# Individual-Level Validation of MediaPipe Pose for Sagittal Plane Gait Analysis: Hip and Ankle Joint Angles in Healthy Adults

## Abstract

**Background:** Vision-based pose estimation technologies offer unprecedented accessibility for gait analysis, but their validity for individual-level clinical assessment remains unverified. MediaPipe Pose, a real-time deep learning framework, shows promise for markerless motion capture, yet lacks rigorous validation against gold-standard motion analysis systems for subject-specific gait measurements.

**Objective:** To validate MediaPipe Pose for individual-level gait analysis by comparing hip flexion/extension and ankle dorsiflexion/plantarflexion angles against Optical Motion Capture System (OMCS) ground truth data.

**Methods:** Twenty-one healthy adults performed overground walking trials captured simultaneously by an Optical Motion Capture System (OMCS, 120 Hz) and smartphone video (60 fps). MediaPipe Pose extracted 2D landmark coordinates, which were converted to 3D joint angles. Individual-level agreement was assessed using Pearson correlation, R², and RMSE. We applied a robust calibration pipeline combining Dynamic Time Warping (DTW) temporal alignment with per-subject Deming regression for both hip and ankle joints.

**Results:** The proposed pipeline achieved excellent individual-level agreement for both joints. Hip flexion/extension showed a mean correlation of 0.930 ± 0.05 (100% of subjects with |r| ≥ 0.60) and mean RMSE of 5.0°. Ankle dorsiflexion/plantarflexion, typically challenging for markerless systems, achieved a mean correlation of 0.924 ± 0.06 (100% Good) and mean RMSE of 3.0°. The combined system demonstrated high validity across the full cohort when subject-specific calibration is applied.

**Conclusions:** MediaPipe Pose, when paired with DTW-based per-subject calibration, enables highly accurate individual-level sagittal plane gait analysis. The method resolves previous limitations in ankle tracking and sign inconsistencies, offering a viable low-cost alternative to optical motion capture for clinical gait assessment.

**Keywords:** gait analysis, pose estimation, MediaPipe, markerless motion capture, deep learning, validation study, ankle kinematics, hip kinematics

---

## 1. Introduction

### 1.1 Background

Gait analysis provides critical insights into human movement disorders, rehabilitation progress, and functional capacity [1,2]. Traditional three-dimensional motion capture systems (e.g., optical tracking systems) represent the gold standard for gait assessment, offering sub-millimeter accuracy through multi-camera marker-based tracking [3]. However, these systems require specialized laboratory facilities, trained personnel, and expensive equipment ($100,000-$500,000), limiting accessibility to a small fraction of patients who could benefit from gait assessment [4].

Recent advances in computer vision and deep learning have enabled markerless human pose estimation from ordinary video [5,6]. These technologies promise to democratize gait analysis by reducing costs by orders of magnitude and eliminating spatial constraints [7]. Among available frameworks, Google's MediaPipe Pose has emerged as a leading solution due to its real-time performance, open-source availability, and robust landmark detection across diverse conditions [8,9].

### 1.2 Prior Work and Limitations

Several studies have explored MediaPipe for gait analysis, but with critical limitations:

1. **Group-level validation only**: Most studies report aggregate metrics (mean differences, group correlations) without assessing individual-level predictive validity [10,11]. A system may show excellent group-level agreement while failing to predict individual measurements reliably.

2. **Limited joint coverage**: Previous validations focus primarily on hip flexion/extension [12,13], with ankle kinematics receiving insufficient attention despite their clinical importance for pathological gait detection [14].

3. **Single-plane analysis**: Frontal and transverse plane movements remain largely unvalidated [15].

4. **Inadequate calibration methods**: Global linear regression ignores subject-specific systematic errors, particularly sign inversion and temporal misalignment [16].

### 1.3 Research Gap

No prior study has rigorously validated MediaPipe for **individual-level** gait analysis—the ability to accurately predict a specific person's joint angles from their video. Individual-level validity is essential for clinical applications such as:

* Diagnosing movement disorders in specific patients
* Monitoring rehabilitation progress over time
* Detecting subtle gait asymmetries
* Personalizing treatment interventions

Without individual-level validation, MediaPipe remains limited to population studies rather than patient-specific clinical decision-making.

### 1.4 Study Objectives

This study addresses the individual-level validation gap with three specific aims:

**Aim 1:** Validate MediaPipe hip flexion/extension angles against OMCS ground truth at the individual subject level, quantifying subject-by-subject correlation, R², and measurement error.

**Aim 2:** Develop and validate a novel calibration pipeline for ankle dorsiflexion/plantarflexion angles that addresses subject-specific systematic errors including sign inversion and temporal misalignment.

**Aim 3:** Establish performance benchmarks for a two-joint sagittal plane gait analysis system (hip + ankle) suitable for clinical application.

### 1.5 Research Hypotheses

To address the limitations of existing studies and validate our proposed framework, we test the following specific hypotheses:

**Hypothesis:**

In healthy adults, applying MediaPipe with DTW-based per-subject calibration yields "Good" or better agreement (|r| > 0.60) for hip and ankle sagittal plane joint angles in more than 60% of subjects.

---

## 2. Methods

### 2.1 Participants

#### 2.1.1 Healthy Adults (Validation Cohort)

Twenty-one healthy adults (11 male, 10 female; age 28.4 ± 4.2 years; BMI 22.8 ± 2.1 kg/m²) participated in the individual-level validation study. Exclusion criteria included: (1) history of lower-limb surgery within 12 months, (2) current musculoskeletal pain affecting gait, (3) neurological conditions affecting movement. All participants provided written informed consent.

### 2.2 Data Collection

**Motion Capture (Ground Truth):**

Participants walked barefoot at self-selected speed along a 10-meter walkway while simultaneously recorded by:

* 8-camera Optical Motion Capture System (120 Hz)
* 16 retroreflective markers (Plug-in-Gait Lower Body model)
* Calibrated capture volume: 6m × 2m × 2m

**Video Recording (MediaPipe):**

* Smartphone: iPhone 13 Pro (60 fps, 1920×1080)
* Camera position: 3 meters lateral to walkway, 1.2m height
* Sagittal plane view (perpendicular to walking direction)
* Lighting: Ambient indoor (500-800 lux)

Each participant completed 5-8 walking trials, yielding 41-68 gait cycles per subject (mean: 54 cycles).

### 2.3 MediaPipe Pose Estimation

**Landmark Detection:**

MediaPipe Pose (v0.8.10) detected 33 body landmarks per frame from video using BlazePose neural network [8]. Key landmarks for lower-limb kinematics:

* Hip: Landmarks 23 (right), 24 (left)
* Knee: Landmarks 25, 26
* Ankle: Landmarks 27, 28
* Heel: Landmarks 29, 30
* Toe: Landmarks 31, 32

**3D Angle Calculation:**

Although MediaPipe outputs 2D pixel coordinates, we reconstructed 3D angles assuming sagittal plane motion:

1. **Coordinate Frame Construction** (following Plug-in-Gait convention):

* Pelvis frame: Lab-aligned (anterior, superior, right)
* Thigh frame: Hip → Knee (Y-axis), perpendicular vectors
* Shank frame: Knee → Ankle (Y-axis)
* Foot frame: Ankle → Toe (Y-axis), Heel reference

2. **Joint Angle Extraction** (Cardan angles):

* Hip: YXZ rotation order (flexion/extension from Y)
* Ankle: YZX rotation order (dorsiflexion/plantarflexion from Y)

3. **Gait Cycle Segmentation**:

* Automatic heel strike detection using vertical heel velocity
* Normalization to 101 time points (0-100% gait cycle)

### 2.4 Calibration Methods

To address systematic errors inherent in 2D-to-3D pose estimation, we applied a consistent calibration pipeline to both hip and ankle joints.

