Visualizing Sound, Sonifying Images

Presented by
Ruoxi Du

What is image sonification ?

1) Sonification refers to the process of converting non-speech information—such as data, images, or numbers—into sound. This technique allows us to transform visual data into an auditory experience, offering a new way to perceive and understand information. Nowadays, sonification is a key topic in various fields, including scientific research, data analysis, and medical monitoring.

2) Image sonification: Image Sonification transforms visual elements like color and brightness into sound, allowing us to "hear" an image.

the most commonly used methods

- **Parameter Mapping**
- **Scanning / Time-Mapping**
- **Cross-modal / Retrieval + Synthesis / Generative Sonification**

1) Parameter Mapping

images/data

pixel brightness, color, position, density, and data values

→

music/ sound

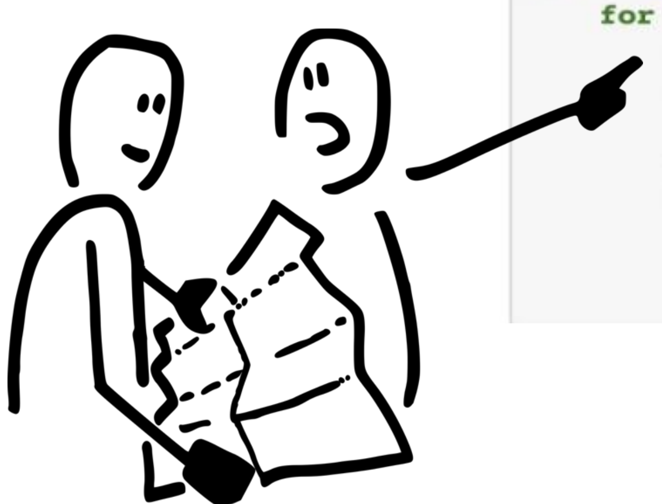
pitch, loudness, timbre, rhythm/duration/density

```python
# Extract the rgb values from an image
img = Image.open("images/bw-animal.jpg", 'r')
pix_val = list(img.getdata())
reds = []
blues = []
greens = []
for rgb in pix_val:
    reds.append(rgb[0])
    blues.append(rgb[1])
    greens.append(rgb[2])

# Image height and width
width, height = img.size
width_step = width // 60
height_step = height // 60
# Reshape RGB to arrays of the image size
reds_arr = np.asarray(reds).reshape(height, width)
blues_arr = np.asarray(blues).reshape(height, width)
greens_arr = np.asarray(greens).reshape(height, width)
reds_avg = []
blues_avg = []
greens_avg = []
# Average small chunks of the image instead of reading each individual pixel
for h in range(0, height, height_step):
    for w in range(0, width, width_step):
        red_slice = reds_arr[h:h + height_step, w:w + width_step]
        reds_avg.append(np.mean(red_slice))

        blue_slice = blues_arr[h:h + height_step, w:w + width_step]
        blues_avg.append(np.mean(red_slice))

        green_slice = greens_arr[h:h + height_step, w:w + width_step]
        greens_avg.append(np.mean(red_slice))
```



```python
s = Server().boot()

# Uncomment for version which does not play unique pitches
# pits_r = rescale(reds, 0, 255, 100, 800, False, False)
# pits_b = rescale(blues, 0, 255, 100, 800, False, False)
# pits_g = rescale(greens, 0, 255, 100, 800, False, False)

# Uncomment for version which uses each pixel
# pits_r = rescale(reds, 0, 255, 200, 550, False, False)
# pits_b = rescale(blues, 0, 255, 212, 650, False, False)
# pits_g = rescale(greens, 0, 255, 212, 700, False, False)

# Version which uses the averaged chunks of the image
# Rescale RGB values to frequency values
pits_r = rescale(reds_avg, 0, 255, 200, 550, False, False)
pits_b = rescale(blues_avg, 0, 255, 212, 650, False, False)
pits_g = rescale(greens_avg, 0, 255, 212, 700, False, False)
# Create a sequence which will iterate the above pitches
seq = Seq(time=.01).play()
it_r = Iter(seq, choice=[pits_r])

it_b = Iter(seq, choice=[pits_b])

it_g = Iter(seq, choice=[pits_g])

# Print the chosen frequency
# pp = Print(it_r, method=1, message="Frequency")

wave_r = Sine(it_r).out()
wave_g = Sine(it_g).out()
wave_b = Sine(it_b).out()

# Uncomment to add a harmonizer on top of the pitches
# hr_r = Harmonizer(wave_r).out()
# hr_g = Harmonizer(wave_g).out()
# hr_b = Harmonizer(wave_b).out()

# Runs the audio system
s.gui(locals())
```

R maps to 200–550
G maps to 212–650
B maps to 212–700

[1]Lucioni, S. (2019). *Digital Image Sonification*. GenEd 1080, December 11, 2019.

Sonifying the colors from Hilma af Klint's 1907 painting Ynglingaåldern.

