Real-Time Conducting Feedback using Computer Vision and MediaPipe

Vivek Bharadwaj
Faculty Mentor: Dr. Andrea Salgian
The College of New Jersey

ABSTRACT

The goal of this project is to detect conductor's movement using MediaPipe, while providing feedback on form and correctness, and helping aspiring conductors to improve their skills.

MOTIVATION & BACKGROUND

Prior projects used sensors Myo armband and Microsoft Kinect to detect conducting movement but were discontinued. The adaptation of newer technology, such as MediaPipe, will provide more reliable collection of joint coordinates for the user's skeleton to detect beats, mirroring, swaying, and gaze direction.

This program provides real-time feedback on:
- Beats per Minute (BPM)
- Number of Beats
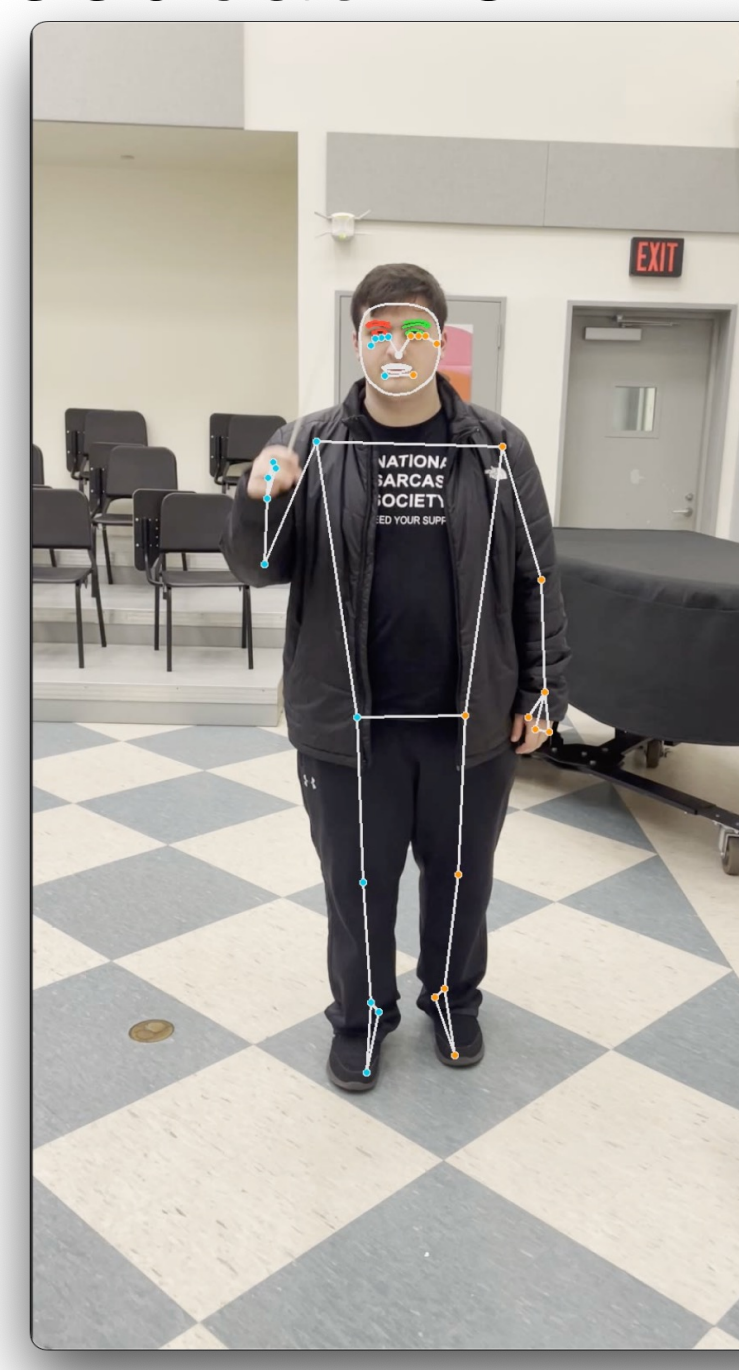- Swaying
- Mirroring
- Gaze Detection

Fig 1: MediaPipe Skeleton (right)

