

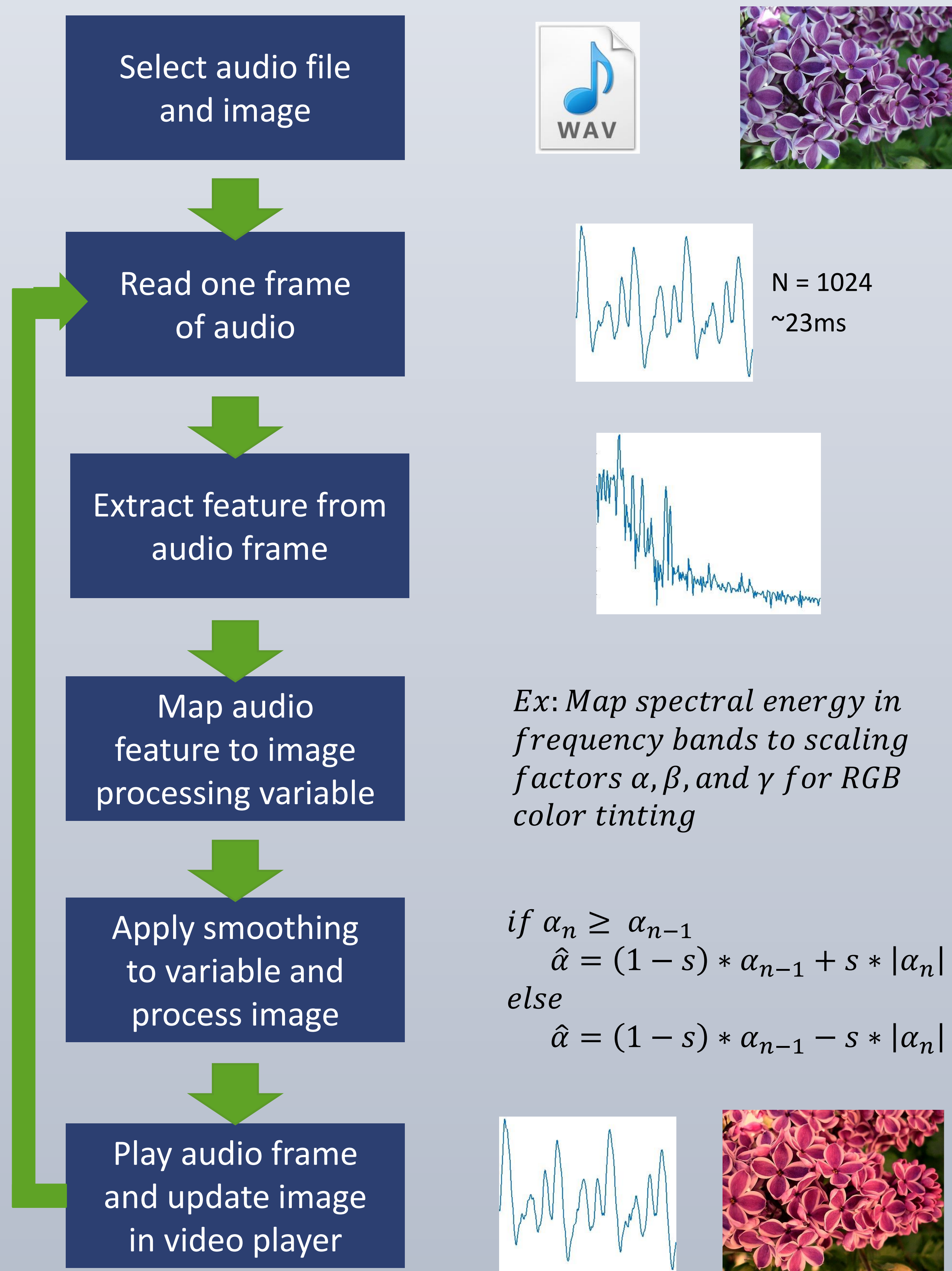
INTRODUCTION

Vision and hearing are two important senses that shape how we perceive the world. Often, what we see and what we hear are strongly connected to one another. In media that has both audio and visual elements, creators work hard to make sure that there is a connection between sound and images, and it is most often the audio that is molded to fit the visual element. I wanted to explore how the opposite might be achieved by manipulating a visual image based on audio features.

This project experiments with several methods of image manipulation: color tinting, contrast control, and blurring. Each method of image manipulation is tied to an extracted audio feature: spectral band energy, spectral flux, and signal energy, respectively. The feature gets mapped to a variable specific to its assigned image processing technique (for example: energy in spectral bands are mapped to scaling factors in RGB color tinting), and a smoothing function is applied to prevent sudden changes in the processed image that would produce unpleasant jitter.

