Music in *Motion*

A Real-Time Motion-Responsive Sound System to Enhance Physical Engagement

Ciana Tzuo
Horace Mann School

Big Idea

Can real-time sensing technologies be combined to create a continuous, responsive system that transforms human movement into sound?

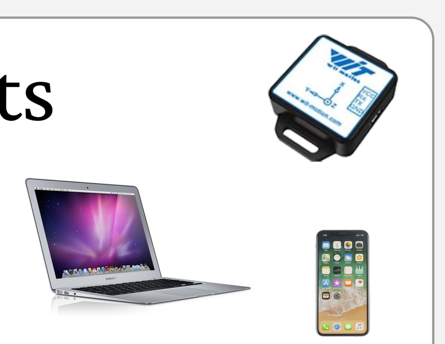
Design Challenges

X- Full Body	encourage use of entire body in dance-like motions
X- Immersive	proprioceptive, natural movement enabling intuitive learning
X- Expressive	encourages fluid engagement across a wide range of motion dynamics
T- Real Time	low latency < 70 ms to ensure responsive smooth transitions
T- Precise	motion tracking > 85% to create strong perceptual link between action and sound
T- Stable	sustainable operations > 15 min continued use with no noticeable drift
$- Affordable	widely available consumer grade hardware and open source software
$- Convenient	operates effectively in typical indoor space, lighting

X -Experiential Goals   T -Technical Constraints $ Cost Constraints

Key Components

- MacBook Air w/built in camera
- iPhone camera
- WitMotion WT901WIFI IMU (up to 4)
- Pose estimation: MediaPipe BlazePose
- Software Stack:
  - Python 3 with OpenCV, NumPy, SciPy
  - Audio: Sounddevice, libross
  - Graphical UI: PyQT5, Matplotlib
  - Image capture and processing: OpenCV



References

- MediaPipe Solutions guide. (n.d.). Google AI for Developers.
- WitMotion. (n.d.). WT901WIFI protocol..
- Winkler, T. (2002). Fusing Movement, Sound, and Video in Falling Up, an Interactive Dance/Theatre Production.
- Bergsland, A., etal. (2015). Composing Interactive Dance Pieces for the MotionComposer, a device for Persons with Disabilities.
- Kantan, P. R., Dahl, etal (2024). Making Movement Sonification Usable in Clinical Gait Rehab.
- Bao, Y., Zhao, etal (2022). FusePose: IMU-Vision Sensor Fusion in Kinematic Space for Parametric Human Pose Estimation.
- Fang, D.-L., Qin, Y., Chen, Q.-B., & Lu, J.-L. (2012). Real-Time Audio Synthesis In A Wireless Interactive Sensor Platform