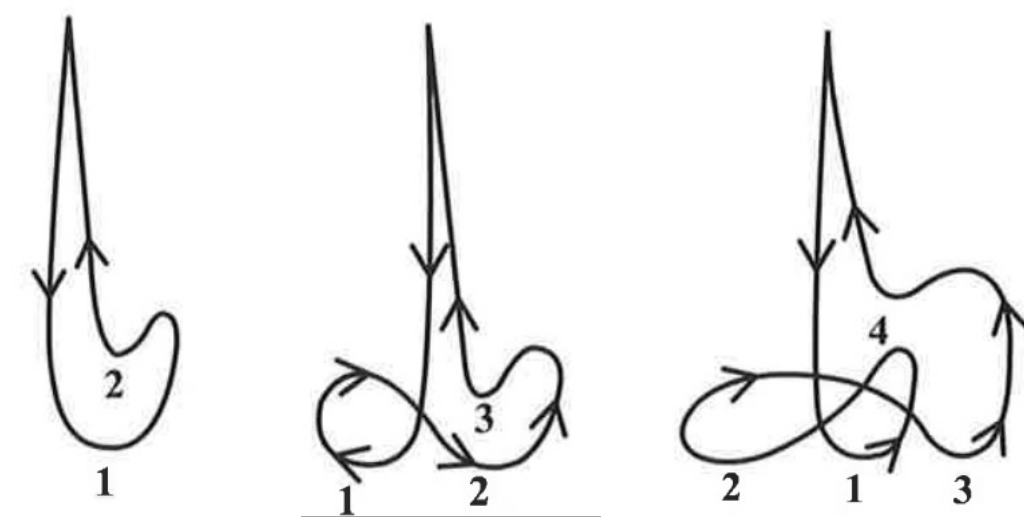Three conducting gestures were referenced to detect beats in beat patterns: 2/4, 3/4, and 4/4.

Fig 2: Conducting gestures for beat patterns: 2/4, 3/4, and 4/4 (left)

The collection of joint coordinates is used to detect beats for 2/4, 3/4, and 4/4 beat patterns from test videos. Wrist, shoulder, and nose coordinates are compared to a threshold to alert the user of swaying, mirroring, and the direction of their view.

SYSTEM REQUIREMENTS

- Anaconda Python v3.9, Google MediaPipe v0.7, OpenCV, Pickle, Matplotlib, NumPy, Scipy.signal, Phone camera

METHODS

To detect a beat, the left wrist y-coordinate is collected from a video. If the current frame consists of a local maxima in the left wrist y-coordinate, then a beat is considered to be detected.

To determine the BPM, the number of frames between local maxima is divided by the frame rate resulting in number of seconds between peaks. Divide 60 by the number of seconds to get the new BPM. If the peak is not detected, the script uses the previous BPM value instead.
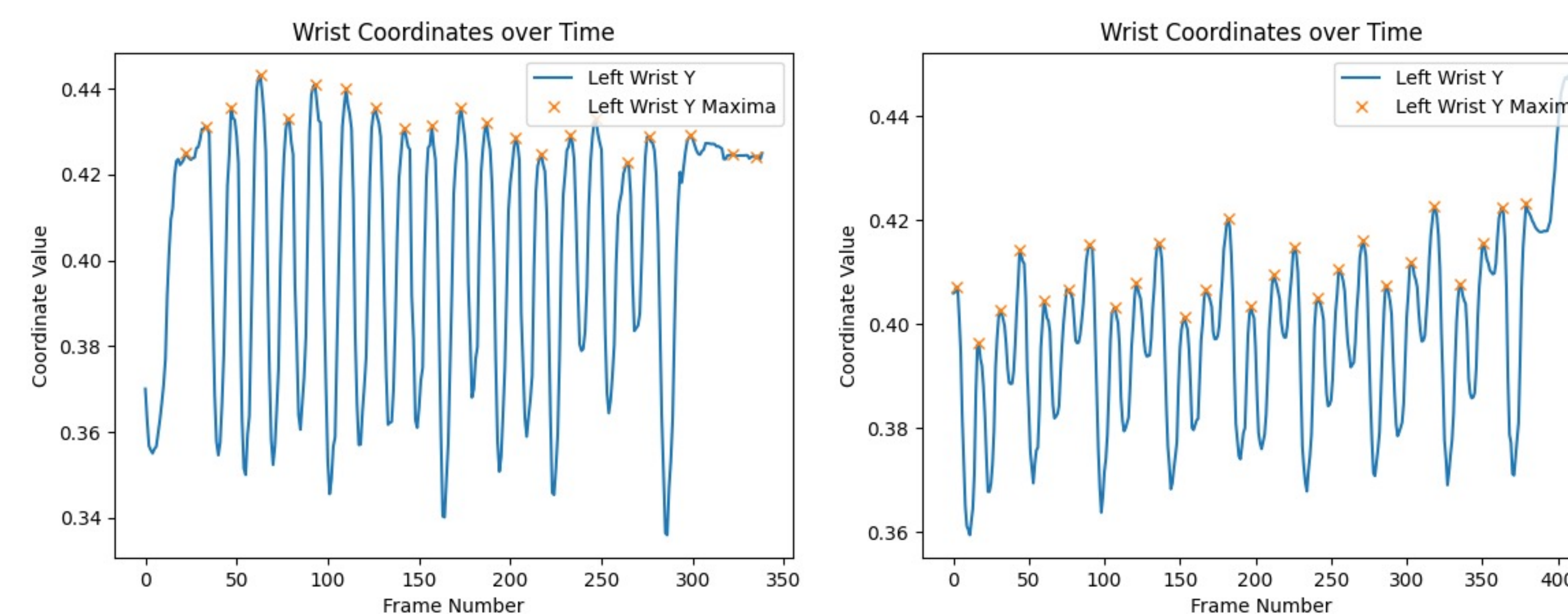
Fig 3: 2/4 pattern's maxima (left) and 3/4 pattern's maxima (right)

