



香港中文大學(深圳)  
The Chinese University of Hong Kong, Shenzhen



SCHOOL OF  
DATA SCIENCE  
數據科學學院

# Introduction of Sound, Speech, and Singing Voice

**Xueyao Zhang**

**2024/04**

# About me



**Xueyao Zhang (张雪遥)**

- ◆ **Second-year PhD student**, Supervised by Prof Zhizheng Wu  
School of Data Science, CUHK-Shenzhen  
Homepage: <https://www.zhangxueyao.com/>
- ◆ **Amphion v0.1's co-founder**  
Project: <https://github.com/open-mmlab/Amphion> **(3.9k stars)**
- ◆ **Research interest: “AI + Music”**, especially on:
  - Singing Voice Processing
  - Music Generation

Spaces | amphion / singing\_voice\_conversion | like 144 | Running on A10G


### Amphion Singing Voice Conversion: *DiffWaveNetSVC*

This demo provides an Amphion [DiffWaveNetSVC](#) pretrained model for you to play. The training data has been detailed [here](#).

#### Source Audio

Hint: We recommend using dry vocals (e.g., studio recordings or source-separated voices from music) as the input for this demo. At the bottom of this page, we provide some examples for your reference.

Source Audio

 **SVC Online Demo:** [https://huggingface.co/spaces/amphion/singing\\_voice\\_conversion](https://huggingface.co/spaces/amphion/singing_voice_conversion)

Target Singer

☐ Adele ☐ Beyonce ☐ Bruno Mars ☐ John Mayer

☐ Michael Jackson ☐ Taylor Swift ☐ Jacky Cheung 张学友

☒ Jian Li 李健 ☐ Feng Wang 汪峰 ☐ Faye Wong 王菲

☐ Yijie Shi 石倚洁 ☐ Tsai Chin 蔡琴 ☐ Ying Na 那英

☐ Eason Chan 陈奕迅 ☐ David Tao 陶喆

Pitch Shift Control

If you want to control the specific pitch shift value, you need to choose "Key Shift"

☒ Auto Shift ☐ Key Shift

Key Shift Values

How many semitones you want to transpose. This parameter will work only if you choose "Key Shift"

0

Diffusion Inference Steps

As the step number increases, the synthesis quality will be better while the inference speed will be lower

1000

Clear

Submit

Let your favorite singer  
sing your favorite song!



香港中文大學(深圳)  
The Chinese University of Hong Kong, Shenzhen

# How to create a sound using Python?

- CSC3160: Fundamentals of Speech and Language Processing
  - Lecture 2: Colab notebook