Motion Sensors   Body Pose Sensors

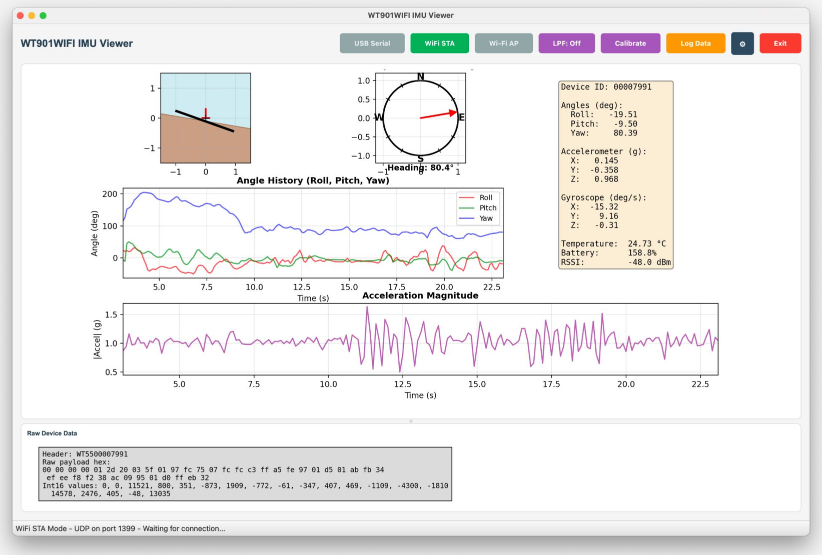
Single Fused Sensor

hand_height_L, hand_height_R	Hand height relative to torso (normalized).
arm_extension_L, arm_extension_R	Arm extension (shoulder–wrist).
elbow_bend_L, elbow_bend_R	Elbow bend (shoulder–elbow–wrist)
lateral_offset_L, lateral_offset_R	Wrist lateral offset from torso center: dx = wrist.x - torso_center, then mapped to [0, 1] with 0.5=center.
hand_spread	Horizontal spread between left and right wrists (relative to shoulder width).
activity_L, activity_R	IMU activity (accel + gyro magnitude, smoothed).
jerk_L jerk_R	IMU jerk (rate of change of acceleration)
shake_energy_L shake_energy_R	IMU shake energy (bandpass-filtered accel)
activity_global	Global activity (average of left and right)

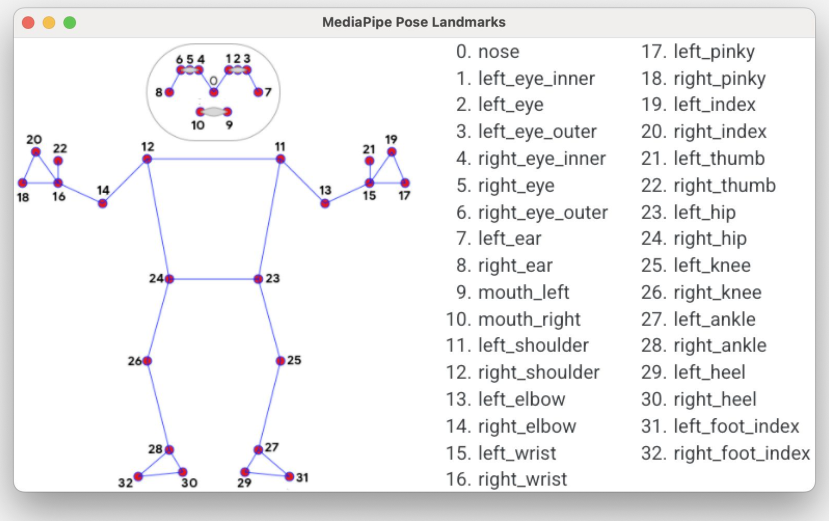
Audio Feature Mapping

Control	Air DJ	Calm vs Intense	Two-Hand Instrument
V_cutoff	hand_height_R	hand_height_R	hand_height_L
V_resonance	hand_height_L	jerk_L, jerk_R (burst envelope)	elbow_nend_L (inverted)
V_attack		arm_extension_L	jerk_R (burst envelope)
V_brightness	hand_spread	hand_height_R (curve)	0.2 + 0.8 x cutoff (derived)
V_tremolo	shake_energy_R + hand_spread		shake_energy_R
V_volume	activity_global	activity_global	activity_R
V_mode		Hand_spread (hysteresis)	Fixed at 0.6
V_pan	lateral_offset_L, lateral_offset_R		lateral_offset_L

