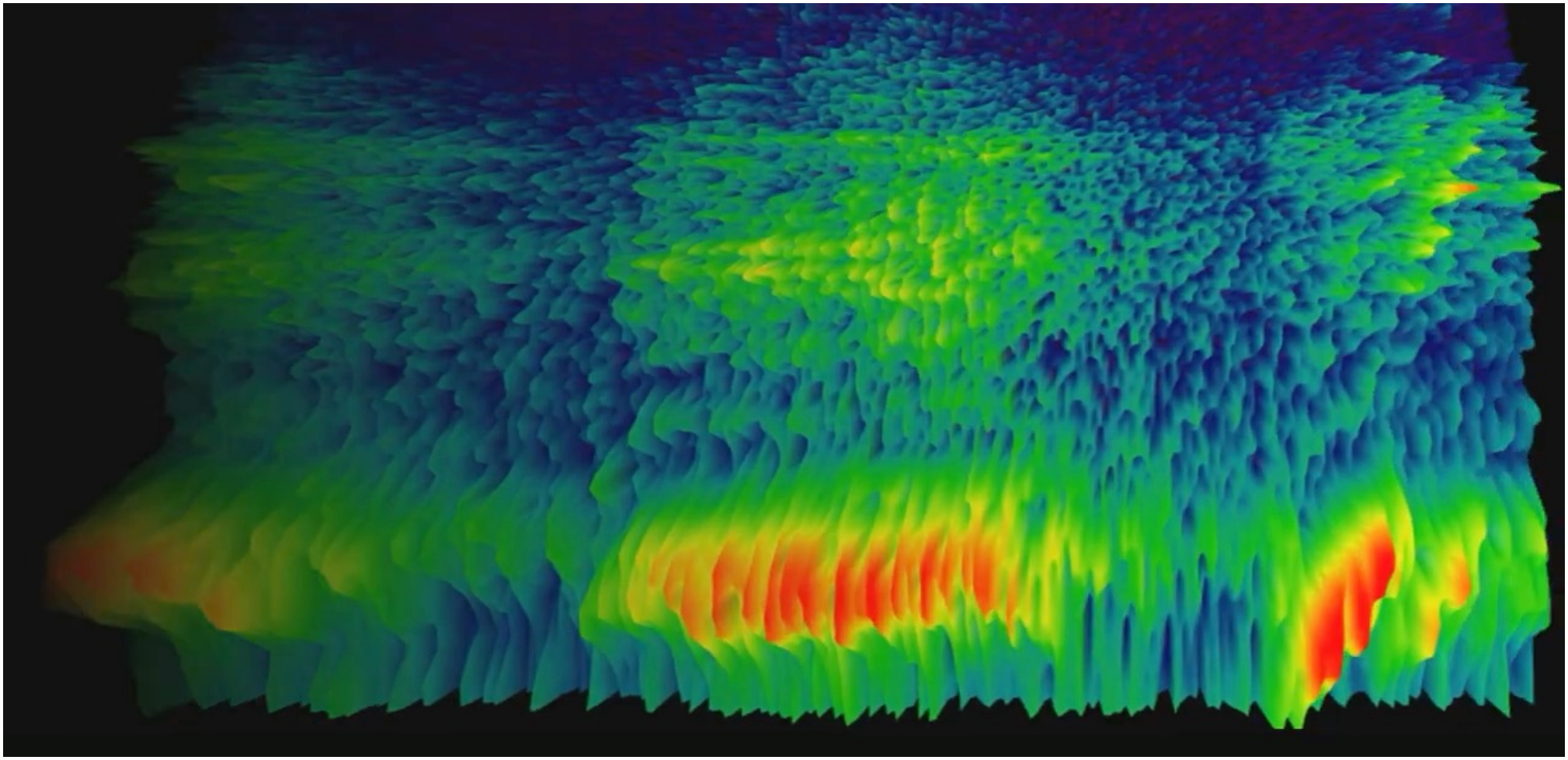
Voice Recognition

- 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.