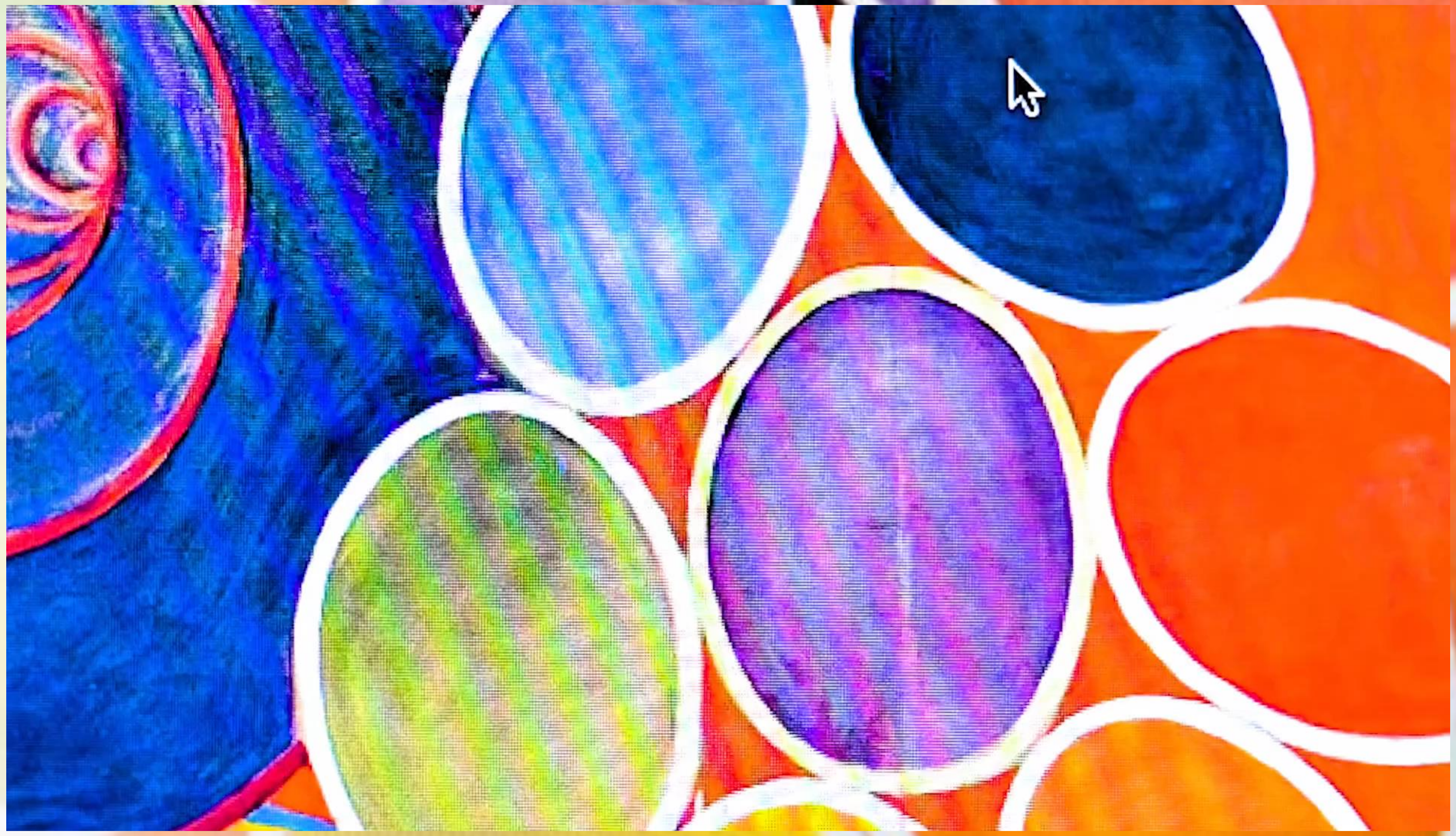
Niklas Rönnberg

Color -----> Musical Notes

· Each pixel's hue, saturation, brightness → mapped to musical notes and harmonies
· Dark or grayscale colors → lower pitches
· Bright or light colors → higher pitches

Interactive Musical Instrument

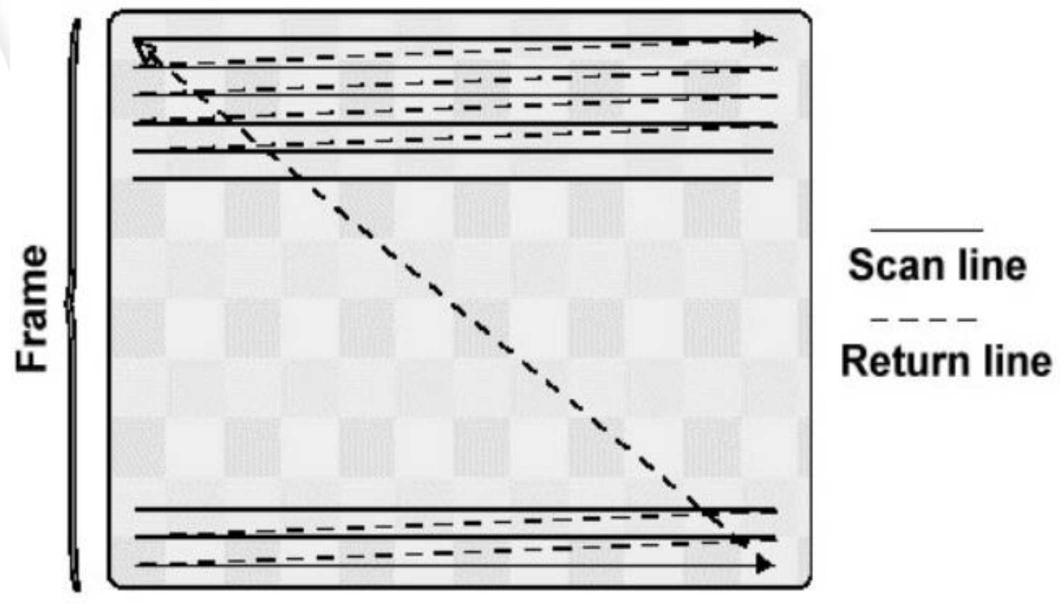
· Clicking on any area → plays a note based on that pixel's color