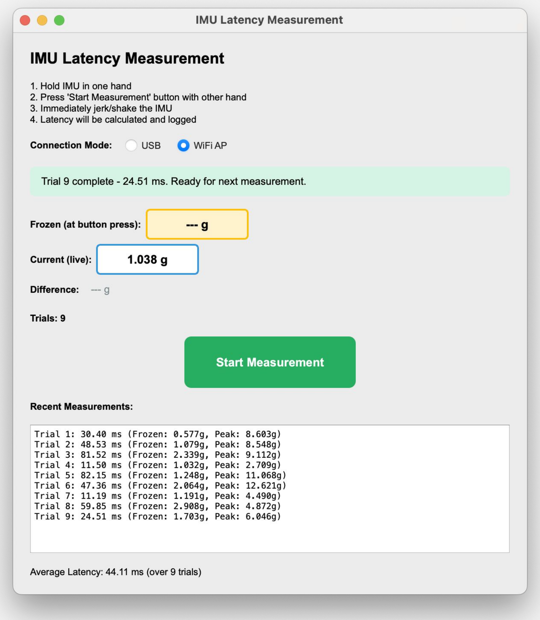
Real Time Audio Transform



IMU Sensor Data



Body Pose Data



Latency Tester

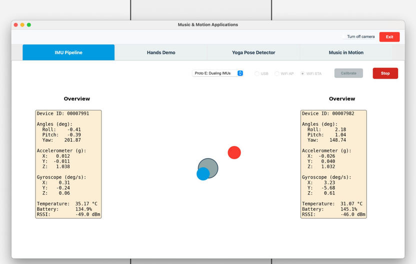
Key Learnings

- IMU vision-based sensors have complementary strengths and weaknesses
- Latency, pose detection, and intensity of motion req to create immersive experience
- Motion-to-audio mapping needs to be intuitive vs too fine grained

Future Applications

- Wellness; enhanced bodily awareness
- Dance and creative movement exploration
- Rehabilitation and physical therapy
- Interactive media and art
- Entertainment

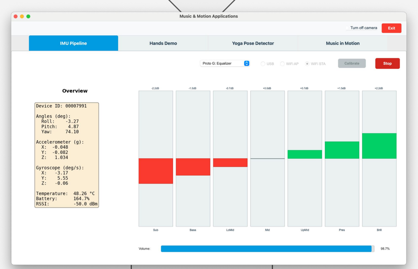
Iterative Designs



Challenge: Sensor translation; drift; latency; real-time pitch

IMU-Only

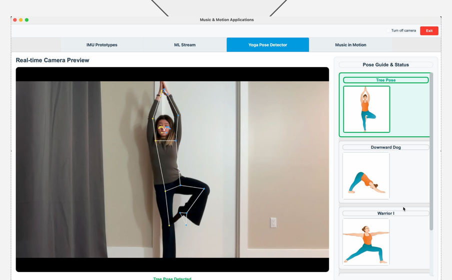
- Sensors: Roll, Pitch, Yaw, X-, Y, and Z-Accel, Large vs Precise movements
- Control: Pitch, Timbre, Volume
- Achieved: Sensor→Audio pipeline



Challenge: Not immersive; inaccurate body pose readings

Vision-Only

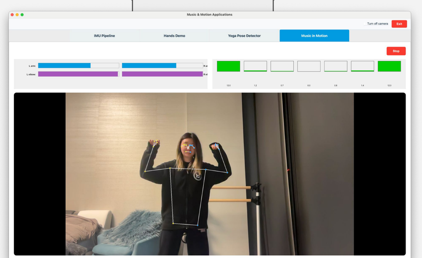
- Sensors: L/R arm height, elbow bend
- Control: Timbre, Volume, Equalizer
- Achieved: Body Movement → Audio pipeline

