Title: SonART

1-line description: "Every sound tells a story. We make sure it's seen." Classify the sounds within a track and convert them into images synchronized with the audio.

Description:

SonART is an application designed to help people with limited resources bring their performances to life, enhancing them through an original technology.

The user can upload an audio track they intend to use during their performance, and SonART will classify the sounds it contains and transform them into images that are always unique, thanks to the generative AI. These visuals will appear on screen in sync with the sounds they represent.

To make the experience even more visually engaging, a custom Processing script modifies the images to give them a unique artistic flair.

The user experience is simplified through a built-in GUI within the notebook, guiding users through the file upload and processing steps. The GUI also allows customization of certain parameters to improve the final result.

While the application works also with musical tracks and instruments, it is primarily designed to accompany theatrical performances, where environmental or narrative-driven sounds are used. In this context, SonART generates dynamic visual backdrops.

This is one of the system's key features, distinguishing it from other applications that are mainly intended to accompany musical performances.

Although this was the original concept that inspired the project, SonART also lends itself to more playful or educational purposes:

"Aren't you curious to see how that track will be transformed?"

It can help children associate sounds with images in a fun way, enhance storytelling during a Dungeons & Dragons session by preparing sounds and letting SonART handle the visuals, or even make performances more inclusive for the hearing-impaired.

You can also use it during relaxation sessions, pairing nature sounds with beautiful, evolving imagery.

In short, the only limit is your creativity.

Challenges, accomplishment and lessons learned:

The main challenges of this project were threefold:
Regarding audio, the accuracy of publicly available audio classifiers, while generally high, never reaches 100%. In particular, the models often struggle to distinguish between similar types of sounds, and the labels found in training datasets are sometimes too generic, sometimes ambiguous, or even overly specific.

As for image generation, one of the main difficulties was minimizing the time required to generate images as much as possible.

Finally, in terms of visualization, synchronizing audio and video, while also managing smooth transitions between images, proved to be a non-trivial task.

We aimed to make the system as robust as possible, balancing responsiveness with the accuracy of the classifier's labels. This involved managing the segmentation size of the audio clip and setting a minimum probability threshold that a class must exceed to be considered valid.

We also minimized image generation wait times and are quite proud of the final visual design used to display the images on screen.

Throughout the project, we learned to work with pre-trained neural networks and to understand both their limitations and their potential. We identified the features that mattered most to our goals and, by refining the input and the output processing, we sought an optimal trade-off tailored to the needs of our application.

Technology:

- Diffusion-based text-to-image generative model
- Convolutional neural network for Audio tagging
- Processing
- Python
- Tcp connection
- Json
- Ngrok
- Torch
- Socket
- Numpy
- Librosa
- Minim
- Google colab

Students:

- Chiara Auriemma: She worked on the audio classification part and on the synchronization between audio and images on the Python side.
- Matteo Benzo: He worked on the image generation part and on creating audio tracks for testing and for the final presentation

- <u>Anna Fusari</u>: She worked on the visualization part in Processing and managed synchronization between audio and images on the Processing side.
- <u>Diego Pini</u>: He worked on the image generation part and on the GUI integrated in the notebook.

Github Link: https://github.com/ChiaraAuriemma/SonART