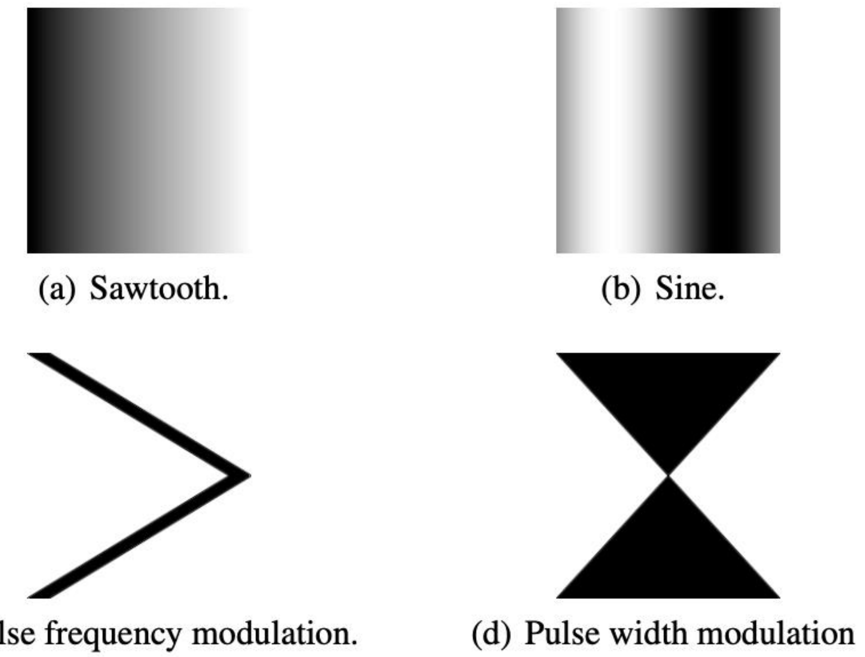


Artwork Link: https://liu.se/en/news-item/sonfiering-for-starkare-konstupplevelser

2) Scanning / Time-Mapping

Raster Scanning





- **Scan image in a fixed sequence**
  - left → right, top → bottom, row by row
  - or follow user interaction

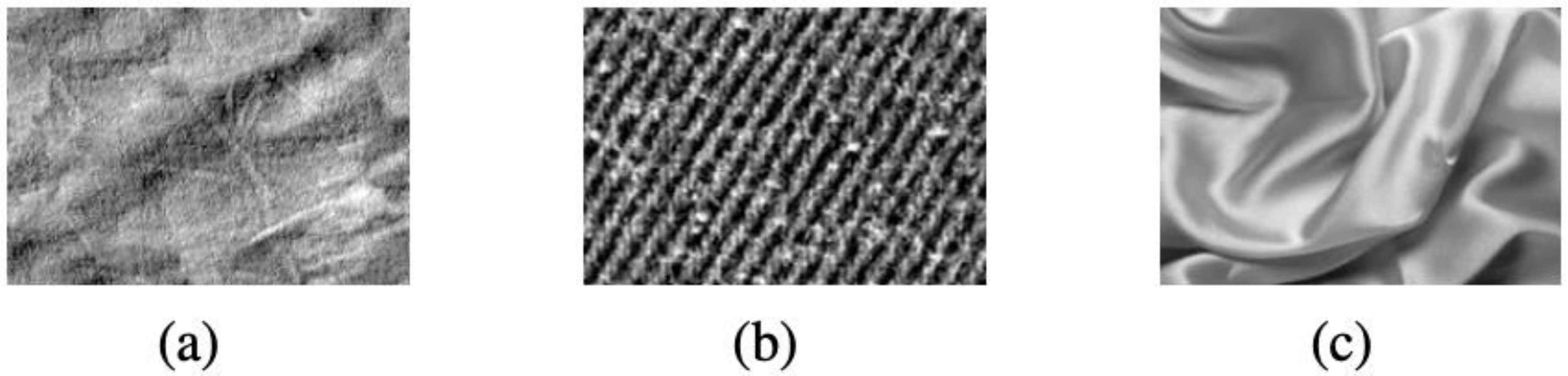


Figure 4: *Images used to compare textures by sound.*

- **Pointer moves through pixels**
  - Reads pixel info one by one
  - Turns pixel values → sound
- **2D → 1D**

[2]Yeo, W. S., & Berger, J. (n.d.). Application of Raster Scanning Method to Image Sonification, Sound Visualization, Sound Analysis and Synthesis. Center for Computer Research in Music and Acoustics, Stanford University.