Currently, this process is implemented in MATLAB using objects from the DSP and Computer Vision toolboxes. Audio is read frame-by-frame from an audio file, which is then analyzed and used in the processing of the originally selected image before both the processed image and the original audio frame are output.

PROCESS OVERVIEW



TYPES OF IMAGE PROCESSING

IMAGE PROCESSING: COLOR TINTING AUDIO FEATURE: SPECTRAL BAND ENERGY

In order to achieve a color tinting affect in an RGB image, three scaling factors (α , β , and γ) were used to scale the intensity values of each of the three layers of the RGG image. A larger scaling factor will mean a stronger presence of a color in the processed image.

The audio feature used to determine the values of these scaling factors is the spectral energy in three frequency bands of the audio frame. The frequency spectrum is divided into low (20-500Hz), mid (500Hz-1500Hz), and high (1500-20,000Hz) frequency bands, and the spectral energy in each band is calculated using (1). The ratio of energy in each band to all of the energy in the spectrum is used as the scaling factor for the color tinting (2) and (3).



$$E_{band} = \sum_{freq.in\ sub-band} X^2(f) \quad (1)$$

$$\alpha = \frac{E_{band}}{E_{total}} \quad (2)$$

$$J(i, j, 1) = I(i, j, 1) * (1 + \hat{\alpha}) \quad (3)$$

Fig. 1 Images processed with red, green, and blue tinting from large amounts of spectral energy in low, mid, and high-frequency spectral bands.

IMAGE PROCESSING: CONTRAST CONTROL AUDIO FEATURE: SPECTRAL FLUX

To control the contrast of an image, the MATLAB function imadjust() was used to specify the image's contrast limits. The contrast limits were calculated by the spectral flux between two consecutive frames. This mapping was chosen with the intention that notes onsets would be visually represented by high-contrast images in the video player.

When using imadjust() with an RGB image, contrast limits for each RGB layer must be specified. Because the spectral energy in three bands are already being calculated for the color tinting effect, these spectral band energies are also used in the calculation of the three layer's contrast limits. To calculate the contrast limit for a layer, the change in spectral energy for that band is calculated (4). It is then multiplied by a scaling factor, ρ , and added to a constant (5) to prevent the image from being reversed (i.e. become a photographic negative). Equation (5) provides the upper contrast limit, and subtracting this from 1 provides the lower contrast limit (6).

$$\Delta E = \frac{\rho}{\sum_{f=f_0}^F [X_n(f) - X_{n-1}(f)]^2} \quad (4)$$

$$\lim_{high} = \Delta E * \rho + 0.05 \quad (5)$$

$$\lim_{low} = 1 - \lim_{high} \quad (6)$$



Fig. 2 Image processed with high contrast due to large spectral flux between audio frames.

TYPES OF IMAGE PROCESSING (CONT.)

IMAGE PROCESSING: BLURRING (GAUSSIAN FILTER) AUDIO FEATURE: SIGNAL ENERGY

A blurring effect was obtained by using a Gaussian filter. The sigma value of the filter was determined based on the amount of signal energy in a frame of audio. The higher the energy, the more blurring effect appears in the processed image. This was a simple way to visually map the loudness of a sound to a visual medium.

The mapping from signal energy to the sigma value for the Gaussian filter is very simple. The sigma value is just a scaled version of the signal energy for the frame (7). The scaling value, ρ , is chosen by inspection so that the processed images aren't all extremely blurry or all hardly have any blurring at all. The optimal amount of scaling varies between audio files. Exploring ways to have self-adjusting scaling for the sigma value could be done during future research.

$$\sigma = \sum_{n=1}^{1024} [x^2(n)] * \rho \quad (7)$$



Fig. 3 Processed image with a large amount of blurring due to high signal energy in an audio frame.

FUTURE WORK

- Explore more complex relationships between audio and image processing
The audio features and image processing techniques used so far were chosen for relative ease of implementation. With more time to explore this concept in the future, I would like to explore other image processing techniques paired with other audio features to represent more complex, higher-level audio features such as genre or mood.
- Port outside of MATLAB into C or C++
By creating a C/C++ program that will do this processing, MATLAB's just-in-time compiler can be bypassed, allowing for faster processing. This will allow us to use more types of image processing at once and more complex algorithms to map more complex relationships between audio and image manipulation.
- Allow for image manipulation from audio captured through a microphone
Currently, only audio files are being used for the sound sources in this project. I would like to extend it to work with microphone input as well. With MATLAB's new Audio System Toolbox available in R2016a, this should be a relatively simple task.

ACKNOWLEDGEMENTS

Thanks to Professor Duan for providing feedback and insight throughout the completion of this project, and the ECE Department for covering the cost of poster printing.