正弦波是一种周期性波动的数学模型，其公式通常写作：

$$y(t) = A \sin(2\pi ft + \phi)$$

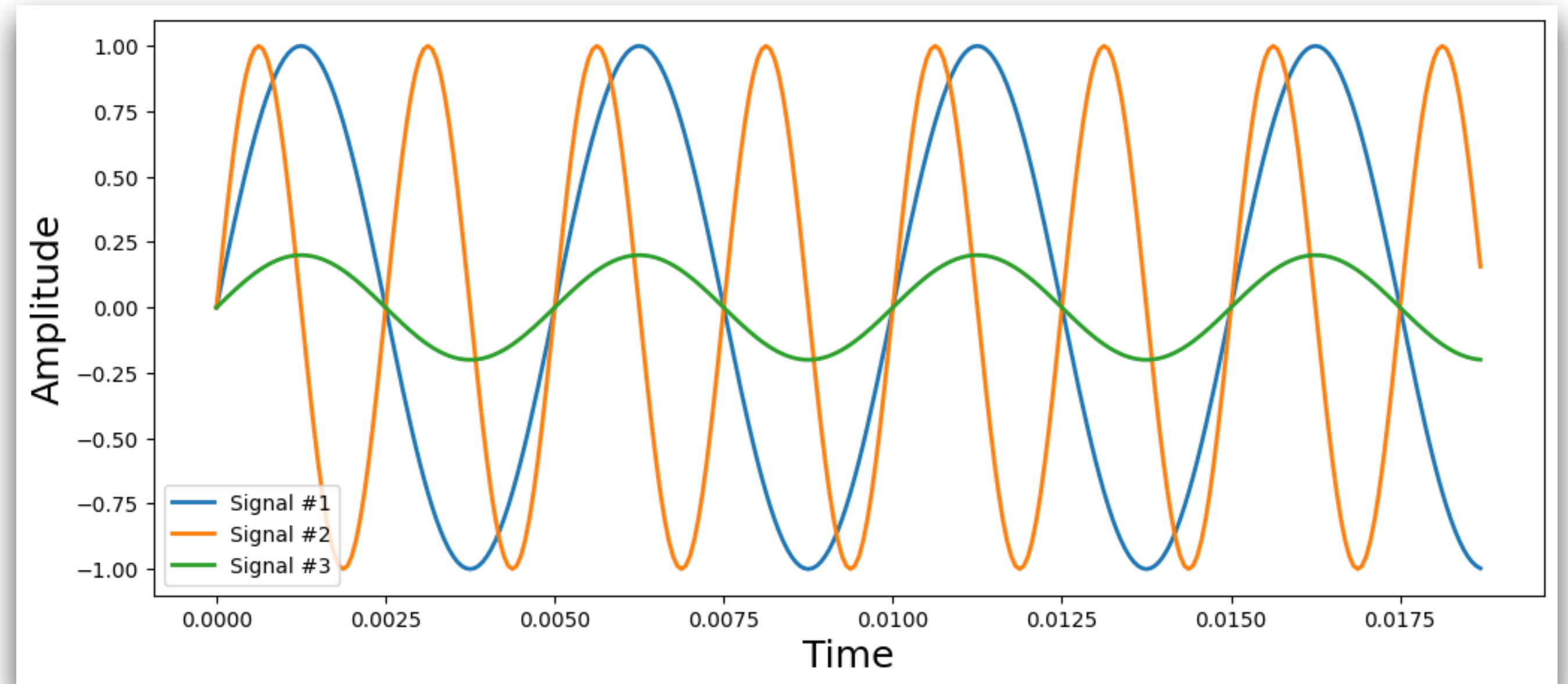
其中：

- $y(t)$  表示时间  $t$  时刻的波形值。
- $A$  是波形的振幅，即波形的最大值和最小值之差的一半。
- $f$  是频率，表示单位时间内波形重复的次数，单位是赫兹（Hz）。
- $t$  是时间，通常以秒为单位。
- $\phi$  是相位，表示波形在时间轴上的偏移量，单位是度或弧度。

正弦波在物理学中非常重要，它可以用来描述许多自然现象，如声波、电磁波、机械振动等。

```
# Time points
time = np.arange(beginTime, endTime, samplingInterval);

# Create three sine waves
signal1 = np.sin(2*np.pi*signal1Frequency*time)
signal2 = np.sin(2*np.pi*signal1Frequency*2*time)
signal3 = 0.2*np.sin(2*np.pi*signal1Frequency*time)
```