Voice of Sisyphus

George Legrady, Ryan McGee, and Joshua Dickinson

•Raster Scanning

•Reads the image **row by row**
•Converts pixels into **1D audio samples**

•Audification

•Maps **grayscale values → floating-point audio**
•Direct sonification of image brightness

• Real-time Controls

•Custom software adjusts **region size & position**
•Sound changes with image area selected

• Spatialized Sound

•Sound placement corresponds to **image position**
•Multi-speaker setup for immersive experience

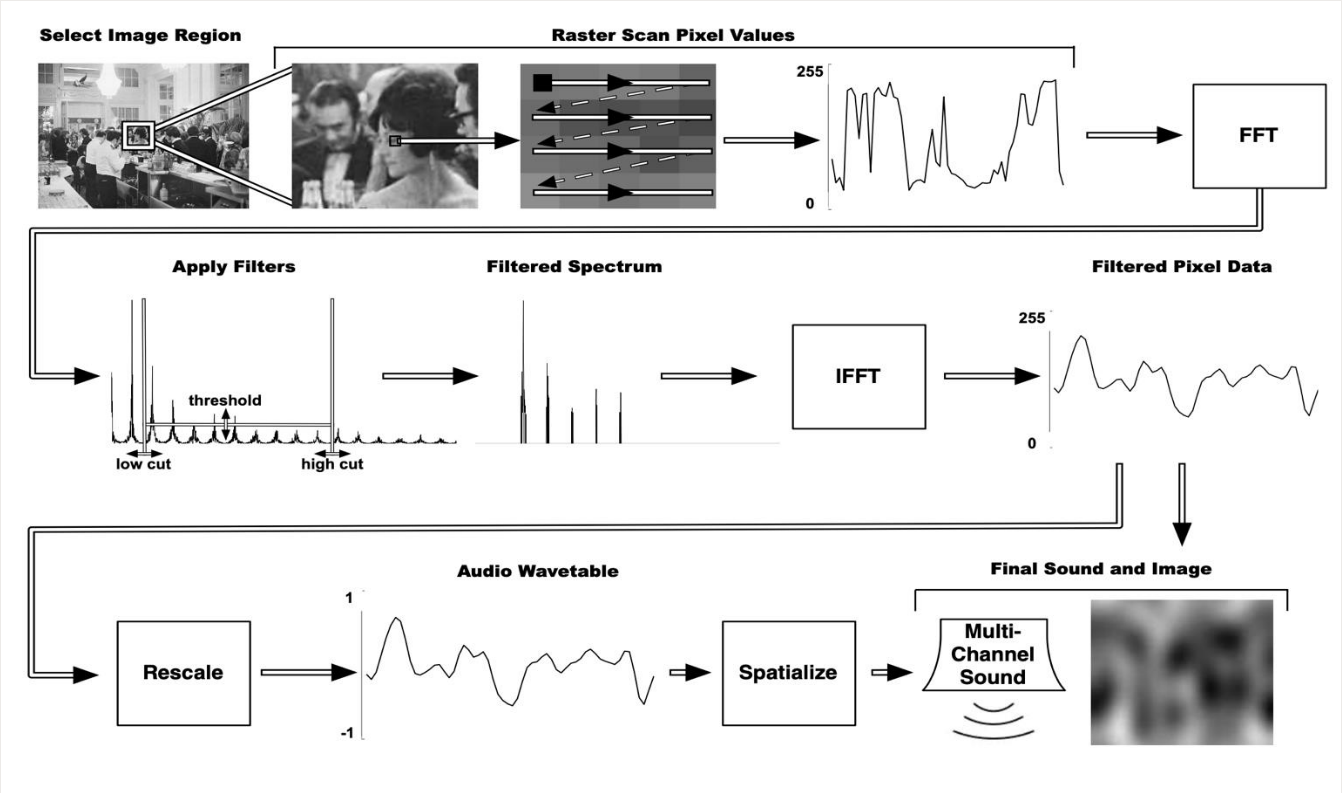
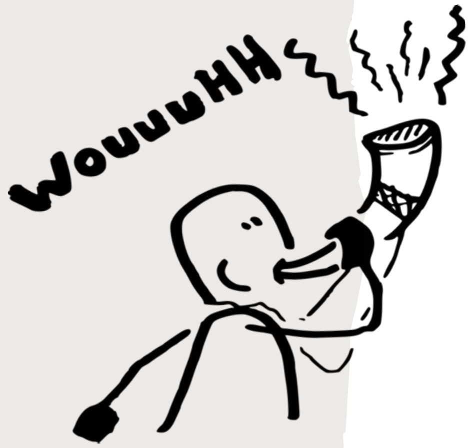