$$BPM = 60 / ((Beat\_frame - prev\_beat\_frame) / frame\_rate)$$

To detect mirroring, the difference of both wrist y-coordinates is compared to a set threshold of 0.05. The difference between both wrist x-coordinates and the shoulder midpoint is compared to a set threshold of 0.05.
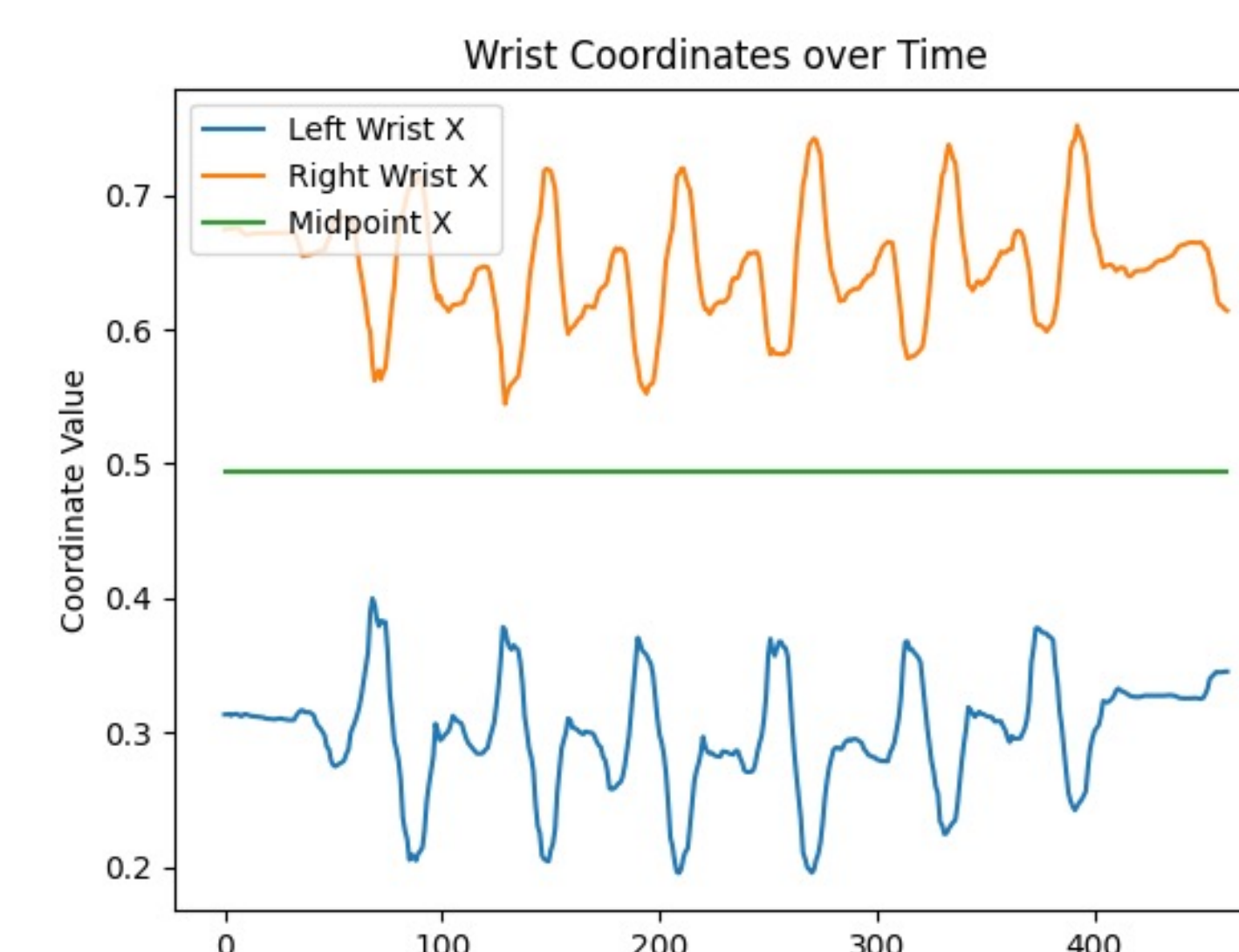
