# Comparative Motion Analysis for Technique Development

*Improving physical performance through applied computer vision*

## Introduction

Developing highly trained physical performance techniques is critical for athletes and musicians. This often involves:

- intricate coordination
- motor planning and muscle memory development
- fine control over simultaneous movements of multiple body parts
- spatial and temporal precision

These all require **practice** to achieve mastery

**Problem**

- During individual practice, self-monitoring of multiple simultaneous movements with precision is challenging.
  - Subtle shifts in position and timing can result in significant outcomes
  - When focusing on improving one or two aspects of motion, one may neglect to control other important simultaneous motions

**Design Goal**

- My project goal was to develop a computer-vision based system which can analyze a user's physical performance technique through widely-available single-camera user input (e.g. a smartphone) and give dynamically generated feedback for more effective practice.
- The scope of the project was limited to a specific subset of violin techniques: the left hand vibrato

## *Vibrato* performance technique

Violin *vibrato* is an expressive left-hand technique that oscillates pitch with varying speed, width and intensity over time, and gives each individual violinist their distinctive signature sound.

Each finger can vibrate on any string in any position, and the violinist must remain in tune and on tempo while simultaneously bowing with the right hand.

The player controls pitch by the contact point of the finger pressing on the string and achieves pitch oscillation by rocking the finger to shift the contact point back and forth, which modulates the length of the vibrating segment of the string. The impetus of finger motion emanates from finger, wrist, arm, or some combination, which leads to a wide variety of types of vibrato.

## Design

My approach was to make recordings of expert violinists performing various types of *vibrato* as a basis for comparison and to determine critical features, then to implement visualizations of comparison metrics in software

**Phase 1: Data Collection**

Experienced Violinists were recorded playing key types of vibrato across tempos, fingers, strings and positions

- Finger vibrato
- Wrist vibrato
- Arm vibrato

**Phase 2: Data Analysis and critical feature extraction**

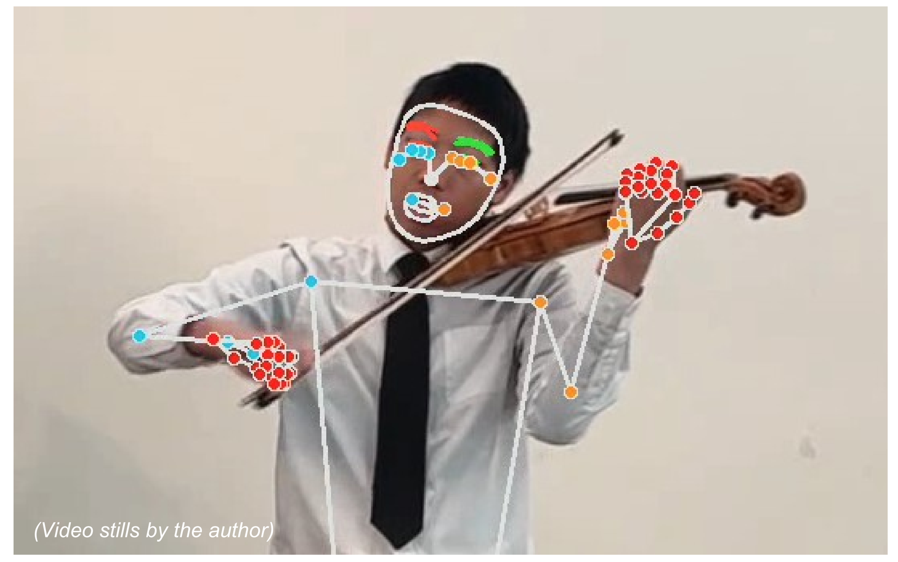
- The MediaPipe Holistic pose detection library was used to extract time-series 3D coordinates of key joint positions from recorded videos
- A hierarchical sequence of 3D translations/rotations was performed on the joint position data to reorient key body components (eg. wrist, palm, thumb, fingers) to canonical reference positions and orientations
- Key composite features were computed
  - Thumb/palm angle and finger curvature
  - Keyframes of poses at start of technique

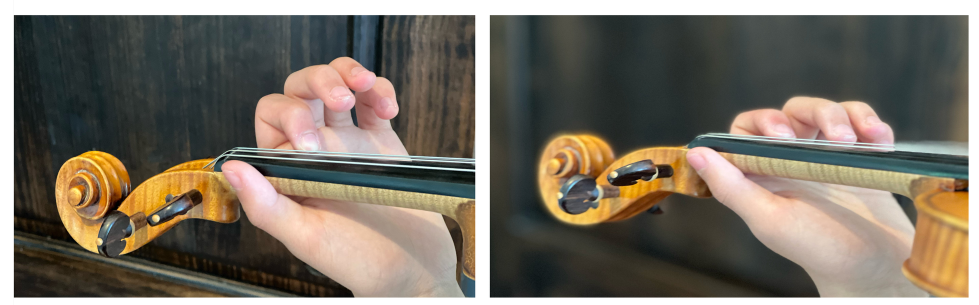
**Phase 3: Expert comparison and feedback**

- Expert and trainee time series data were aligned in time based on composite feature comparison between expert key frames and trainee time series features
- Expert and trainee data were spatially aligned and overlaid to give visual feedback.
- Differences are computed and displayed.