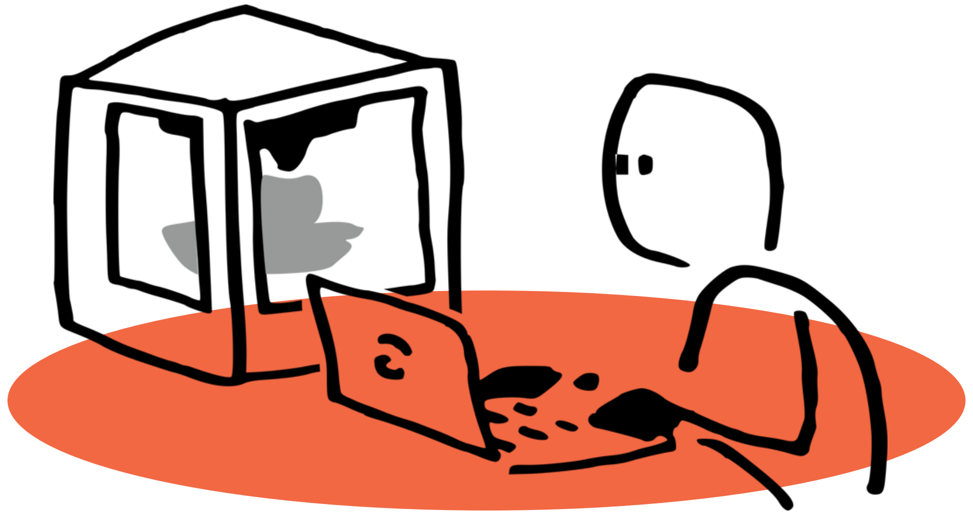
Figure 2: Sound Synthesis Algorithm for *Voice of Sisyphus*

Artwork Link: https://vimeo.com/767684234



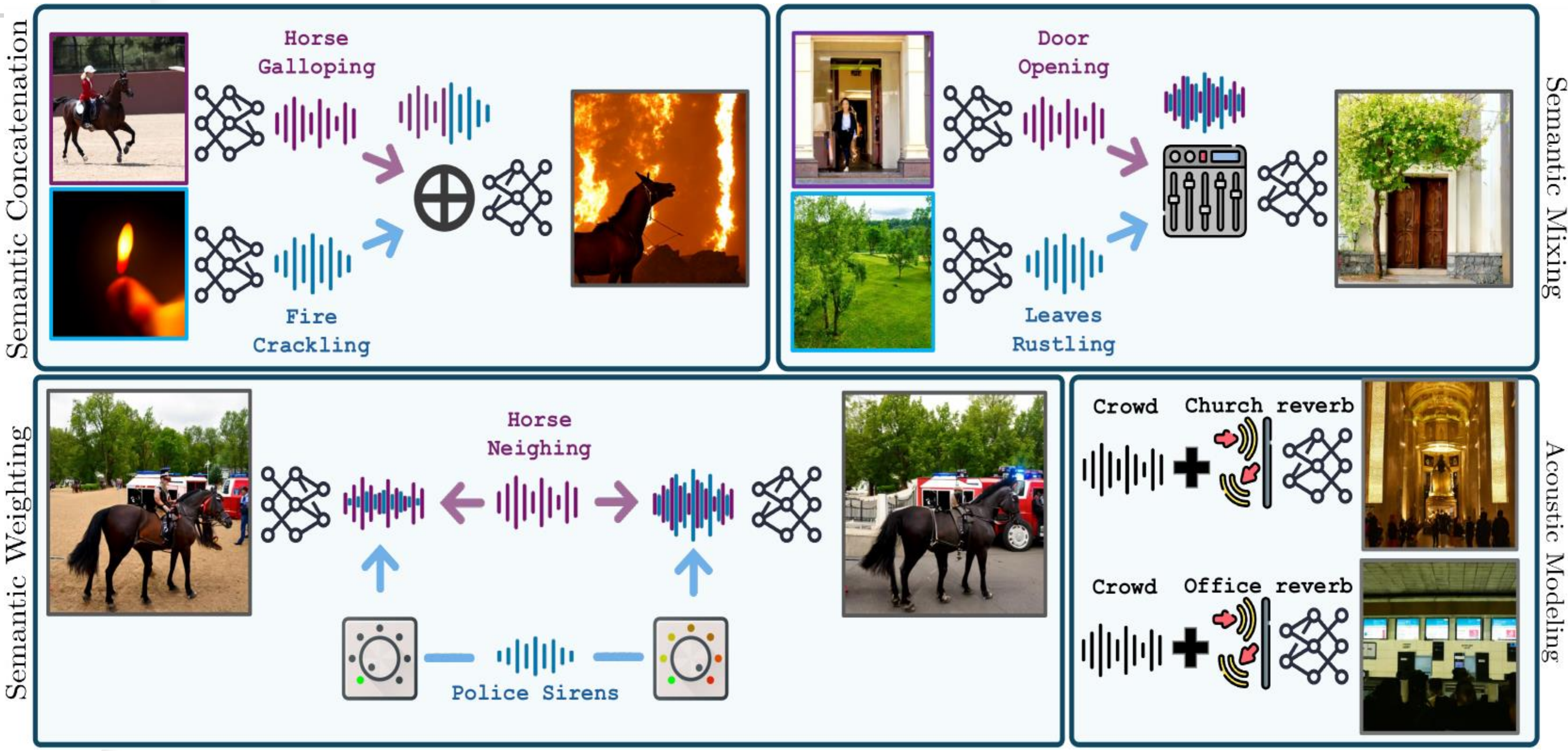
[3] McGee, R., Dickinson, J., & Legrady, G. (n.d.). Voice of Sisyphus: An Image Sonification Multimedia Installation. Experimental Visualization Lab, Media Arts and Technology, University of California, Santa Barbara.