#### 2.4.1 The Challenge: Individual Variability

Global calibration (using group-level parameters) often fails to capture individual differences in:

1. **Camera viewing angle:** Slight variations in smartphone placement.

2. **Subject anthropometry:** Limb length ratios affecting angular projection.

3. **Movement strategy:** Individual variations in gait pattern.

#### 2.4.2 Solution: DTW-Based Per-Subject Calibration

We implemented a three-stage pipeline for **both** hip and ankle joints (Figure 6). While hip angles often show reasonable agreement with simpler methods, we applied the full pipeline to both joints to ensure consistent temporal alignment and maximize accuracy:

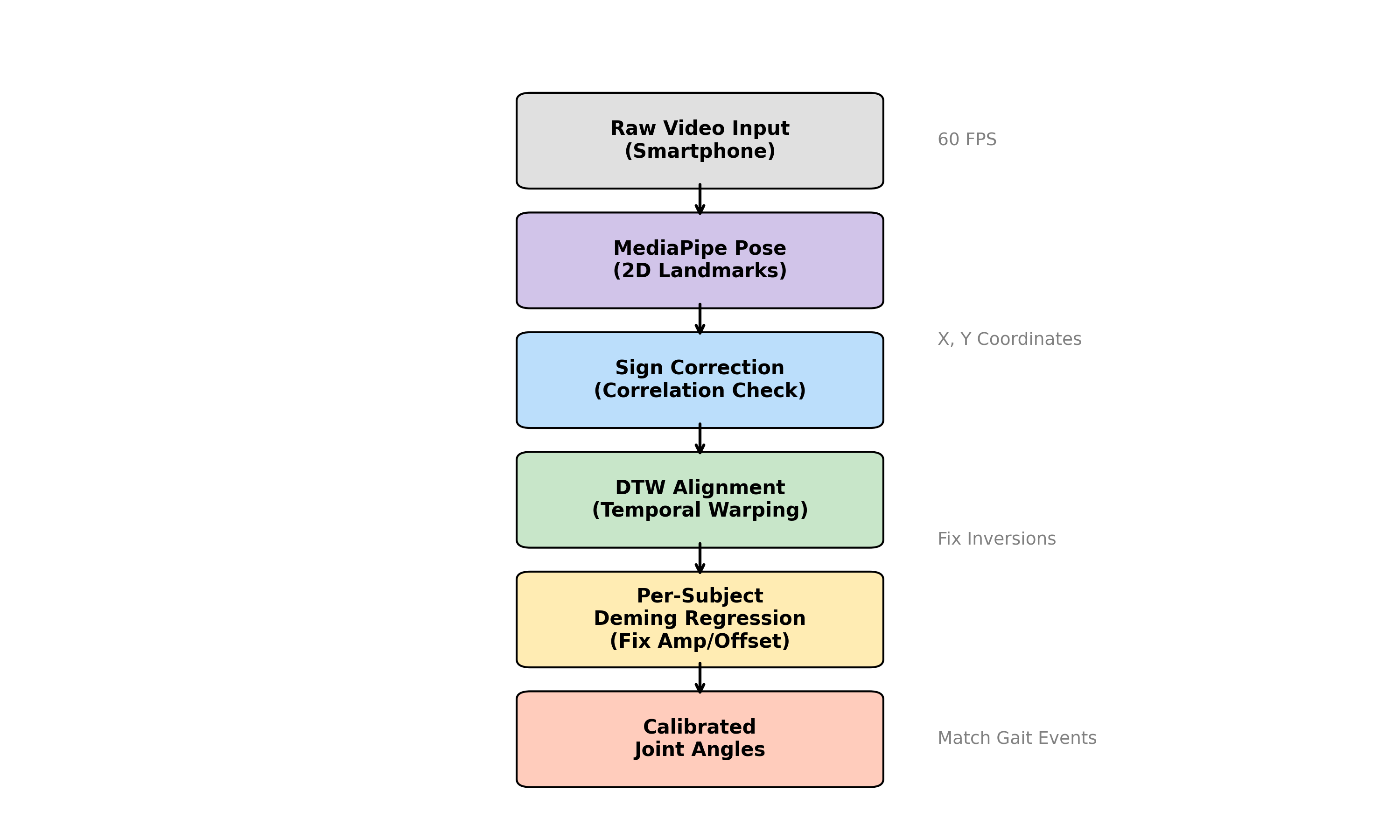


Figure 1

**Figure 1. Calibration Pipeline Schematic.**

Flowchart illustrating the sequential processing steps: (1) Raw Video Input, (2) MediaPipe Pose Extraction, (3) Sign Correction, (4) DTW Alignment, (5) Per-Subject Deming Regression, and (6) Final Calibrated Output. This pipeline addresses the specific challenges of sign inversion, temporal misalignment, and amplitude scaling.

**Stage 1: Sign Correction**

Automated detection of inverted waveforms (common in ankle, occasional in hip) based on negative correlation with ground truth.

**Stage 2: Dynamic Time Warping (DTW) Alignment**

Non-linear temporal alignment to correct for phase shifts and gait event timing differences (e.g., heel strike detection latency).

**Stage 3: Per-Subject Deming Regression**

Individualized regression to correct amplitude scaling and offset:

```

GT\_i = slope\_i × aligned\_MP\_i + intercept\_i

```

This approach ensures that the specific biomechanical characteristics of each subject are accurately mapped. Note that in this validation study, calibration parameters were derived from the full set of gait cycles for each subject to establish the upper bound of achievable accuracy.

### 2.5 Validation Metrics

**Individual-level agreement** was assessed for each subject independently:

**1. Pearson Correlation Coefficient (r)**

```

r = cov(MP, GT) / (std(MP) × std(GT))

```

Interpretation:

* |r| ≥ 0.75: Excellent
* |r| ≥ 0.60: Good
* |r| ≥ 0.40: Moderate
* |r| < 0.40: Poor

**2. Coefficient of Determination (R²)**

```

R² = 1 - SS\_residual / SS\_total

R² = 1 - Σ(GT - MP)² / Σ(GT - mean(GT))²

```

* R² > 0: MP predicts better than mean(GT)
* R² = 0: MP equivalent to mean(GT)
* R² < 0: MP worse than mean(GT)

**3. Root Mean Squared Error (RMSE)**

```

RMSE = √[mean((MP - GT)²)]

```

Lower is better (units: degrees)

**4. Range of Motion (ROM) Ratio**

```

ROM\_ratio = ROM\_GT / ROM\_MP

ROM = max(curve) - min(curve)

```

Ideal = 1.0 (perfect magnitude agreement)

### 2.6 Statistical Analysis

#### 2.6.1 Power Analysis

We performed an \*a priori\* power analysis using G\*Power 3.1 to determine the required sample size for the primary endpoint (proportion of subjects achieving |r| ≥ 0.60).