## Analysis pipeline

### Joint Position Extraction

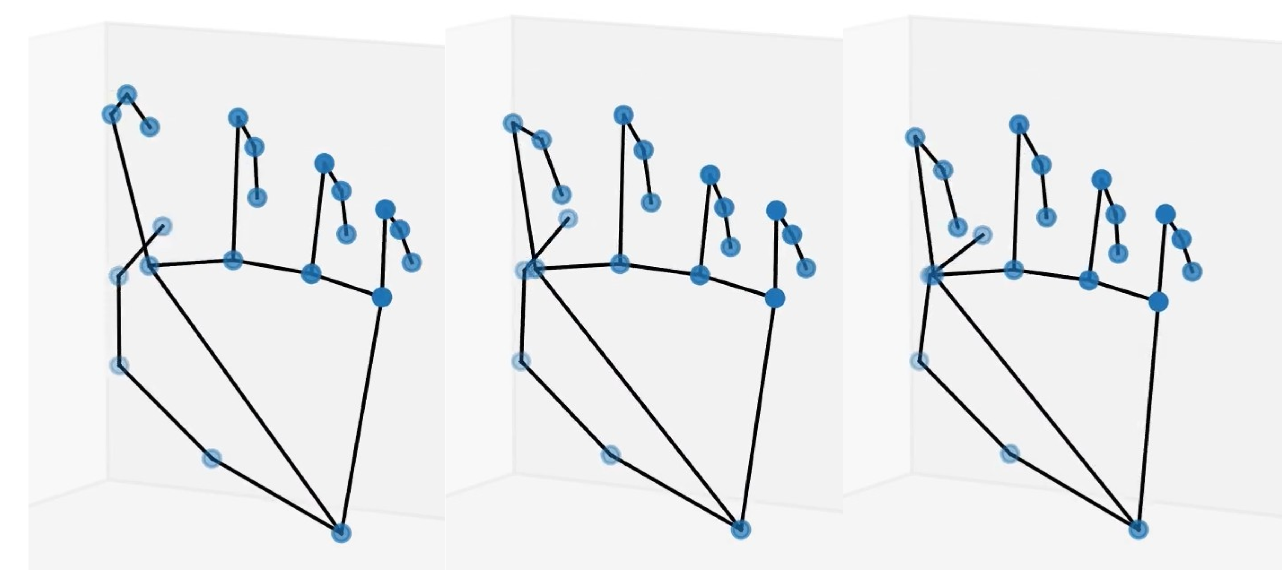


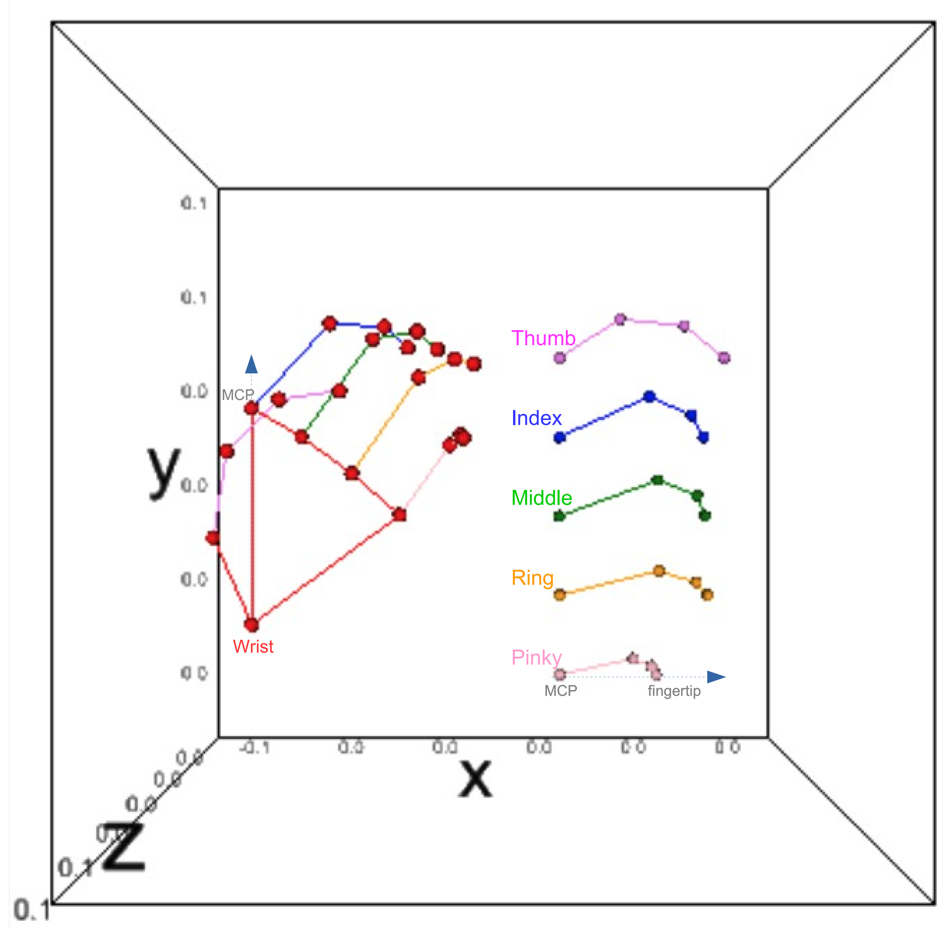
A) Start: Squared finger B) Collapsed finger