# Three elements of sound: Pitch, Loudness, and Timbre

---

## Perceptual Property

Pitch

Loudness

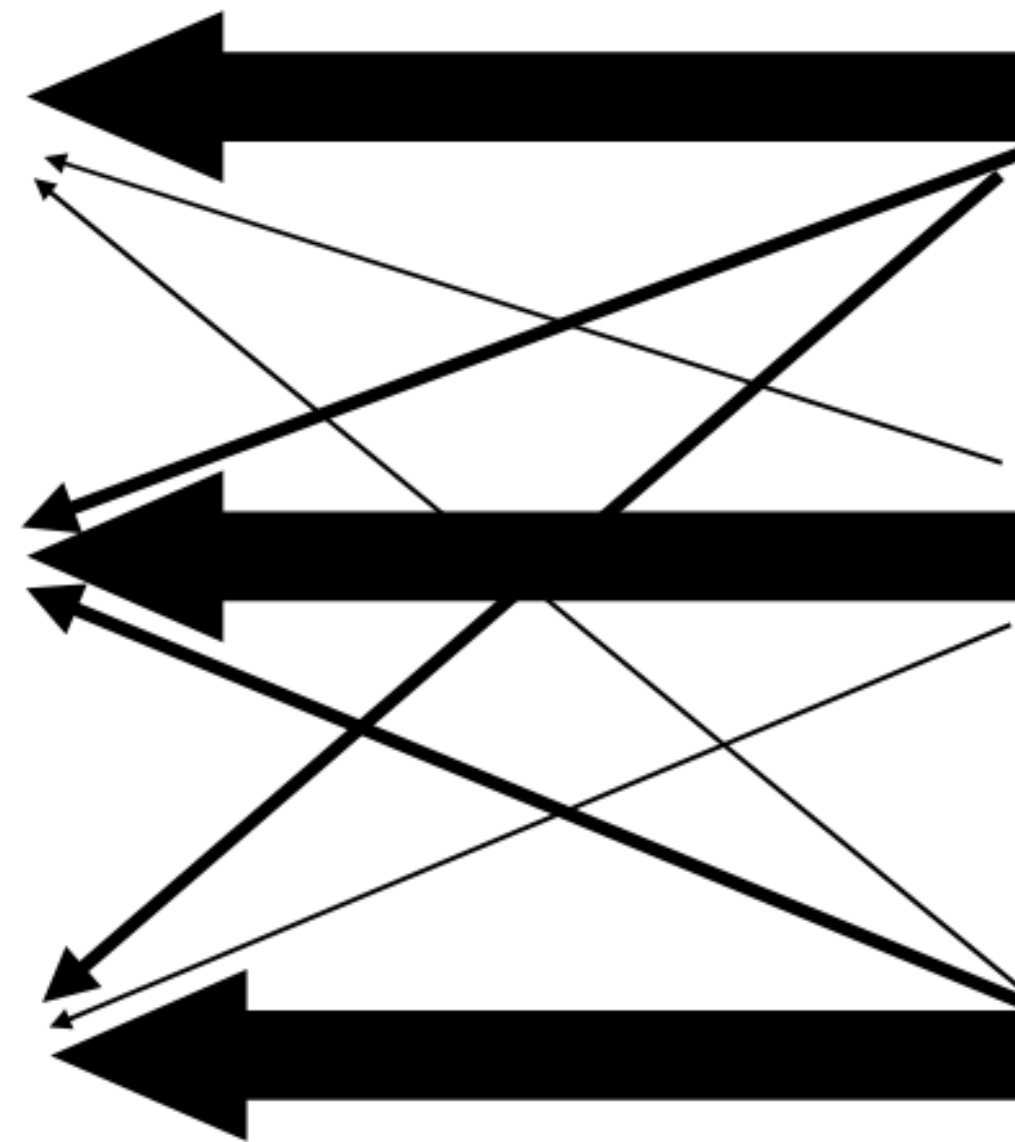