To detect the direction of gaze, the difference between the current and initial nose midpoint is compared to a threshold of 6% of the initial midpoint.
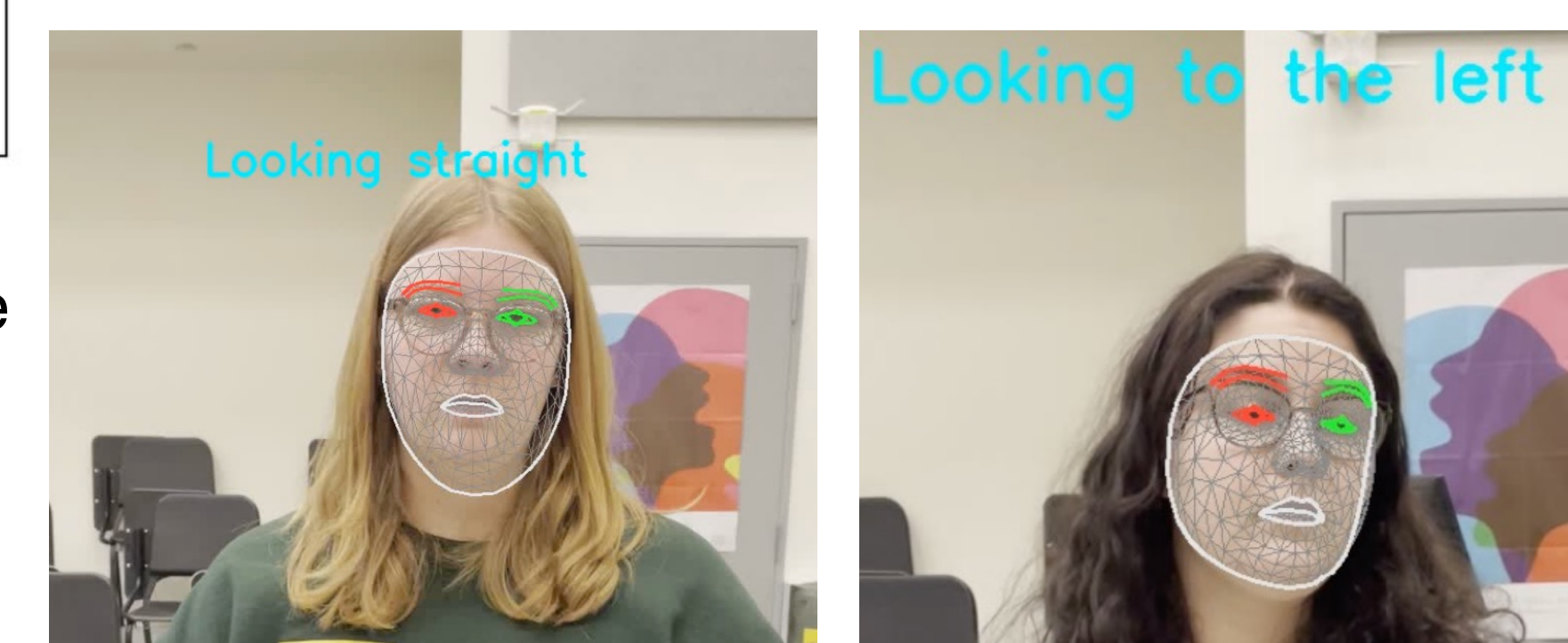
Fig 4: Mirroring coordinates with a line of symmetry (above)

Fig 5: Gaze direction using FaceMesh (above)