The **Joint Position Extraction** phase includes two important aspects:

1) Video is analyzed using the MediaPipe holistic model, which computes a time-series 3D point cloud of joints of the subject in motion
   - Hand model: 21 points x {left, right}
   - Body model: 33 points
2) Since video capture data is often noisy, missing coordinates and dropped frames are reconstructed through interpolation:



### Data Alignment




The trainee position data must be **aligned** with the reference expert position data in order to perform the comparison.

**Spatial alignment** is performed hierarchically, eg.

Torso → Shoulder → Elbow → Wrist → Palm → Thumb/Fingers

At each level of hierarchy, the upstream component serves as an anchor and downstream components are reoriented to align with a predefined canonical orientation.

For the hand's canonical orientation,

- The wrist serves as the local origin (0,0,0)
- The wrist → index MCP (metacarpophalangeal joint) vector is aligned with the Y axis
- The wrist → pinky MCP vector lies in the XY plane
- The palm normal vector points towards the viewer

For the fingers, the joints are approximated as hinges with single degrees of freedom

- The finger MCP serves as the local origin (0,0,0)
- The finger MCP → fingertip vector is aligned with the X axis
- The curvature of the finger lies in the XY plane above the X axis

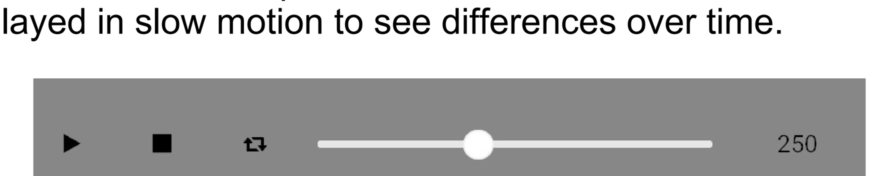
Spatial comparison is performed by placing both the trainee and expert pose components into their canonical orientations and computing distance

**Temporal alignment** is performed by finding the minimum distance frame across the trainee time series compared to key frames from the expert's pose data

### Comparison and Feedback

Once aligned, the expert's canonical pose (gray) is overlaid on the trainee's pose (multicolored) to enable a visual **comparison**.

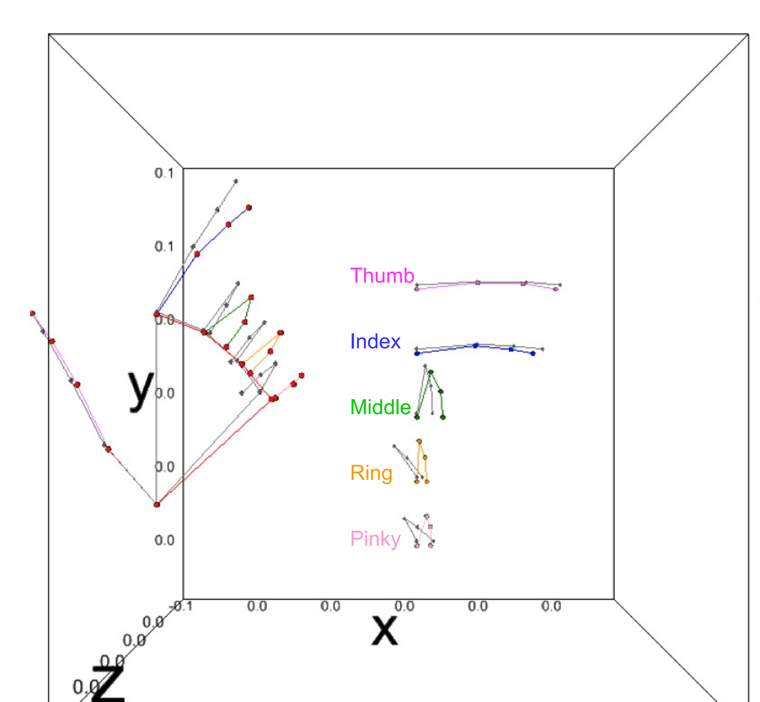
Animation can be paused and rolled backward/forward or played in slow motion to see differences over time.