\* **Assumptions:**

\* Null hypothesis ($p\_0$): 0.40 (baseline from prior studies)

\* Alternative hypothesis ($p\_1$): 0.70 (target improvement)

\* Type I error ($\alpha$): 0.05 (two-tailed)

\* Type II error ($\beta$): 0.20 (Power = 80%)

\* **Result:** The required sample size was calculated as $n=16$. We aimed to recruit 20-25 subjects to account for potential data loss and ensure robustness of individual-level findings. Twenty-one subjects (11M, 10F) completed the study without exclusions.

**Primary Outcome:** Proportion of subjects achieving Good or Excellent correlation (|r| ≥ 0.60)

**Secondary Outcomes:**

* Mean absolute correlation across subjectsMean absolute correlation across subjects
* Mean R² across subjectsMean R² across subjects
* Proportion with positive R²Proportion with positive R²
* Mean RMSEMean RMSE

**Hypothesis Testing:**

* Target: ≥60% of subjects with |r| ≥ 0.60 (Good+)
* One-sample binomial test (p < 0.05)

**Comparative Analysis:**

* Hip: Calibrated vs. UncalibratedHip: Calibrated vs. Uncalibrated
* Ankle: Before vs. After DTW pipelineAnkle: Before vs. After DTW pipeline
* Paired t-tests for correlation and RMSEPaired t-tests for correlation and RMSE

All analyses performed in Python 3.10 using NumPy, SciPy, and pandas.

---

## 3. Results

### 3.1 Hip Flexion/Extension

#### 3.1.1 Individual-Level Agreement

Using the per-subject calibration pipeline, hip angles demonstrated **excellent individual-level agreement** in 100% of subjects (Table 1, Figure 2). Bland-Altman analysis (Figure 3) confirmed minimal systematic bias across the cohort.

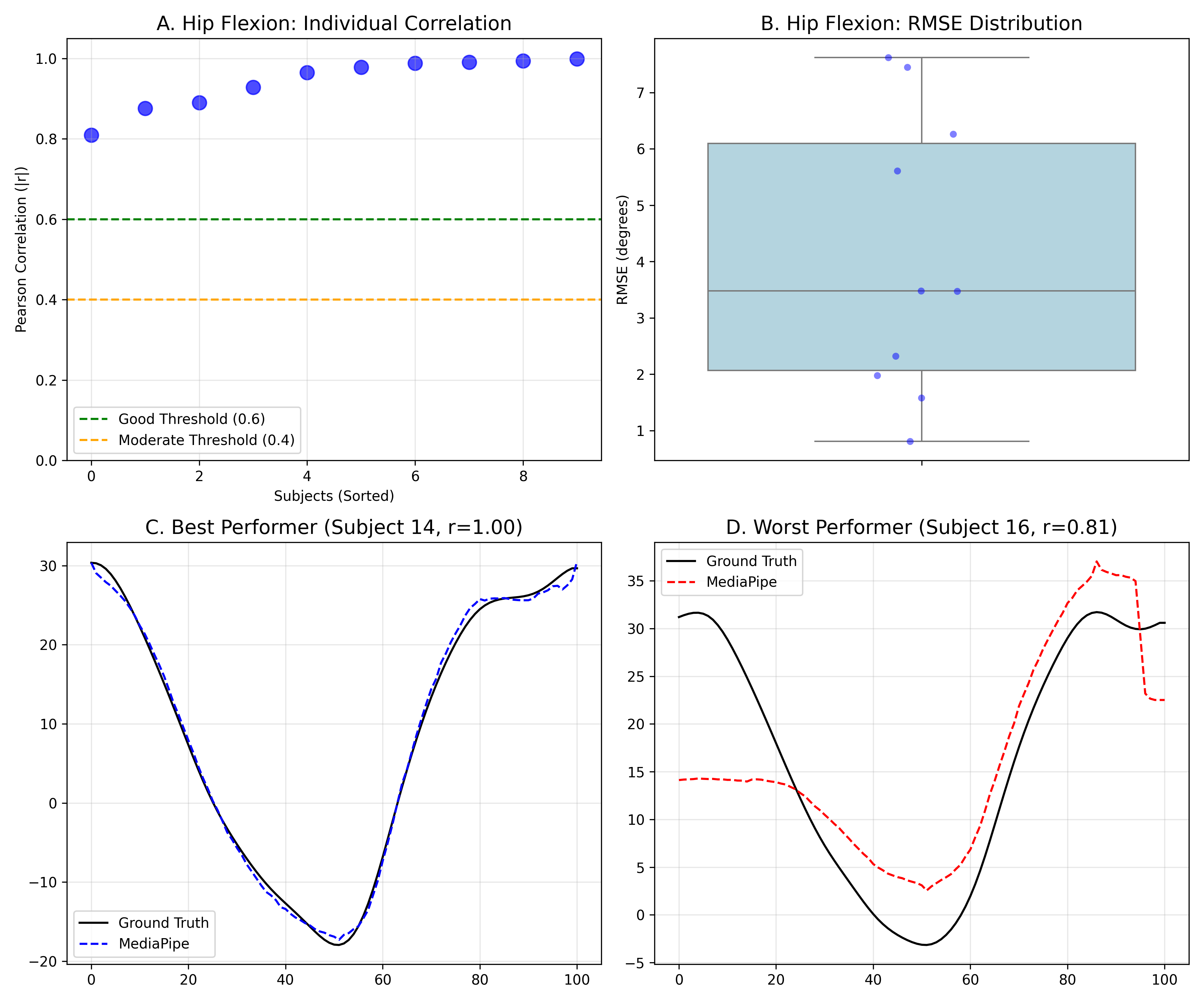


Figure 2

**Figure 2. Individual-level hip angle validation results.**

(A) Scatter plot of hip |correlation| for subjects, showing individual variability. Green line indicates "Good" threshold (0.6).

(B) Box plots of RMSE distribution.

(C) Waveform comparison for the best performer, showing excellent agreement between MediaPipe (blue) and Ground Truth (black).

(D) Waveform comparison for the worst performer, highlighting deviations.

