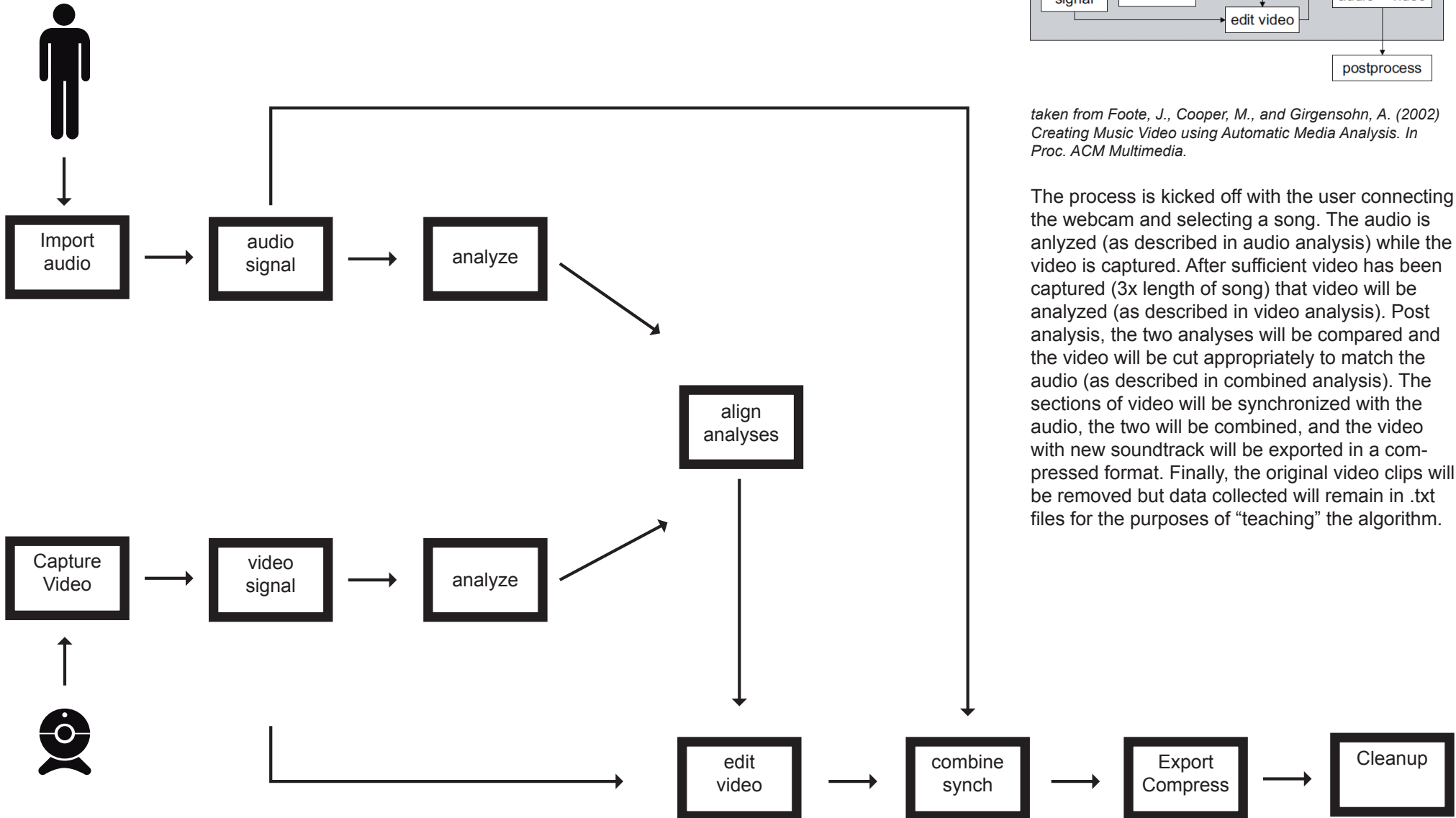# Sedentary Panopticon

Use / Build Sketches

High Level Process
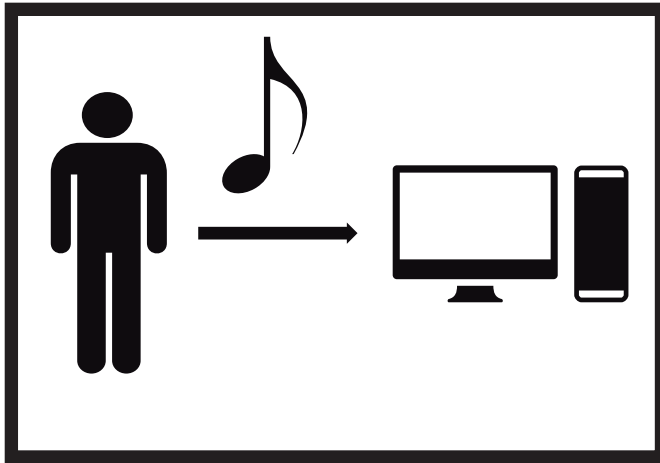## Broad overview of how the application functions

taken from Foote, J., Cooper, M., and Girgensohn, A. (2002) Creating Music Video using Automatic Media Analysis. In Proc. ACM Multimedia.

The process is kicked off with the user connecting the webcam and selecting a song. The audio is anlyzed (as described in audio analysis) while the video is captured. After sufficient video has been captured (3x length of song) that video will be analyzed (as described in video analysis). Post analysis, the two analyses will be compared and the video will be cut appropriately to match the audio (as described in combined analysis). The sections of video will be synchronized with