3) Cross-modal / Retrieval + Synthesis / Generative Sonification



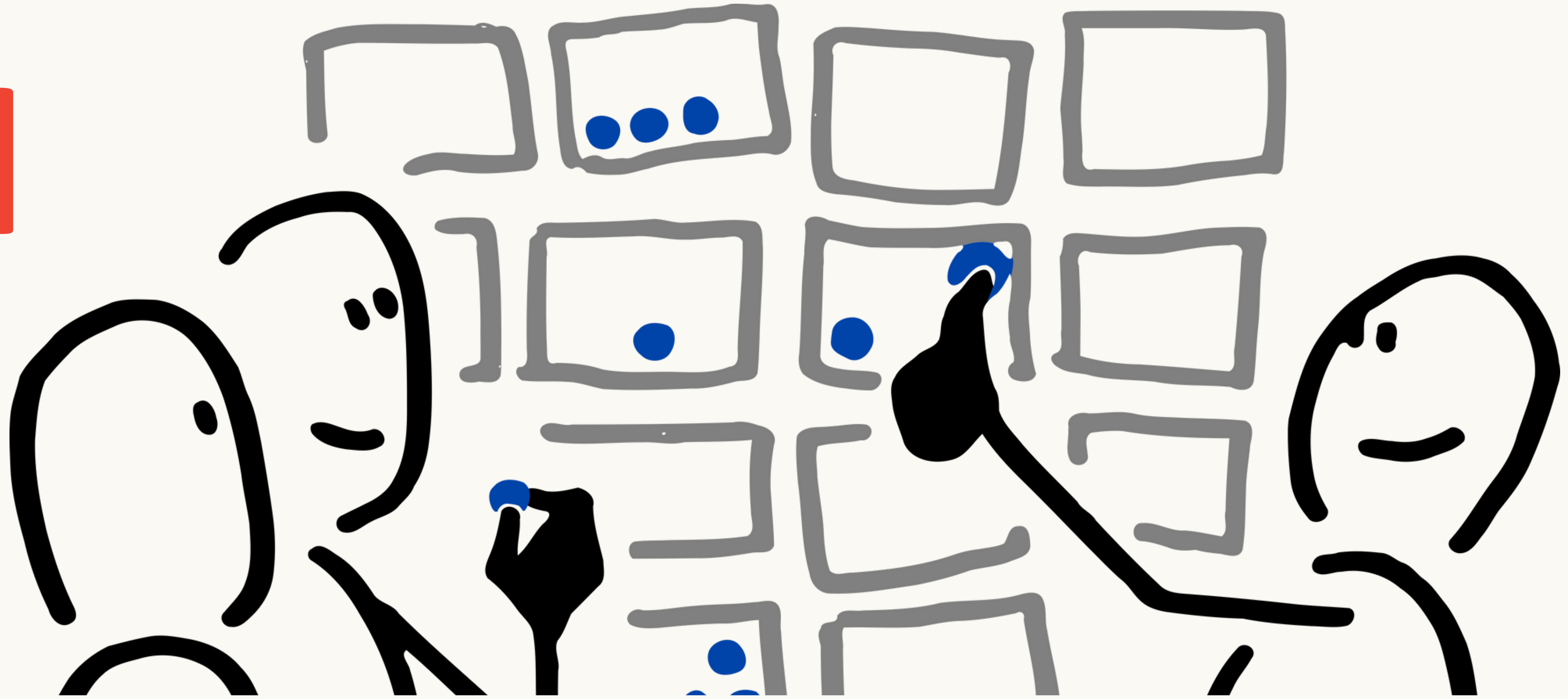
Key Idea: Use ML models to align image + sound semantically

- Uses modern generative models (image→sound, multimodal models)
- Learns semantic relationships — not just numeric mapping
- Retrieves or generates sound from large unpaired datasets
- Focus on content, meaning, style alignment
- Goes beyond simple mappings → aims for intelligent, meaningful sonification



[4] Petermann, D., & Kalayeh, M. M. (n.d.). The work conducted during author's internship at Netflix. Indiana University Bloomington, IN & Netflix Los Gatos, CA.