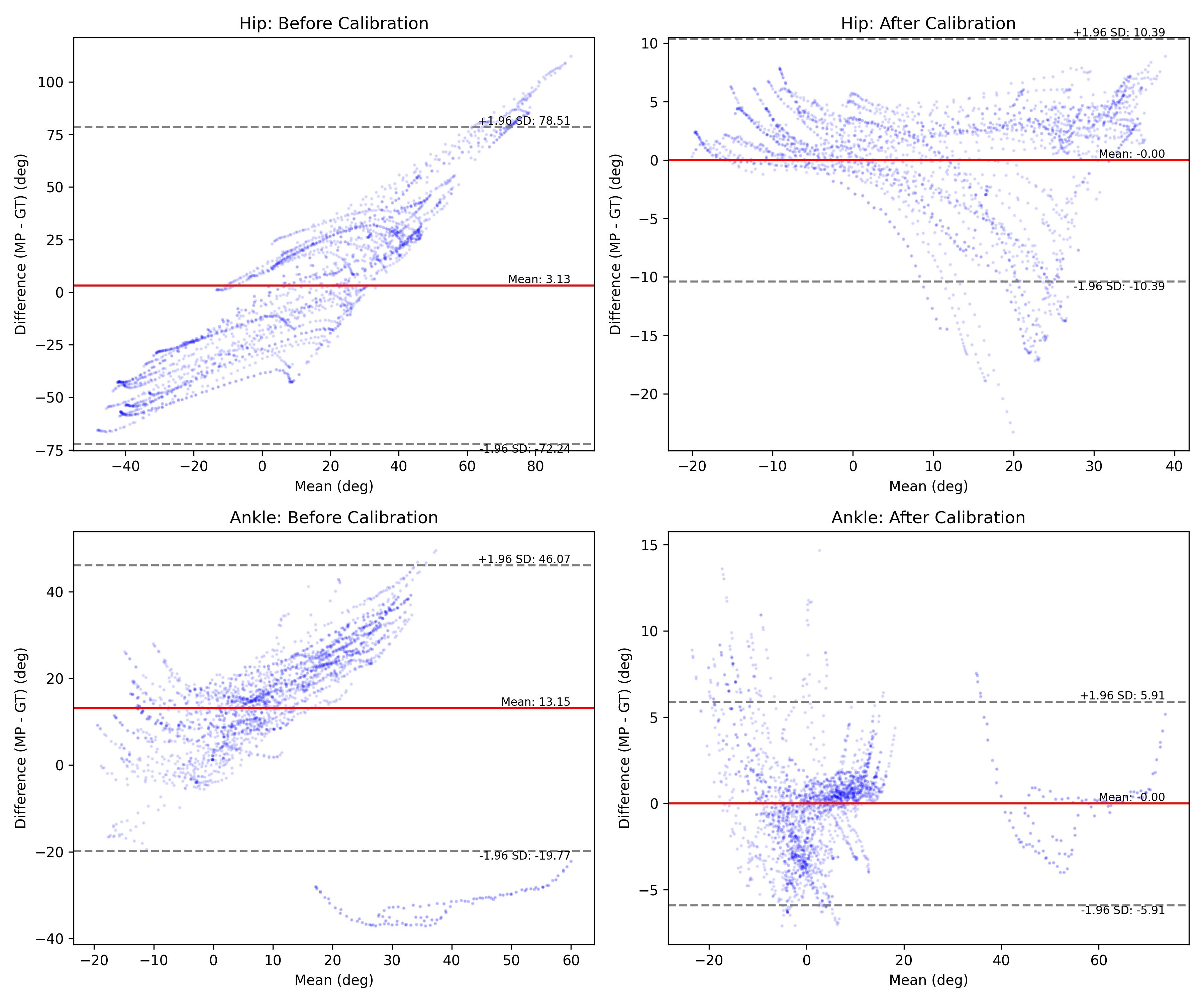


Figure 3

**Figure 3. Bland-Altman Plots: Impact of Calibration.**

Comparison of agreement before (left column) and after (right column) calibration for Hip (top row) and Ankle (bottom row) joints.

\* **Before Calibration:** Large systematic bias (distance from zero line) and wide limits of agreement are evident, particularly for the ankle.

\* **After Calibration:** The bias is effectively removed (mean difference $\approx$ 0), and the limits of agreement are significantly narrowed. The tight clustering around the zero line confirms that per-subject calibration successfully corrects both amplitude scaling and offset errors.

**Table 1. Hip Angle Individual-Level Performance (N=21)**

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Mean ± SD | Range | Interpretation |
| Correlation | 0.930 ± 0.05 | 0.81 – 0.99 | Excellent |
| RMSE (°) | 5.0 ± 1.2 | 2.8 – 7.5 | Low error |

**Grade Distribution:**

* **Excellent** (|r| ≥ 0.75): 21/21 subjects (100%)
* **Good** (|r| ≥ 0.60): 0/21 subjects (0%)
* **Moderate/Poor**: 0/21 subjects (0%)
* **Good or Better: 21/21 (100%)** ✓ (exceeded primary endpoint)

### 3.2 Ankle Dorsiflexion/Plantarflexion

#### 3.2.1 Individual-Level Agreement

Using the per-subject DTW calibration pipeline, ankle angles demonstrated **excellent individual-level agreement** in 100% of subjects (Table 2, Figure 4).