the audio, the two will be combined, and the video with new soundtrack will be exported in a compressed format. Finally, the original video clips will be removed but data collected will remain in .txt files for the purposes of "teaching" the algorithm.
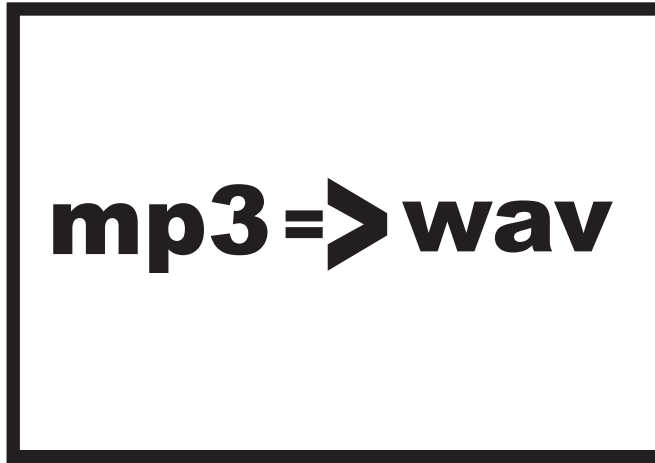
Import audio → audio signal → analyze → align analyses → edit video → combine synch → Export Compress → Cleanup

Capture Video → video signal → analyze → align analyses

Audio Analysis

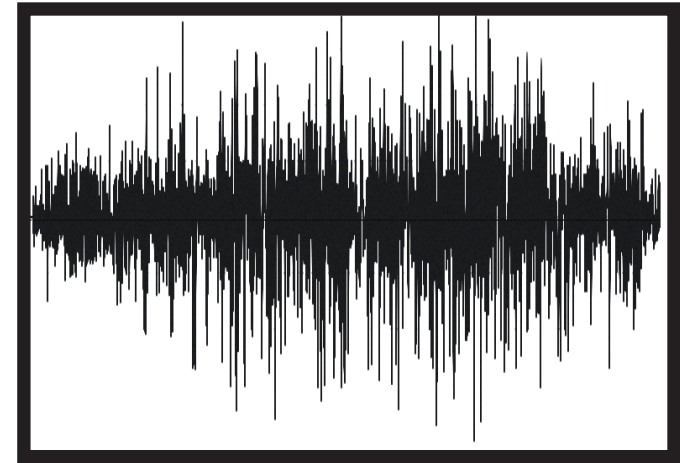## How the audio will be turned into useful data

The user will input a song, it will be analyzed and numeric representation of the signal (waveform) will be generated

User selects a song because of personal meaning, popularity, or just enjoyment and uploads it through the application's interface.