RESULTS

- Beats are correctly detected, the BPM is correctly calculated, and the overall feedback provided is correct
- Monitors user's conducting form and rhythm through their shoulder and wrists positions
- Outputs video that includes feedback analysis which alerts user of current BPM, number of beats, swaying, gaze direction, and mirroring
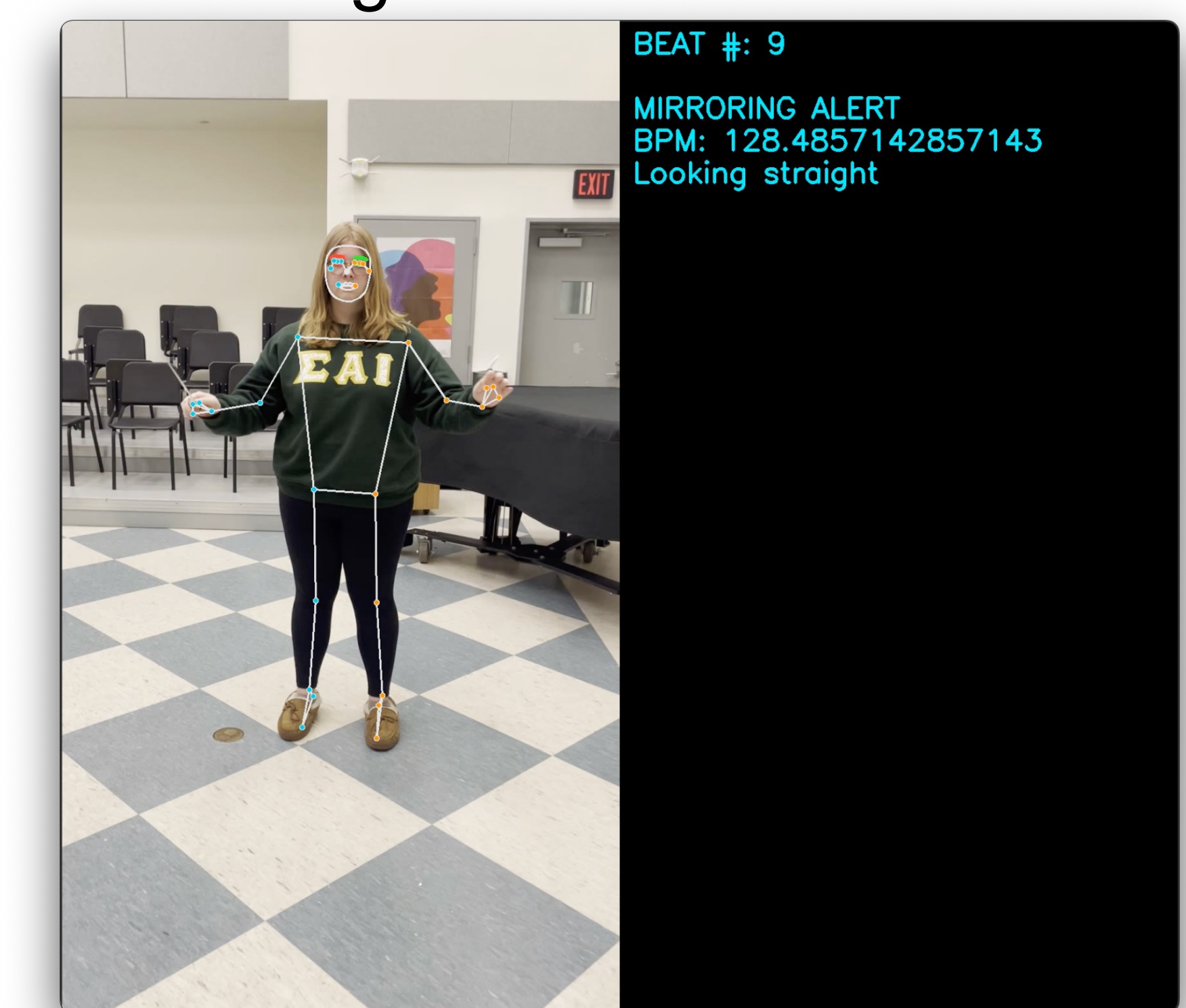
Fig 6: Overlay of Feedback Analysis on Output Video

FUTURE WORK

- Integrate z-coordinate for swaying to track if user is moving back and forth
- Conduct more testing with different patterns and more videos
- Create real-time application that provides similar feedback analysis

REFERENCES

[1] https://github.com/google/mediapipe
[2] Andrea Salgian, David Vickerman, David Vassallo, "A Smart Mirror for Conducting Exercises", *The 25th ACM Multimedia Conference – Thematic Workshops*, Mountain View, California, October 2017.
[3] Andrea Salgian, David Vickerman, "Computer-Based Tutoring for Conducting Students", *The 42nd International Computer Music Conference (ICMC 2016)*, Utrecht, The Netherlands, September 2016.
[4] Nowak, J., & Nowak, H. (2002). *Conducting the music, not the musicians*. Carl Fischer.

Presented at Celebration of Student Achievement, Spring 2023