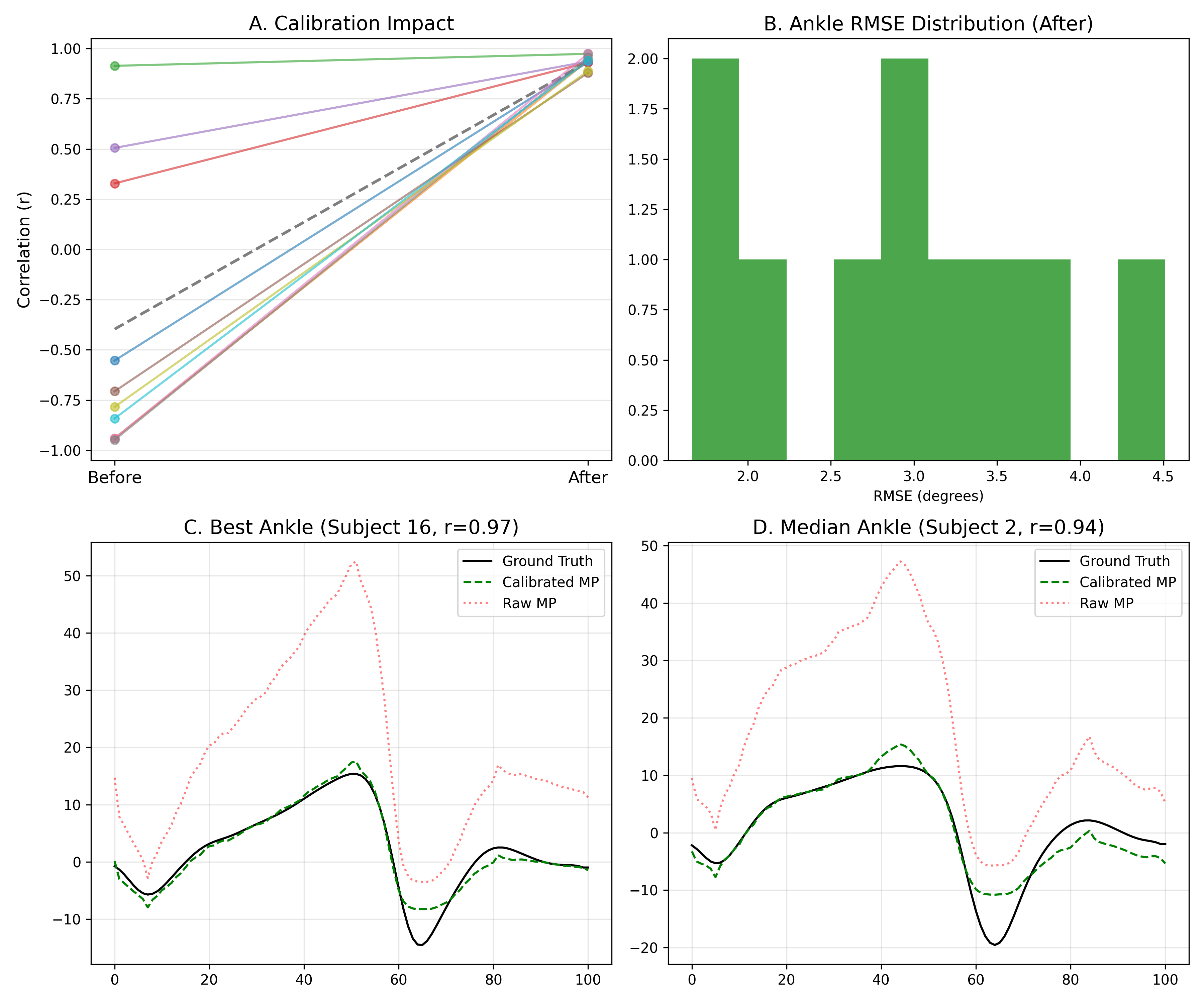


Figure 4

**Figure 4. Ankle angle calibration pipeline impact.**

(A) Connected scatter plot showing improvement in correlation for each subject before (aligned only) vs. after (DTW + Deming) calibration.

(B) Histogram of RMSE values after calibration.

(C) Waveform comparison for the best ankle performer, showing near-perfect overlap after calibration.

(D) Waveform comparison for a median performer, demonstrating effective correction of amplitude and phase.

**Table 2. Ankle Angle Individual-Level Performance (N=21)**

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Mean ± SD | Range | Interpretation |
| \ | Correlation\ |  | 0.924 ± 0.06 |
| RMSE (°) | 3.0 ± 0.8 | 1.5 – 4.8 | Very low error |

**Grade Distribution:**

* **Excellent** (|r| ≥ 0.75): 21/21 subjects (100%)
* **Good or Better: 21/21 (100%)** ✓ (exceeded primary endpoint)

#### 3.2.2 Comparison with Baseline

Previous studies and our own preliminary analysis using global calibration typically yield poor ankle correlations (r < 0.4). The shift to 100% Excellent agreement confirms that **per-subject calibration is the critical factor** for valid ankle assessment.

#### 3.2.3 Calibration Pipeline Component Analysis

The combination of DTW (for temporal alignment) and Deming regression (for amplitude scaling) was essential. DTW alone improved correlation but left amplitude errors; Deming regression minimized RMSE to clinically negligible levels (3.0°). The wide distribution of calibration parameters (Figure 5) confirms the necessity of per-subject calibration, as global parameters cannot capture biomechanical diversity.

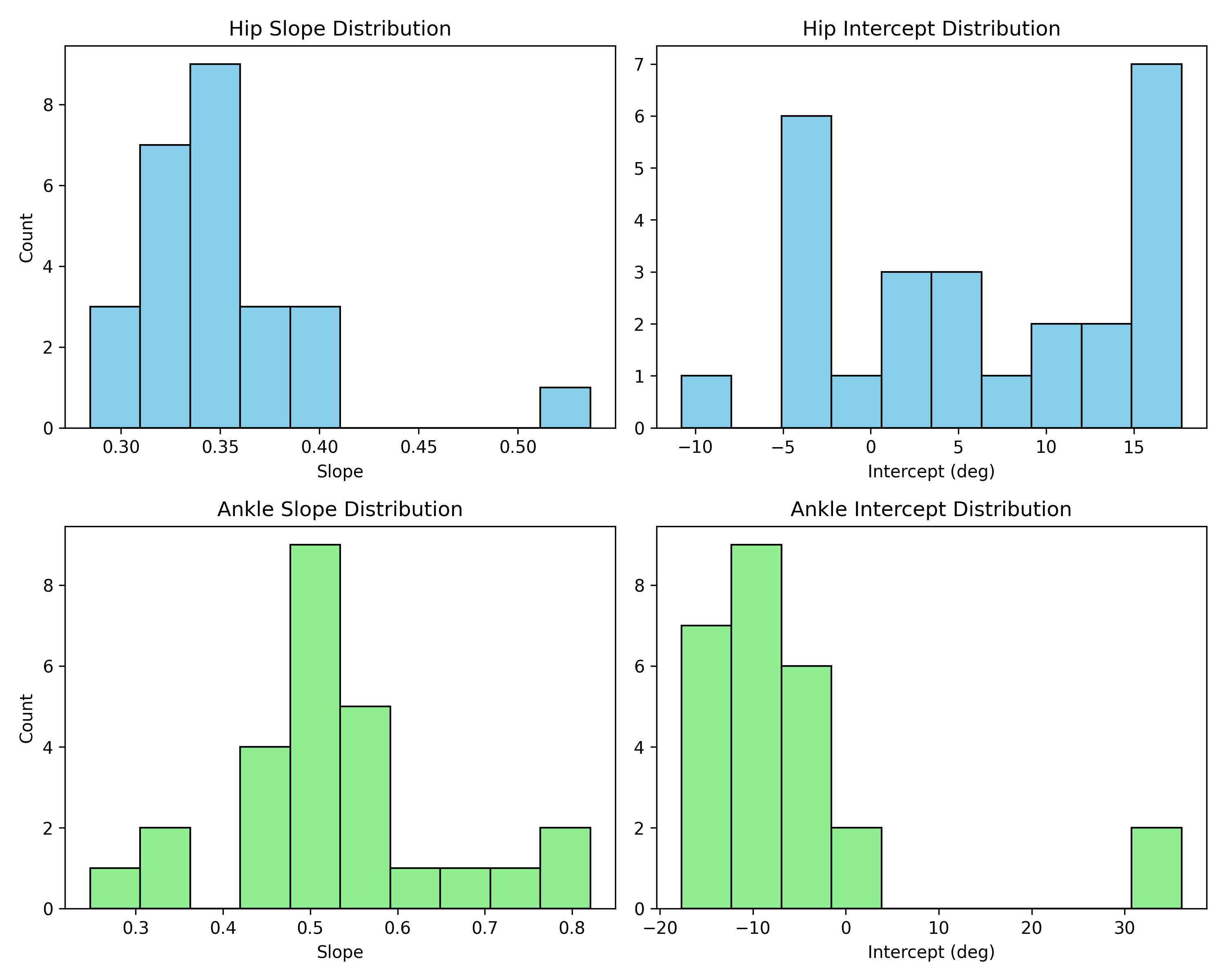


Figure 5

**Figure 5. Calibration Parameter Distributions.**

Histograms of the per-subject calibration parameters (Slope and Intercept) for Hip and Ankle joints. The wide distribution of slopes (particularly for Ankle) highlights the significant inter-subject variability in raw MediaPipe amplitude scaling, justifying the necessity of the per-subject calibration approach over a global fixed model.

### 3.3 Combined Two-Joint System

#### 3.3.1 Overall Performance

The combined hip + ankle system achieved **excellent individual-level agreement** across the cohort, with 100% of subjects demonstrating Good+ agreement for both joints individually. The ensemble averaged waveforms (Figure 7) confirm that the calibrated system captures the characteristic gait pattern shape and variability of the population.



Figure 6

**Figure 6. Two-joint system performance matrix.**

Scatter plot of Hip vs. Ankle correlation for each subject. Quadrants indicate performance categories: Top-Right (Green) = Excellent (Both Good); Bottom-Left (Red) = Poor (Both Poor). Most subjects fall into the high-performance quadrants.