What is sound visualization



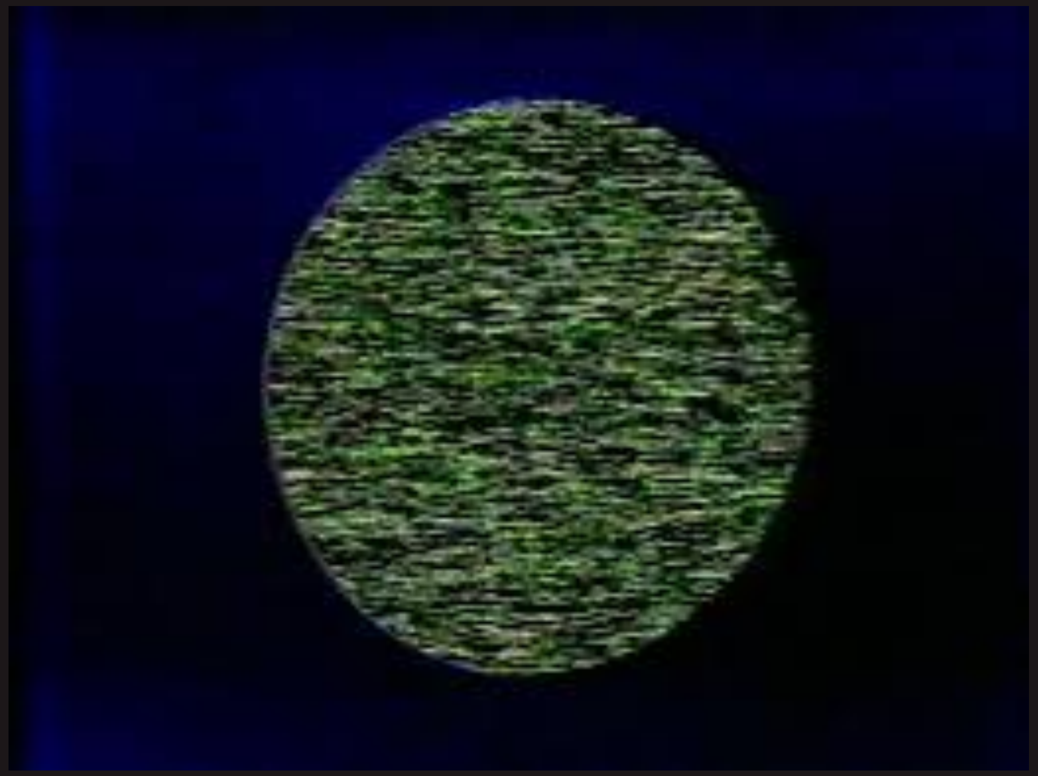
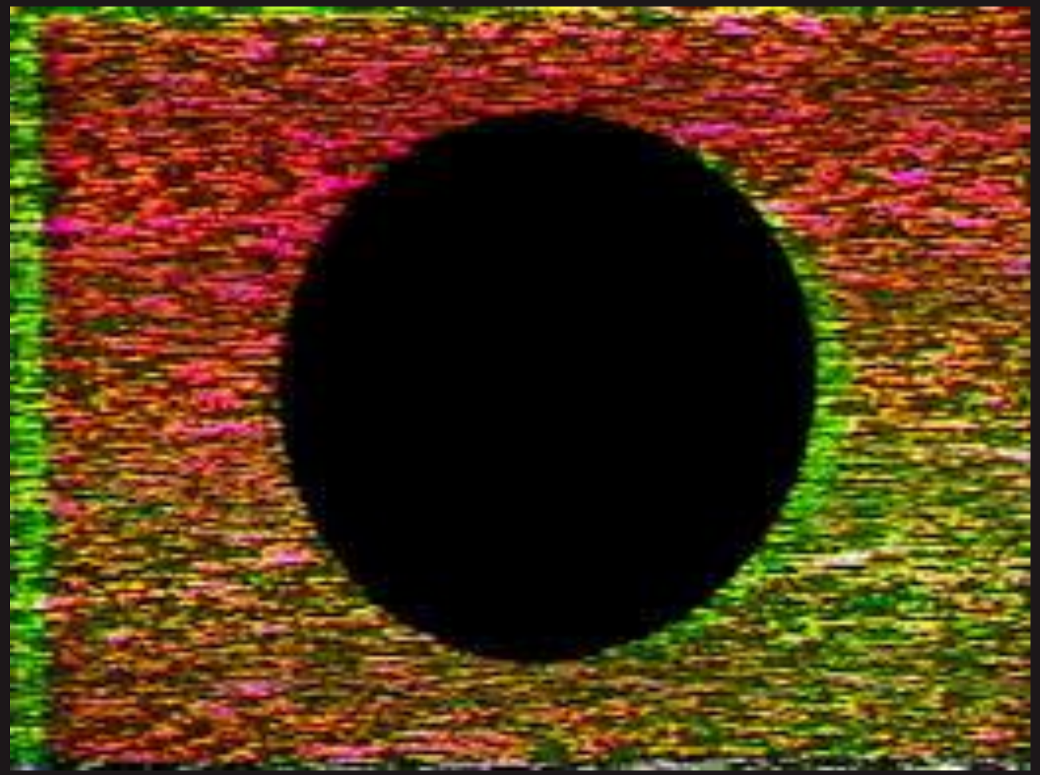
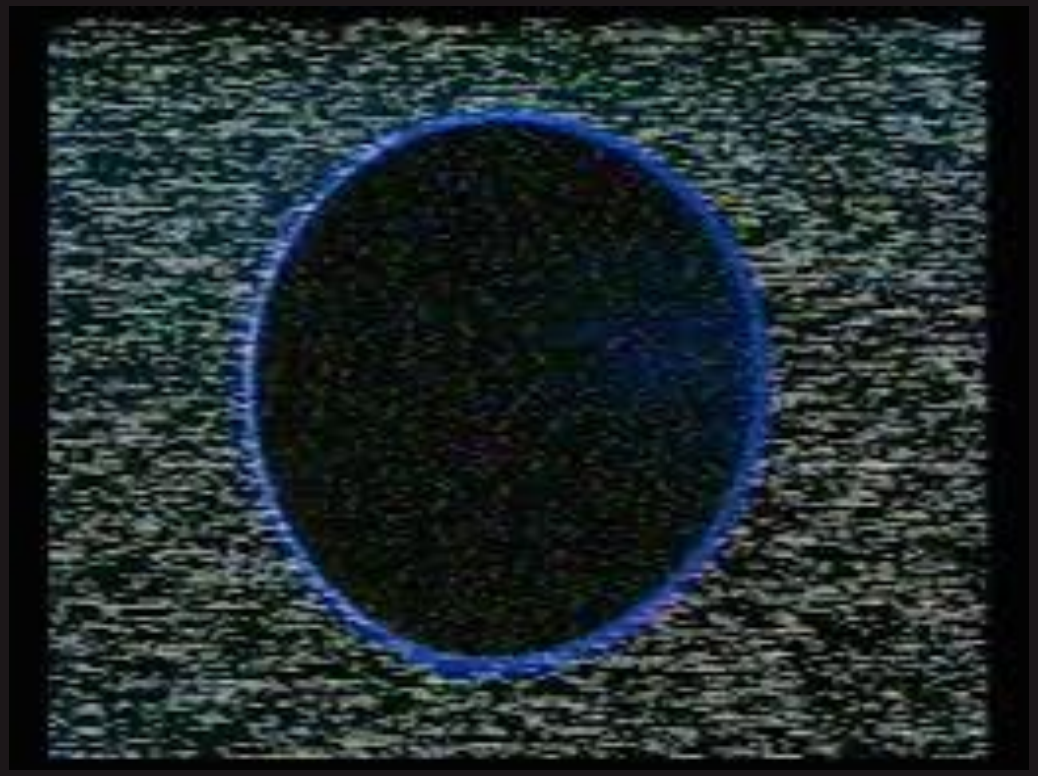
Sound Visualization is the process of converting sound signals (such as audio, sound waves, or vibration data) into visual representations. Through sound visualization, sounds that are typically perceived through hearing are presented in a visual form, allowing people to "see" sound. Characteristics of sound, such as amplitude, frequency, waveform, and sound field, can be displayed through graphics, waveforms, color changes, and other forms. Sound visualization is not only a visual mapping of sound but also a tool for exploring the structure and spatial propagation of sound.