Timbre

## Physical Property

Frequency

Intensity

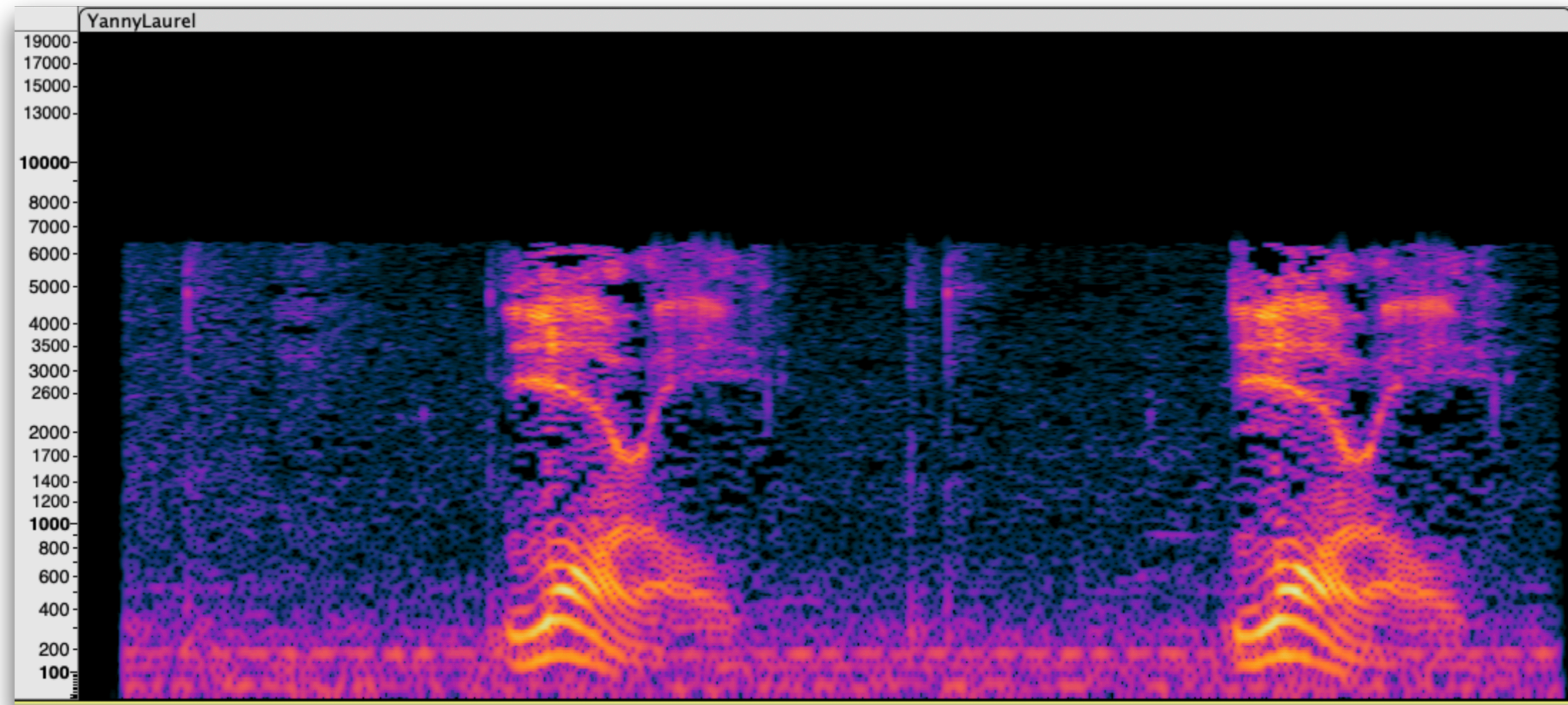
Time variation  
&  