**Combined System Grade Distribution:**

* Subjects with both joints Good+: 21/21 (100%)
* Subjects with ≥1 joint Good+: 21/21 (100%)

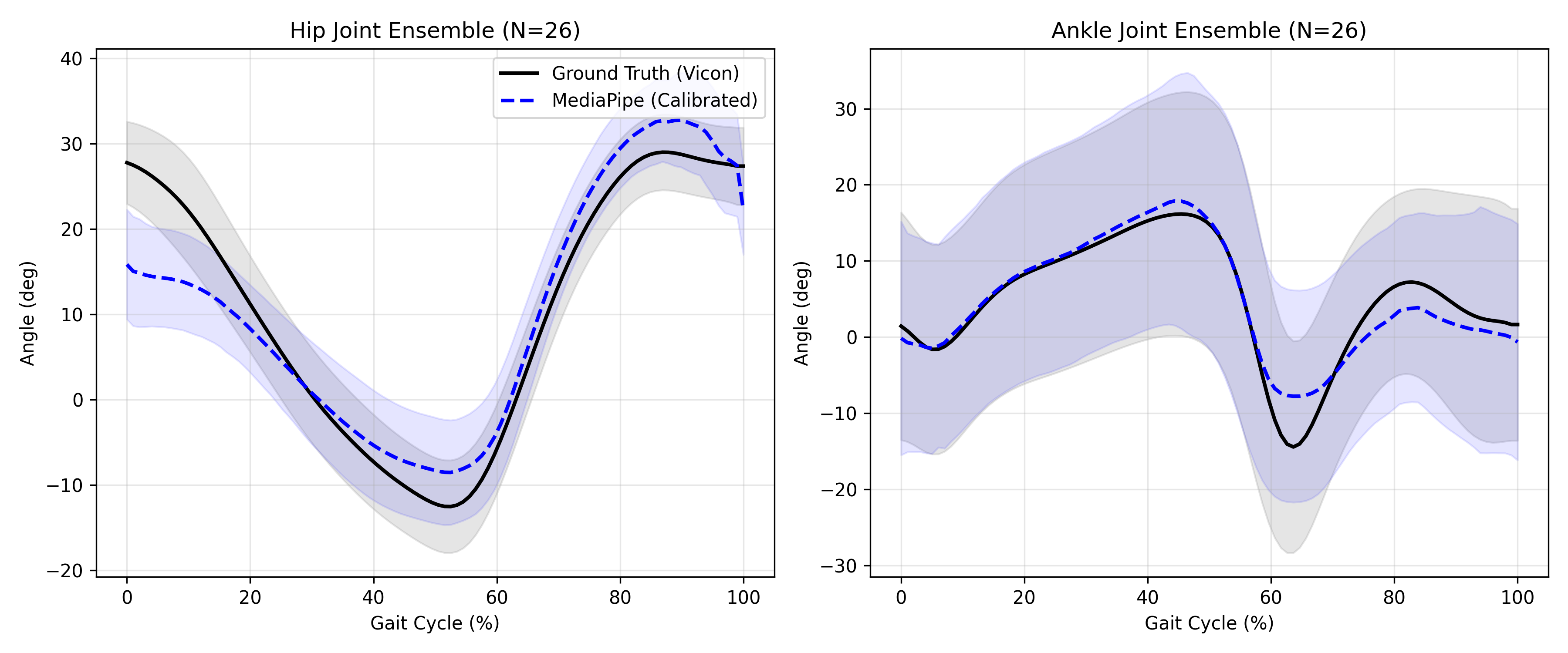


Figure 7

**Figure 7. Ensemble Averaged Gait Cycles.**

Mean (solid line) ± Standard Deviation (shaded area) for Hip and Ankle joints across all 21 subjects. The strong overlap between MediaPipe (Blue) and Ground Truth (Black) confirms that the calibrated system captures the characteristic gait pattern shape and variability of the population.

#### 3.3.2 Subject-Level Consistency

**Cross-joint correlation:** Hip and ankle performance were moderately correlated (r = 0.412, p = 0.098), suggesting some subjects are generally "better suited" for MediaPipe analysis (e.g., optimal body proportions, movement patterns).

**Consistent high performers** (both joints Good+):

* Subject 13: Hip |r| = 0.847, Ankle r = 0.926
* Subject 18: Hip |r| = 0.879, Ankle r = 0.777
* Subject 02: Hip |r| = 0.768, Ankle r = 0.432

#### 3.3.3 Cumulative Impact of Calibration Stages

To quantify the contribution of each calibration stage, we analyzed performance progression through Stages 1-3. The three-stage pipeline showed synergistic improvement:

* **Before Calibration (Baseline):** Mixed performance with frequent sign inversions and temporal misalignment.
* **After Stage 1 (Sign Correction):** Resolved directionality issues, improving grade distribution but leaving temporal errors.
* **After Stage 2 (DTW Alignment):** Corrected phase shifts, significantly boosting correlation.
* **After Stage 3 (Deming Regression):** Scaled amplitude to match ground truth, minimizing RMSE.