Noisefields

Steina and Woody Vasulka, 1974

Noisefields" deconstructs the video signal to reveal its immaterial energy. The video imagery is shaped by the modulation and variation of signal frequency, creating a dynamic and evolving visual experience.

The Vasulkas intervened in the direction of the video signal, disrupting its standard horizontal and vertical flow. This disruption led to the creation of a three-dimensional form, deviating from traditional raster images.


Artwork Link: https://www.youtube.com/watch?v=qCYpNtRMh_g

Test Pattern

Ryoji Ikeda

In this artwork, the system converts various types of data—such as text, sounds, photos, and movies—into barcode patterns and binary sequences (0s and 1s). The primary aim of the project is to examine the relationship between device performance and human perception, pushing the boundaries of how we understand both sound and image.

Artwork Link: https://www.youtube.com/watch?v=jCR7KJQtwGE

LIFE—fluid, invisible, inaudible...

Ryuichi

For this work, approximately 400 audio/video sequences extracted from the original "LIFE" were combined with around the same amount of newly produced material. Images are divided into 15, sounds into 30 categories, whereas both components are algorithmically combined to create an infinitely transforming audio-visual setting.

Artwork Link: https://www.ycam.jp/en/archive/works/life/

REFERENCE

1) [1]Lucioni, S. (2019). *Digital Image* Sonification. GenEd 1080, December 11, 2019.
Article link :https://www.sarahlucioni.com/documents/1080FinalProjectReport.pdf

2) Artwork: Sonifying the colors from Hilma af Klint's 1907 painting Ynglingaåldern.
Link: https://liu.se/en/news-item/sonfiering-for-starkare-konstupplevelser

3) [2]Yeo, W. S., & Berger, J. (n.d.). Application of Raster Scanning Method to Image Sonification, Sound Visualization, Sound Analysis and Synthesis. Center for Computer Research in Music and Acoustics, Stanford University.
Article Link: https://ccrma.stanford.edu/~woony/publications/Yeo_Berger-DAFx06.pdf

4) Artwork: Voice of Sisyphus.
Link: https://digitalartarchive.at/database/work/2771/
5) [3] McGee, R., Dickinson, J., & Legrady, G. (n.d.). Voice of Sisyphus: An Image Sonification Multimedia Installation. Experimental Visualization Lab, Media Arts and Technology, University of California, Santa Barbara.
Article Link: https://lifeorange.com/writing/McGee_ICAD_2012.pdf

6) Artwork: Noisefields--Steina and Woody Vasulka, 1974
Link: https://www.youtube.com/watch?v=qCYpNtRMh_g

7) Artwork: Test Pattern---Ryoji Ikeda
Link: https://www.youtube.com/watch?v=jCR7KJQtwGE

8)Artwork: LIFE—fluid, invisible, inaudible... ---- Ryuichi

Link: https://www.ycam.jp/en/archive/works/life/

BOOK

1) O'Keeffe, L. (2024). Sound Art. In The Bloomsbury Encyclopedia of New Media Art, Volume 1: History and Theory (1st ed.). Bloomsbury.
2) Art and Electronic Media---Edward A. Shanken
3) Electric Dreams: Art and Technology Before the Internet----Dennis Del Favero
4) Art and the Thinking Machine: Coded: Art Enters the Computer Age, 1952-1982----Cinqué Hicks