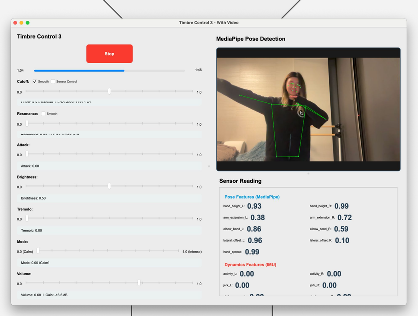


Challenge: Not expressive, no intensity of motion

Multi-Modal

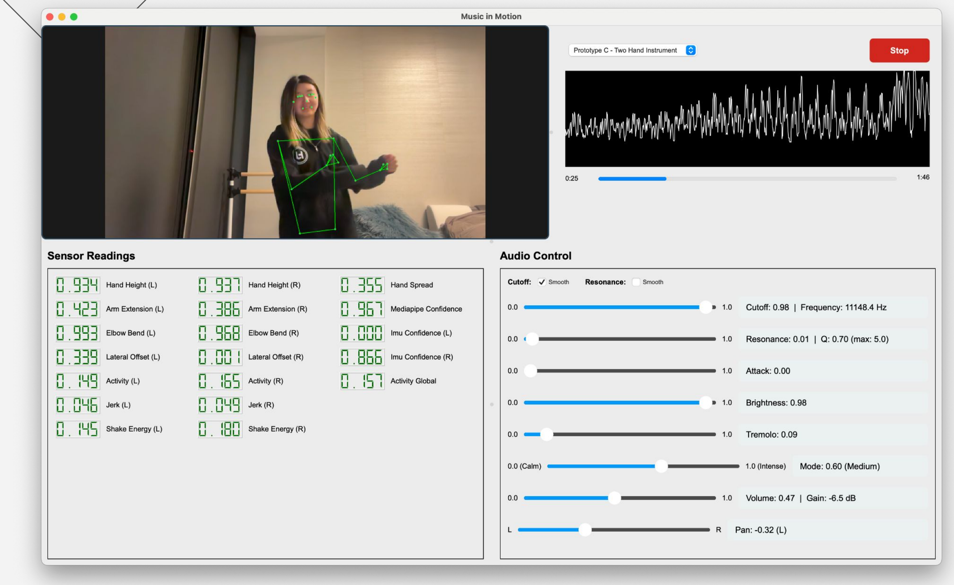
- Motion Layer combines IMU and Media Pose into single Fused Sensor
- Simplified Timbre + Volume Control
- DSP Layer for audio control