The song (most likely in mp3 format) will be converted to a wav. While this will do nothing to retrieve data loss from compression, it will allow for more robust manipulations.
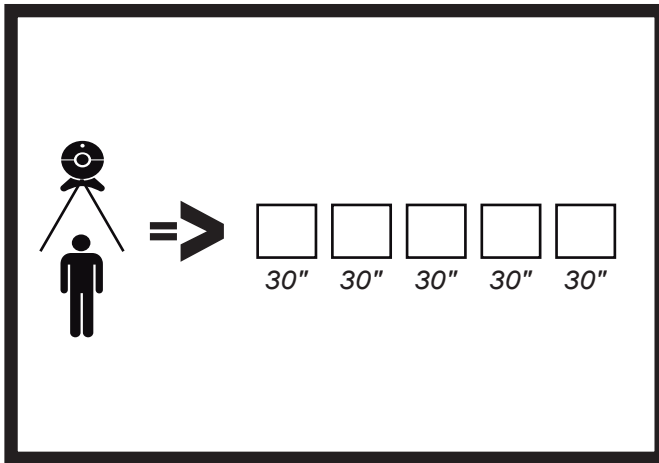

The song will be translated into its corresponding waveform, which is essentially a mathematical model of the way that a sound changes the air. Bottom line - a representation of the music that can be analyzed mathematically.

Video Analysis

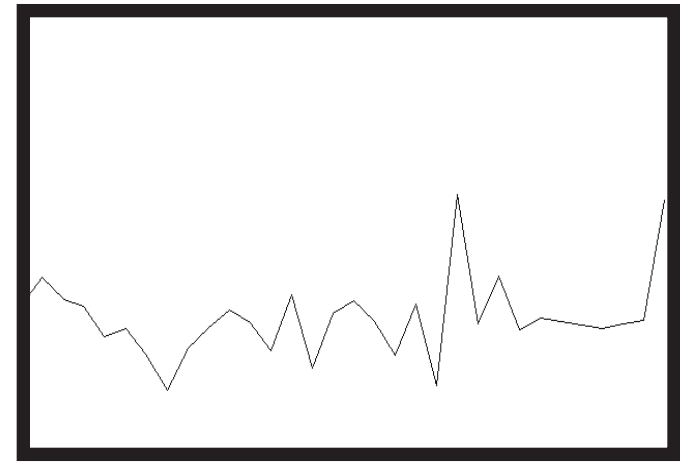## How the video will be turned into useful data

Having generated and analyzed a waveform for the audio, creating a similar form for the video would be the next step.

Clips are captured in 30 second chunks, for the sake of memory. These clips are saved, numerically, to the hard drive along with a text file corresponding to the clip.

The contents of the text file are a numeric evaluation of the clip. During recording, data is captured on a basis of the number of pixel changes from frame-to-frame. This creates a numeric representation of the amount of movement between two frames (Numbers range between 0 and 307200 for 640 x 480 video).
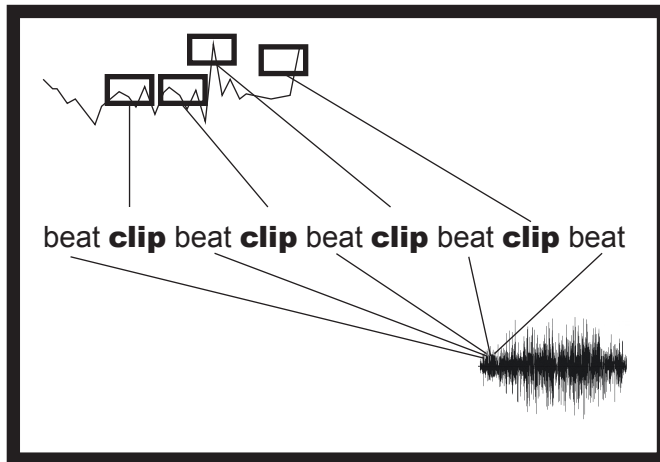

This information regarding the amount of change between frames can be used to create a histogram, similar to the audio section. Since these files are smaller than the full video files, and therefor a little easier to handle, all of the clips will be combined, in order, to make one long histogram.