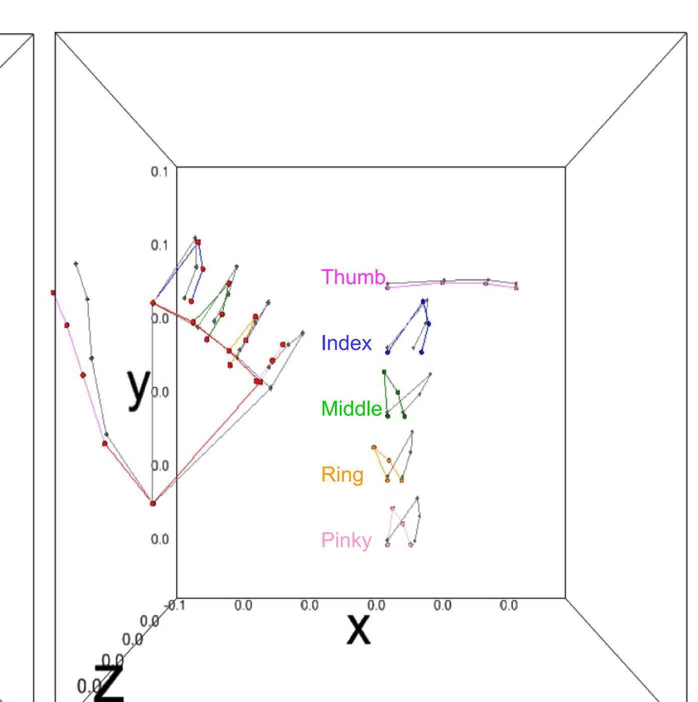

The figures show two examples of a ring finger vibrato at different string positions

A) In this high position vibrato, the expert's palm is curved slightly further inward to achieve a better finger angle

B) In this low position vibrato, the expert's palm is flattened, also to achieve a better finger angle. The thumb is also in a more neutral relaxed position.




A) High position vibrato




B) Low position vibrato

## Results

My application of computer vision and development of a comparison system to analyze motion has resulted in a software feedback system that can benefit individual practice by visualizing differences between practice trials and an expert's reference motion.

This work moves beyond traditional motion capture schemes which typically involve considerable overhead such as physical markers which may interfere with or obscure hand motion, as well as multiple camera schemes requiring calibration. This work uses single-camera input from widely available sources such as modern smartphones.

My hierarchical comparison scheme enables visualization and comparison of fine motion details and allows motions of particular components of interest such as individual fingers to be isolated and examined in detail.

Key challenges included noisy data from the MediaPipe 3D point cloud extraction, which often dropped frames in which hands or other components were occluded or otherwise not recognized I applied inter-frame interpolation to reconstruct the missing data.

Additional challenges were encountered in alignment, where similar (but not identical) motions were executed in sequence, especially in playing violin scales, where finger positions shifted in rapid succession. Future work may include incorporating secondary inputs such as audio pitch detection to improve temporal alignment.

## Conclusion


**Summary**

- My system enhances the trainee's ability to perceive the details of their vibrato technique by generating visualizations of comparisons with an expert demonstration so they can see their relative movements and to highlight the specific motions in their technique that should be adjusted


**Future directions**

- My system can be augmented to analyze other physical performance techniques involving hands or other body parts, as the framework can be readily adapted. Similar recordings of experts would have to be done to provide references for comparison for the additional techniques.

## References


- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.
- Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.
- Gustus, A., Stillfried, G., Visser, J. et al. (2012). Human hand modelling: kinematics, dynamics, applications. Biol Cybern 106, 741–755). https://doi.org/10.1007/s00422-012-0532-4
- Jarque-Bou, N.J., Atzori, M. & Müller, H. (2020. A large calibrated database of hand movements and grasps kinematics. Sci Data 7, 12). https://doi.org/10.1038/s41597-019-0349-2
- Galamian, Ivan. Principles of Violin Playing & Teaching. Shar Products Co., 1999.
- Flesch, Carl *The Art of Violin Playing*. Carl Fischer. (2000).