Spectral content





# Perceptual Property: “Yanny” or “Laurel”?

---



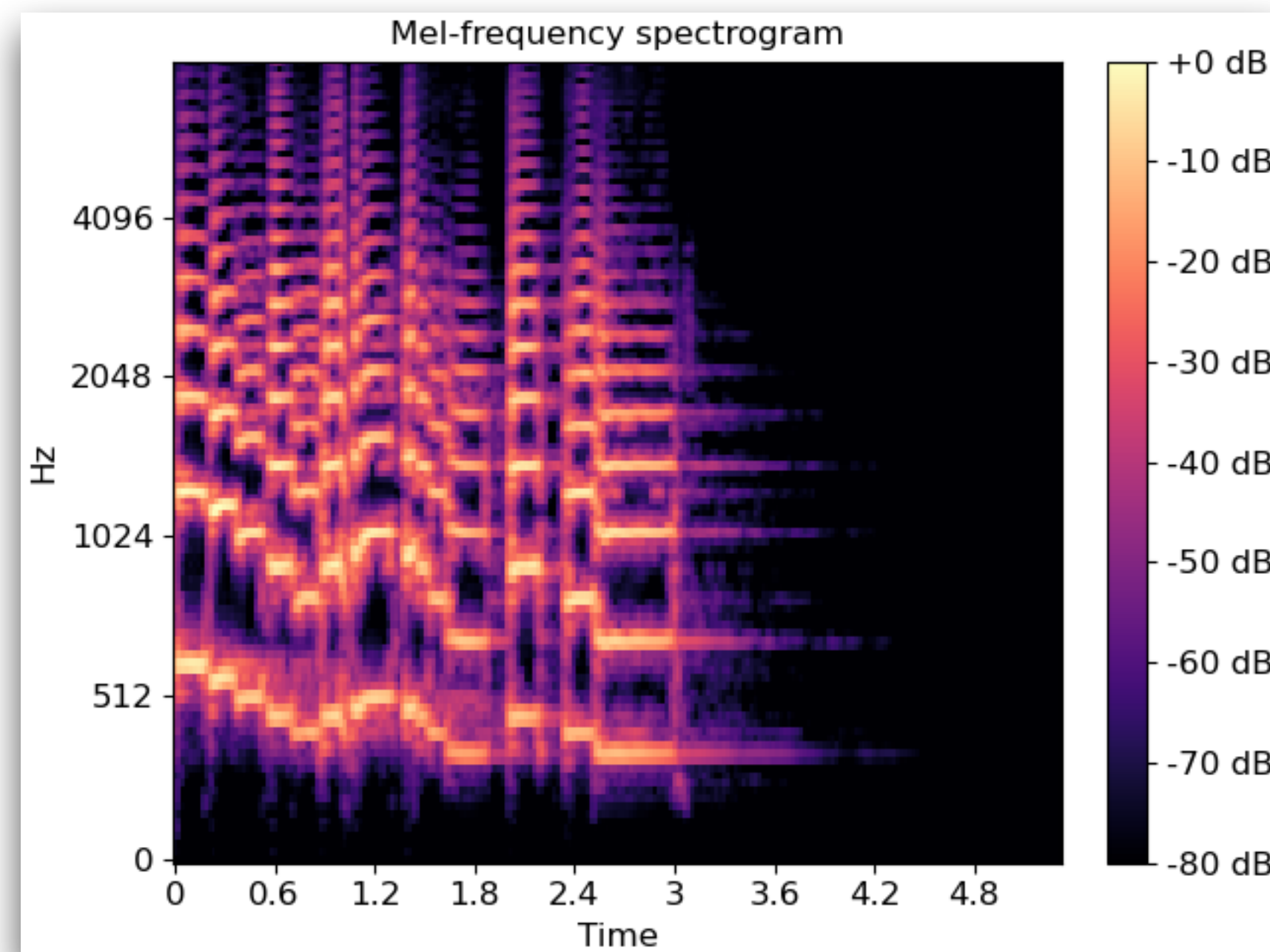
“Yanny” or “Laurel”?