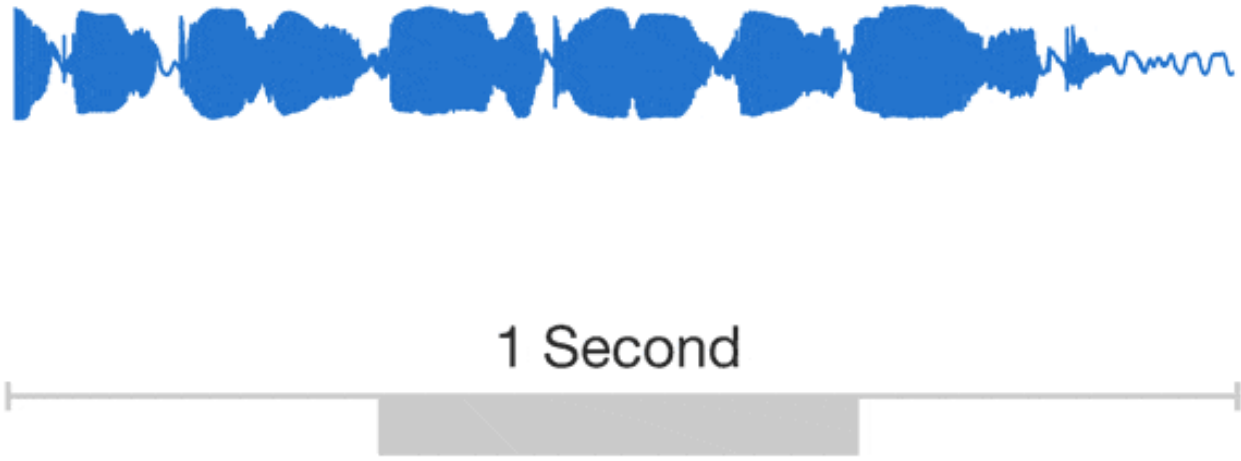
Computation time on Intel Xeon 3rd Gen Scalable cpu: 12.294 s

```
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 |]
```