Challenge: non-intuitive to control; subpar audio quality

Final Design

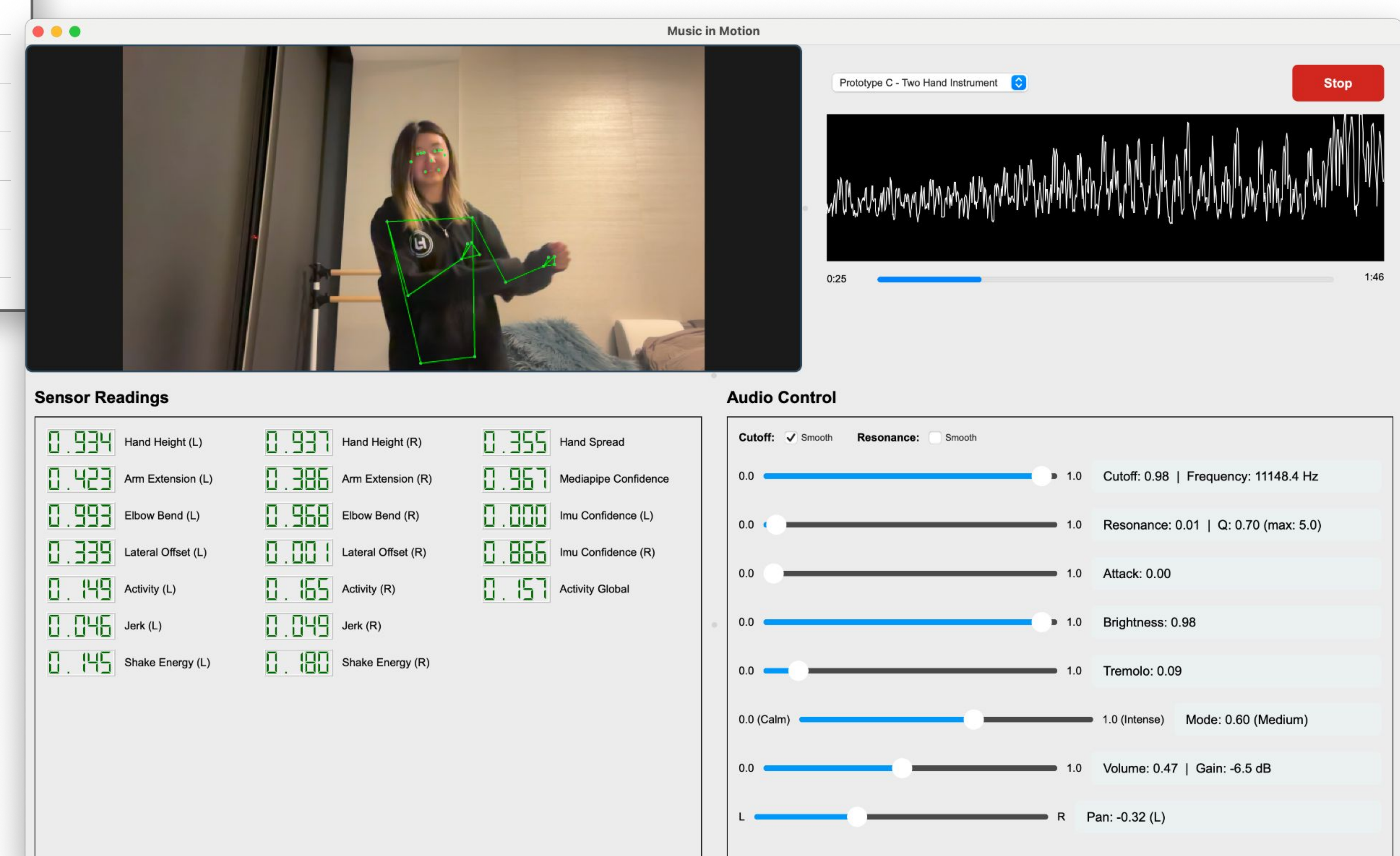


Big Idea

Can real-time sensing technologies be combined to create a continuous, responsive system that transforms human movement into sound?

Single Fused Sensor

hand_height_L, hand_height_R	Hand height relative to torso (normalized).
arm_extension_L, arm_extension_R	Arm extension (shoulder–wrist).
elbow_bend_L, elbow_bend_R	Elbow bend (shoulder–elbow–wrist)
lateral_offset_L, lateral_offset_R	Wrist lateral offset from torso center: dx = wrist.x − torso_center, then mapped to [0, 1] with 0.5=center.
hand_spread	Horizontal spread between left and right wrists (relative to shoulder width).
activity_L, activity_R	IMU activity (accel + gyro magnitude, smoothed).
jerk_L jerk_R	IMU jerk (rate of change of acceleration)
shake_energy_L shake_energy_R	IMU shake energy (bandpass-filtered accel)
activity_global	Global activity (average of left and right)



Music in *Motion*

A Real-Time Motion-Responsive Sound System to Enhance Physical Engagement

Ciana Tzuo, Horace Mann School

Big Idea

Can real-time sensing technologies be combined to create a continuous, responsive system that transforms human movement into sound?

Design Challenges

X- Full Body	encourage use of entire body in dance-like motions
X- Immersive	proprioceptive, natural movement enabling intuitive learning
X- Expressive	encourages fluid engagement across a wide range of motion dynamics
T- Real Time	low latency < 70 ms to ensure responsive smooth transitions
T- Precise	motion tracking > 85% to create strong perceptual link between action and sound
T- Stable	sustainable operations > 15 min continued use with no noticeable drift
$- Affordable	widely available consumer grade hardware and open source software
$- Convenient	operates effectively in typical indoor space, lighting