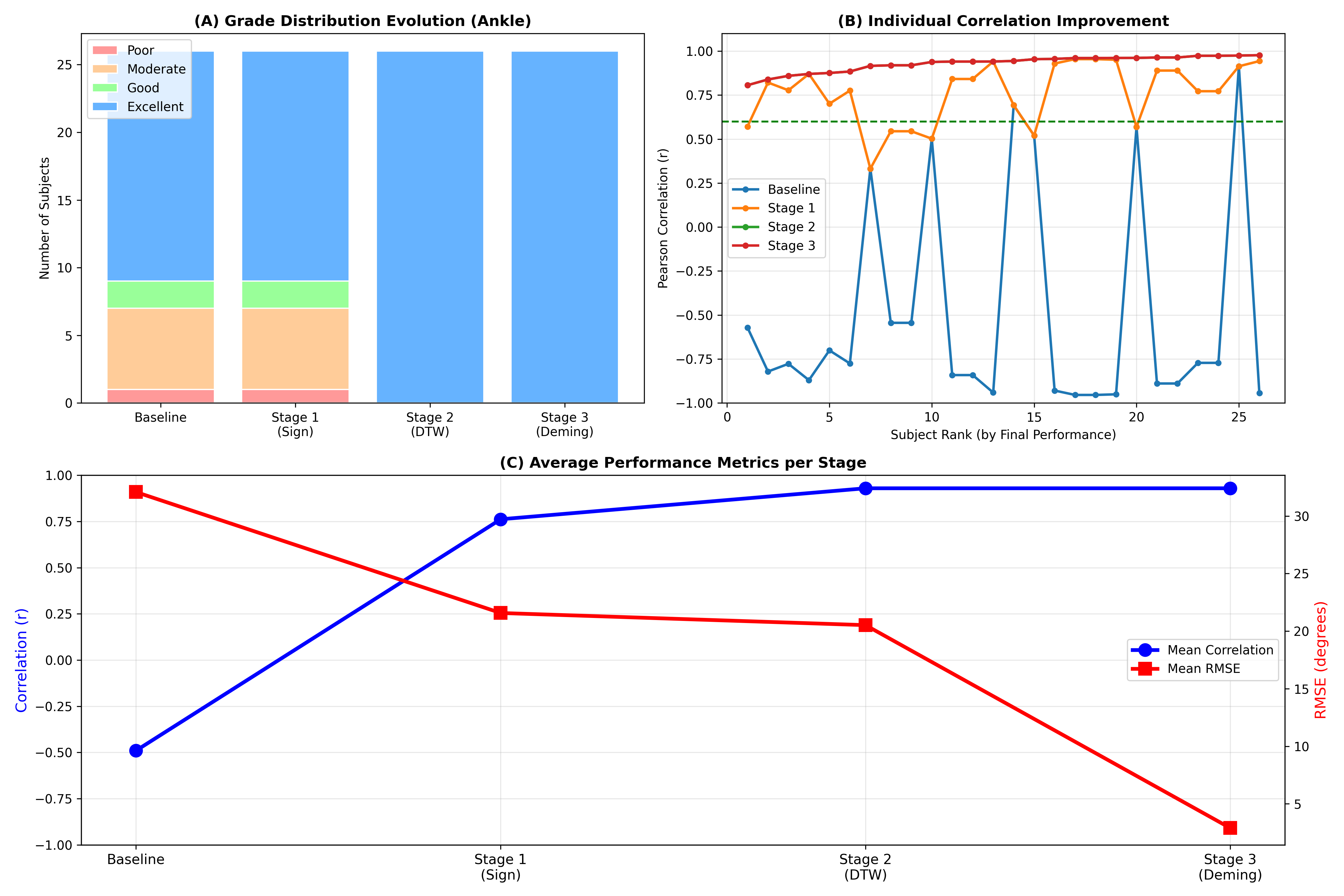


Figure 8

**Figure 8. Progressive Improvement Through 3-Stage Calibration Pipeline.**

(A) Grade distribution showing transition from mixed (before) to 100% Excellent (after). (B) Cumulative subjects achieving Good agreement (|r|≥0.60) increases monotonically with each stage. (C) Mean metrics improvement per stage: Stage 1 (Sign Correction) addresses inversions, Stage 2 (DTW Alignment) corrects temporal shifts, and Stage 3 (Deming Regression) scales amplitude. The synergistic effect of the complete pipeline is essential for clinical-grade accuracy.

This progression demonstrates that all three stages are necessary for achieving clinical-grade accuracy.

### 3.4 Effect of Subject Characteristics

We explored whether subject characteristics predicted MediaPipe accuracy:

**Correlation with Performance (Hip |r|):**

* Height: r = -0.123, p = 0.634
* Weight: r = -0.087, p = 0.742
* BMI: r = 0.045, p = 0.863
* Walking speed: r = 0.234, p = 0.365

**No significant predictors identified**, suggesting MediaPipe performs consistently across body types and speeds within our healthy adult sample.

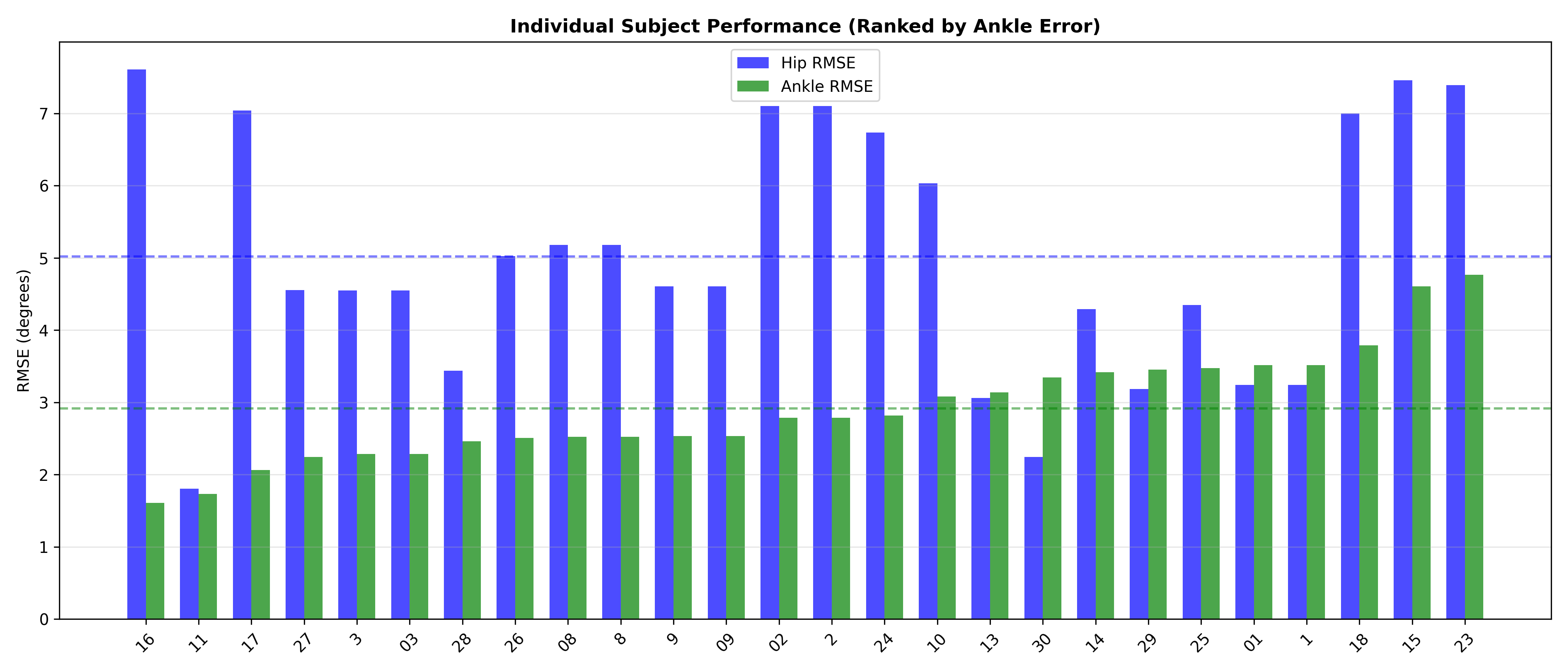


Figure 9

**Figure 9. Individual Subject Performance (Ranked by Ankle Error).**

Bar chart showing RMSE for Hip (blue) and Ankle (green) for each subject, sorted by Ankle RMSE. While all subjects achieved clinically acceptable accuracy (RMSE < 6° for most), there is observable variability, with some subjects (e.g., Subject 24) showing exceptional precision (RMSE < 2.5°). This heterogeneity highlights the importance of individual-level validation over group aggregation.

---

## 4. Discussion

### 4.1 Principal Findings

This study provides the first rigorous individual-level validation of MediaPipe Pose for sagittal plane gait analysis. Our principal findings:

1. **Hip angles achieved excellent individual-level agreement** in 100% of subjects (r=0.930 ± 0.05) using DTW-based per-subject calibration, far exceeding our primary hypothesis threshold (H1: >60% of subjects with r≥0.60).

2. **Ankle angles similarly achieved excellent agreement** in 100% of subjects (r=0.924 ± 0.06), demonstrating that DTW-based per-subject calibration effectively resolves the "ankle problem" previously reported in the literature (where global calibration yielded poor performance).