# Spectrogram: Visualization of Sound

---

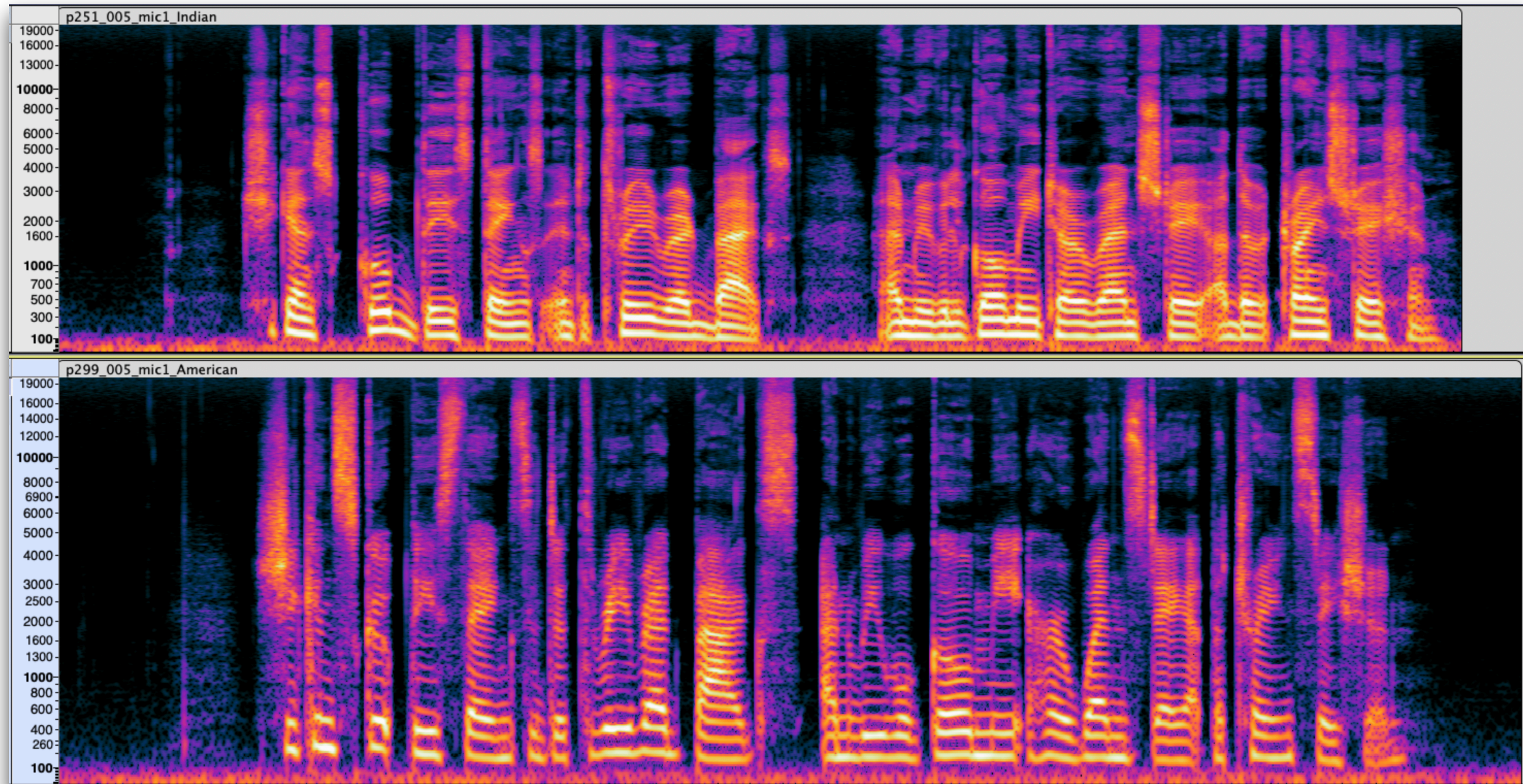
- Librosa
  - [librosa.feature.melspectrogram](#)