text-to-music: https://huggingface.co/sander-wood/text-to-music

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.*

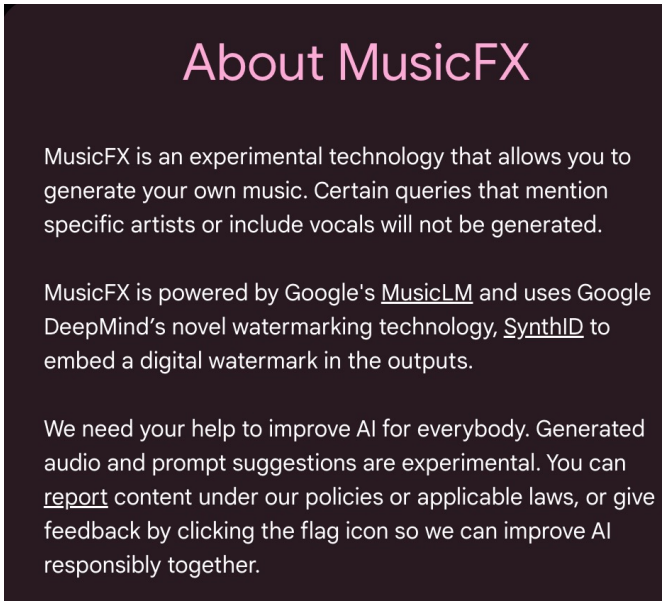


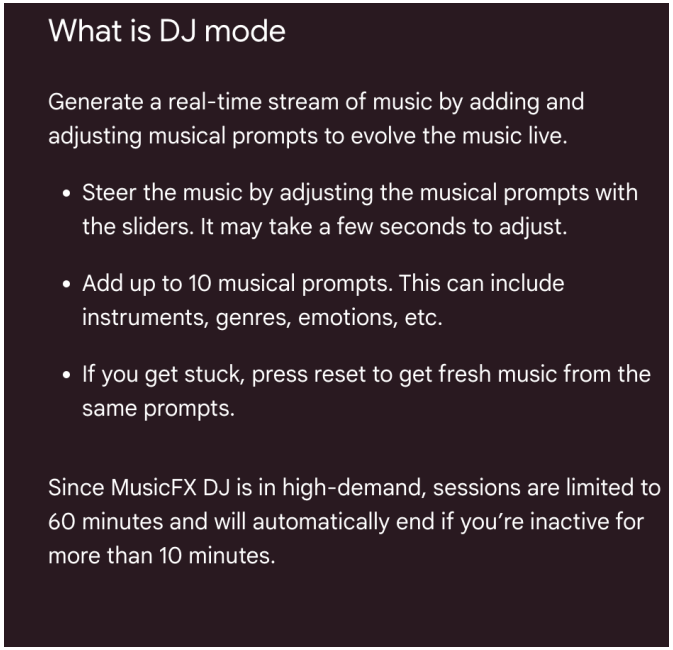
Tone Transfer

Source

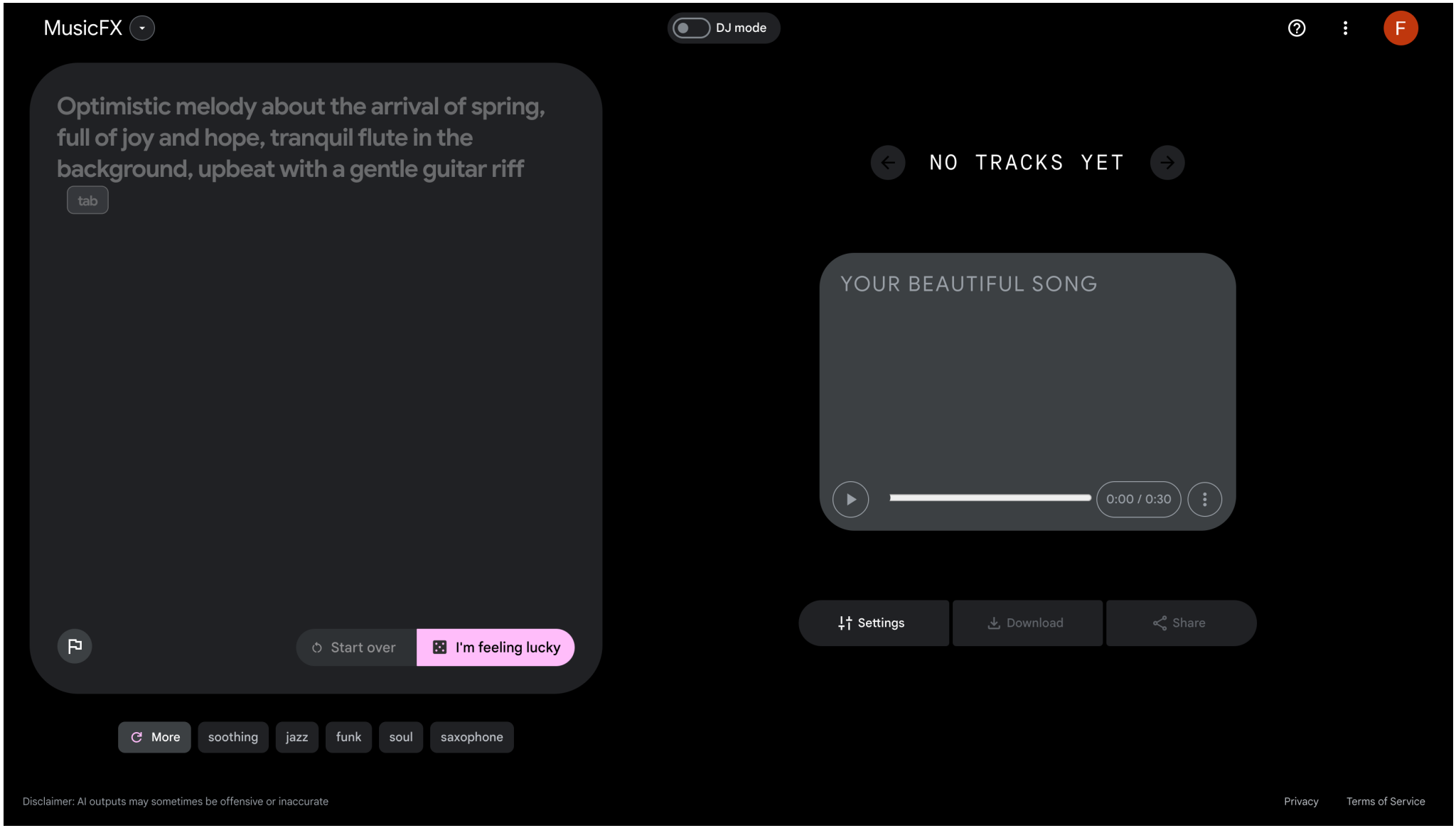
MusicLM: Generating Music From Text

- 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.