Combined Analysis
How the video will be mapped to the audio

While both sets of data can be mapped to a wave, the automation process hinges on the ability to compare those waves in a meaningful way, which cannot be done without some sort of further manipulation.

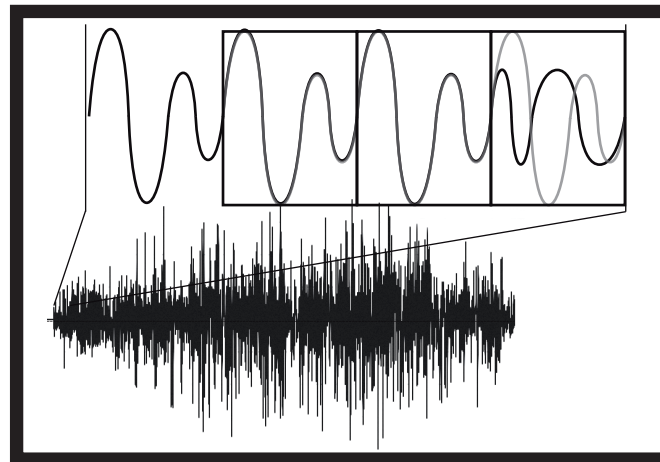beat **clip** beat **clip** beat **clip** beat **clip** beat

clip

clip

v1
One of the main differences between the two waves is that the waveform from the audio comprises the entire character of the sound, while the data from the video only encompasses motion. An easy (and readymade for prototyping) approach would be audio beat analysis. By determining where the beat is, clips can be cut to be on-beat, with an emphasis on high-motion clips.
Benefit: Relatively easy / fast to implement
Drawback: lacks novelty over time
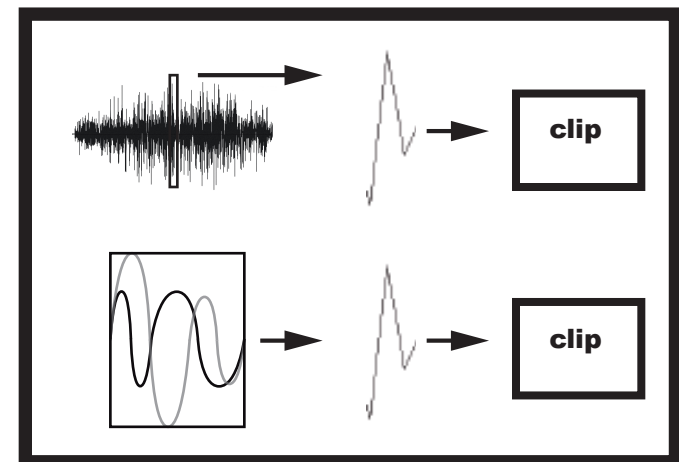Conclusion: Useful for early testing, but not an end in itself.

v2
Rather than take in just the beat, this approach will examine the waveform for patterns (repetitions that separate music from noise) then use those patterns to compare the rest of the song to that pattern, arriving at a novelty score for different sections of the song (Foote et al. describes this). The highest "novelty" sections could then be mapped to the parts of the video with the greatest movement.
Benefit: Uniquely appropriate video/audio combination
Drawbacks: Lengthy development
Conclusion: Use v1 to test parallel to this version; inform further development from findings.

Since the video editing is "driven" by the content of the audio, it's editing is more straight forward. Units will be taken from the previous analysis, applied to the video and optimization will be applied to find appropriately lengthed clips that have the most "action" (based off of the motion analysis). For v1 clips can be placed with higher movement clips placed in areas of increased amplitude, then beat speed, then arbitrarily assigned. For v2, the evaluation of "novelty" in the audio should map to the amount of clip "action."

Foote, J., Cooper, M., and Girgensohn, A. (2002) Creating Music Video using Automatic Media Analysis. In Proc. ACM Multimedia.