# Speech — Sound that owns semantic information

---

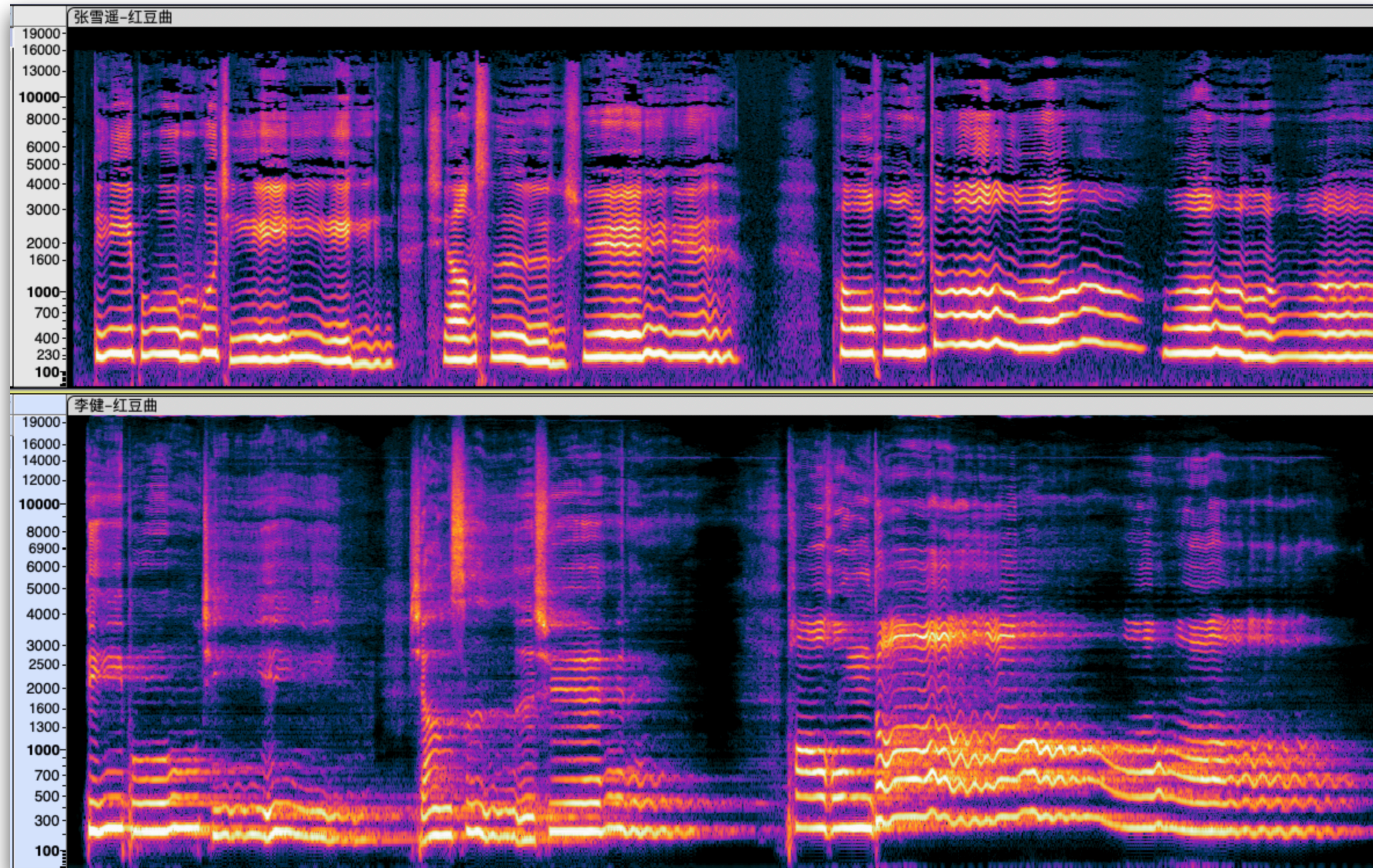


**A same utterance spoken by different people in different accents**



# Singing Voice — A More Beautiful Speech

---



**Xueyao**

**Li Jian**



# Magic of Singing Voice Conversion

	Source	Conversion Results <sup>[1]</sup>	Ground Truth
韩红 to 李健			
齐秦 to 李健			
张学友 to 李健			-
林志炫 to 李健			-
陶喆 to 李健			-

Source	Reference	Results
李健 《异乡人》	Xueyao	
	Peking Opera Performer	
Peking Opera 《苏三起解》	Xueyao	

[1] **Xueyao Zhang**, et al. Leveraging Content-based Features from Multiple Acoustic Models for Singing Voice Conversion. Machine Learning for Audio Workshop, NeuIPS 2023.



香港中文大學(深圳)  
The Chinese University of Hong Kong, Shenzhen



SCHOOL OF  
DATA SCIENCE  
數據科學學院

# THANKS



香港中文大學(深圳)  
The Chinese University of Hong Kong, Shenzhen