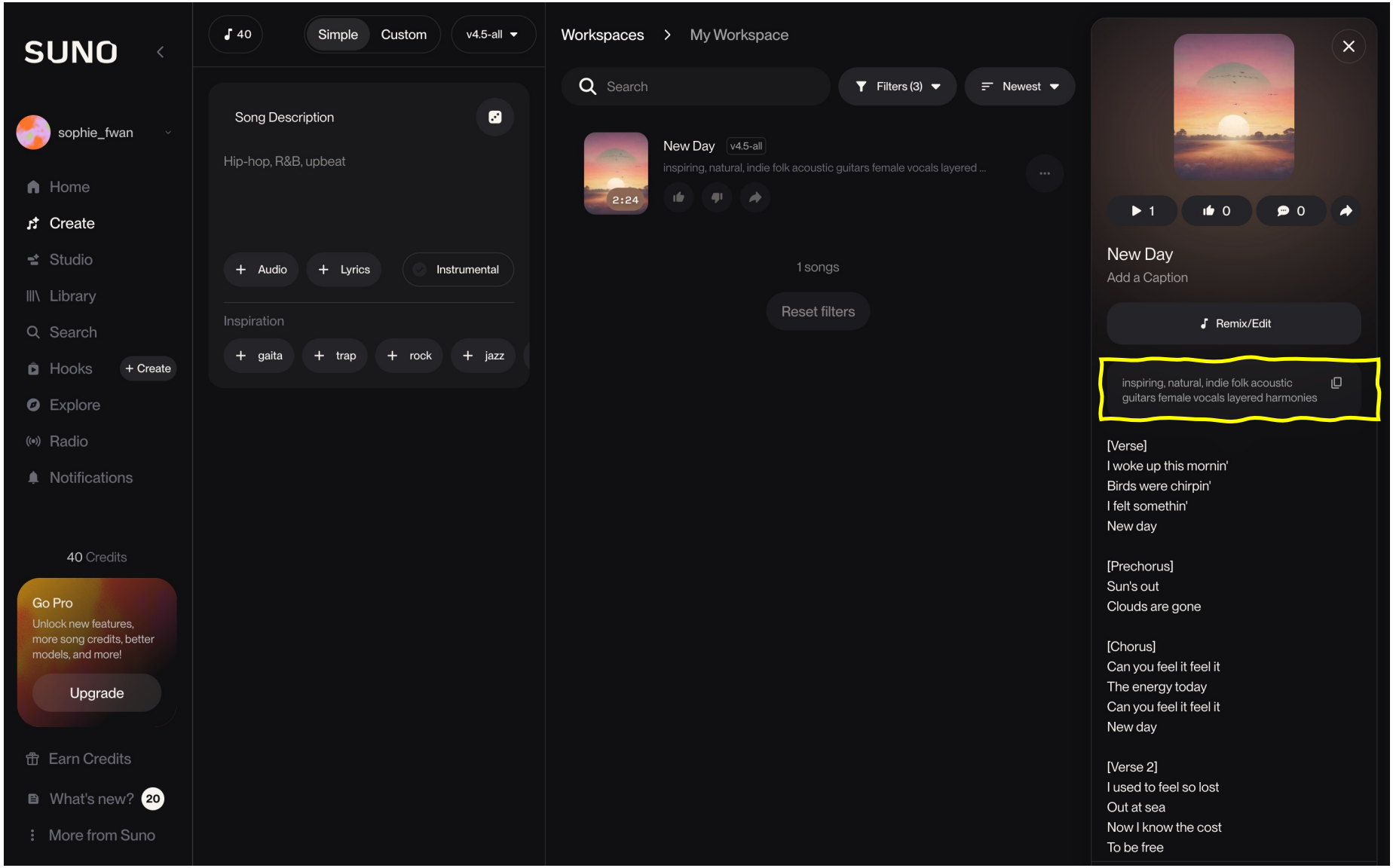




MusicFX: Generating Music From Text



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.

Resource

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