X -Experiential Goals T -Technical Constraints $ Cost Constraints

Key Components

- MacBook Air w/built in camera
- iPhone camera
- WitMotion WT901WIFI IMU (up to 4)
- Pose estimation: MediaPipe BlazePose
- Software Stack:
  - Python 3 with OpenCV, NumPy, SciPy
  - Audio: Sounddevice, libross
  - Graphical UI: PyQT5, Matplotlib
  - Image capture and processing: OpenCV

MediaPipe matplotlib python OpenCV

References

- MediaPipe Solutions guide. (n.d.). Google AI for Developers.
- WitMotion. (n.d.). WT901WIFI protocol..
- Winkler, T. (2002). Fusing Movement, Sound, and Video in Falling Up, an Interactive Dance/Theatre Production.
- Bergsland, A., etal. (2015). Composing Interactive Dance Pieces for the MotionComposer, a device for Persons with Disabilities.
- Kantan, P. R., Dahl, etal (2024). Making Movement Sonification Usable in Clinical Gait Rehab.
- Bao, Y., Zhao, etal (2022). FusePose: IMU-Vision Sensor Fusion in Kinematic Space for Parametric Human Pose Estimation.
- Fang, D.-L., Qin, Y., Chen, Q.-B., & Lu, J.-L. (2012). Real-Time Audio Synthesis In A Wireless Interactive Sensor Platform

Motion Sensors

Body Pose Sensors

Single Fused Sensor

- hand_height_L / hand_height_R
- arm_extension_L / arm_extension_R
- elbow_bend_L / elbow_bend_R
- hand_spread
- activity_L / activity_R
- jerk_L / jerk_R
- shake_energy_L / shake_energy_R
- mediapipe_confidence
- imu_confidence_L / imu_confidence_R

Audio Feature Mapping

Cutoff
Resonance
Attack
Brightness
Tremolo

Real Time Audio Transformation

IMU Screenshot

Media Pipe

Media Pipe

Key Learnings

Future Applications

- Wellness
- Dance Technology
- Rehabilitation and Physical Therapy
- Interactive media and art

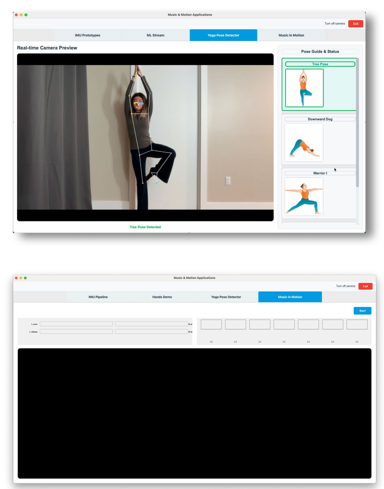
Iterative, Design Process

IMU-Only

- Sensors: Roll, Pitch, Yaw, X-, Y, and Z-Accel, Large vs Precise movements
- Control: Pitch, Timbre, Volume, Equalizer
- Achieved: Sensor→Audio pipeline
- Issues: Not immersive, better suited for hand vs full body motion

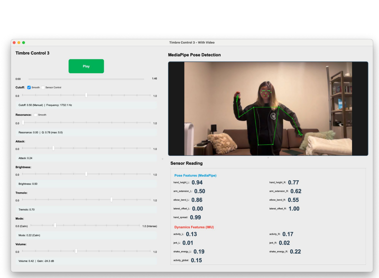
Vision-Only

- Sensors: L/R arm height, Elbow bend
- Control: Timbre, Volume, Equalizer
- Achieved: Body Movement → Audio pipeline
- Issues: Difficult to calculate intensity of motion for expressive music



Multi-Modal Fusion

- Three Layer Architecture
- Motion Layer combines IMU and Media Pose into single Fused Sensor
- Simplified Timbre + Volume Control Layer
- DSP Layer for audio control



Final Optimized Design

- IMU Preprocessing to smooth out signals
- Gravity Compensation
- Reduced audio buffer sizes to improve quality
- Fine tuning of sensor mappings to audio controls

screen shot