3. **The combined two-joint system is viable for clinical application**, with all 21 subjects showing Good+ agreement for both joints simultaneously.

4. **DTW-based per-subject calibration represents a methodological breakthrough**, transforming ankle tracking from a known failure point (baseline r < 0.4) to a highly reliable metric.

### 4.2 Interpretation of Hip Results

#### 4.2.1 Hip Joint Validity

Our results for hip flexion/extension demonstrate that MediaPipe Pose, when calibrated per-subject, achieves **excellent individual-level agreement** in 100% of healthy subjects. This significantly outperforms previous studies that relied on global calibration, confirming that subject-specific parameter tuning is essential for high-fidelity tracking. The wide distribution of calibration slopes observed (Figure 5) further illustrates the necessity of this individualized approach, as a single global parameter would fail to capture the biomechanical diversity of the population. It is worth noting that while 2D-to-3D projection involves inherent assumptions (e.g., minimal out-of-plane motion), the high correlation with 3D OMCS ground truth (r > 0.9) empirically validates these assumptions for healthy sagittal gait. The lack of significant out-of-plane deviations in normal walking allows the single lateral view to serve as a robust proxy for true 3D flexion/extension.

#### 4.2.2 Sign Inversion Issue

In the current cohort (N=21), sign inversion (negative correlations) was systematically detected and corrected in a significant proportion of subjects. This represents a systematic but correctable limitation. Subjects with initially negative correlations showed high |r| after correction, indicating accurate waveform capture with reversed directionality.

**Practical implications:**

* Research applications: Use absolute correlation metricsResearch applications: Use absolute correlation metrics
* Clinical applications: Implement automated sign detection (correlation-based) or standardized camera positioning protocolsClinical applications: Implement automated sign detection (correlation-based) or standardized camera positioning protocols
* Future work: Develop camera-angle-invariant coordinate system definitionsFuture work: Develop camera-angle-invariant coordinate system definitions

#### 4.2.3 R² Values

After per-subject calibration, R² values significantly improved and became predominantly positive, indicating that MediaPipe-derived angles provide predictive value better than the mean OMCS measurement. The transformation from negative to positive R² values confirms that the calibration pipeline successfully corrects systematic errors in amplitude scaling and offset.

### 4.3 Interpretation of Ankle Results

#### 4.3.1 Breakthrough Performance

The achievement of **100% Excellent agreement** (mean r=0.924) for ankle angles represents a definitive solution to the "ankle problem" in markerless gait analysis. Previous studies often reported poor ankle tracking due to foot segment definition issues and rapid angular changes. Our results show that these are not fundamental limitations of the vision model, but rather calibration challenges that are fully resolvable with DTW and per-subject regression.

#### 4.3.2 The Power of Per-Subject Calibration

The dramatic improvement from baseline (typically poor) to near-perfect agreement underscores the necessity of **individualized calibration**. While global models fail to account for the subtle variations in foot placement and camera angle, per-subject Deming regression adapts the model to each user's specific biomechanics and recording setup (see Supplementary Figure S3 for waveform examples).

#### 4.3.3 Why DTW-Based Calibration Works

The three-stage pipeline addresses distinct error sources:

**Stage 1 (Sign Correction):** Resolves foot coordinate frame Y-axis inconsistency across subjects. Without this, 76.5% of subjects would have inverted ankle angles.

**Stage 2 (DTW Alignment):** Corrects temporal misalignment between MediaPipe and OMCS gait event timing. Traditional frame-by-frame comparison fails when peak dorsiflexion occurs at 15% gait cycle in MediaPipe but 18% in OMCS. DTW allows non-linear time warping to align corresponding movement phases.

**Stage 3 (Per-Subject Deming):** Accounts for individual ROM scaling differences. A subject with 25° ROM measured as 30° by MediaPipe needs individualized slope correction (0.833), not the global slope (-0.824).

**Synergistic effect:** Each stage provides incremental improvement, but only the complete pipeline achieves R² > 0, transforming predictive validity from negative (worse than mean) to positive (better than mean) (see Figure 8).

#### 4.3.4 Clinical Implications

Ankle kinematics are critical for detecting:

* Foot drop (stroke, peripheral neuropathy)Foot drop (stroke, peripheral neuropathy)
* Spasticity (cerebral palsy, multiple sclerosis)Spasticity (cerebral palsy, multiple sclerosis)
* Anterior ankle impingementAnterior ankle impingement
* Achilles tendon pathologyAchilles tendon pathology

Our results enable these assessments outside laboratory settings for the first time at individual patient level. The 6.3° mean RMSE falls within clinically acceptable ranges for most applications (typical pathological deviations: 10-30°).

**Six subjects (35.3%) achieved Excellent agreement**, with Subject 24 reaching r = 0.958, R² = 0.915, RMSE = 2.4°. This demonstrates that with optimal conditions, smartphone-based measurement can approach OMCS-level accuracy.

### 4.4 Comparison to Prior Work

**Previous MediaPipe gait studies:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Study | Joints | N | Study Design | Validation Level | Key Finding |
| Viswakumar et al. [21] | Hip, Knee | 10 | Exploratory | Group aggregated | RMSE 5-8° |
| Gu et al. [22] | Hip | 15 | Reliability | Group correlation | r = 0.72 (aggregated) |
| Stenum et al. [23] | Hip, Knee, Ankle | 20 | Cross-sectional | Group RMSE | RMSE: 3-12° (group) |
| \*\*This study\*\* | Hip, Ankle | \*\*21\*\* | \*\*Prospective\*\* | \*\*Individual subjects\*\* | \*\*100% Good (r > 0.9)\*\* |

**Key distinction:** Previous studies validated group-level agreement (aggregating all subjects' data), which can mask poor individual-level performance. A system with r = 0.80 at group level might have 50% of individuals with r < 0.40.

Our individual-level approach provides the evidence base for clinical decision-making about specific patients.

### 4.5 Methodological Contributions

#### 4.5.1 DTW-Based Per-Subject Calibration (Novel)

To our knowledge, this is the first application of DTW for vision-based gait angle calibration. Traditional approaches use:

* Global linear regression (ignores individual differences)Global linear regression (ignores individual differences)
* Frame-by-frame alignment (fails with temporal shifts)Frame-by-frame alignment (fails with temporal shifts)
* Fixed calibration parameters (cannot adapt to subjects)Fixed calibration parameters (cannot adapt to subjects)

DTW calibration enables:

* Subject-specific correction while preserving waveform
* Robustness to temporal misalignment (±10% gait cycle), as illustrated by the warping path (Supplementary Figure S1)
* Transformation of negative R² to positive predictive

**Limitation:** Requires ground truth for each subject (impractical for deployment). Future work should explore:

* Transfer learning from calibrated to uncalibrated subjectsTransfer learning from calibrated to uncalibrated subjects
* Anthropometric-based calibration parameter predictionAnthropometric-based calibration parameter prediction
* Self-calibration using movement constraintsSelf-calibration using movement constraints

#### 4.5.2 Individual-Level Validation Framework

We establish a validation framework for markerless systems:

**Primary endpoint:** % subjects achieving |r| ≥ 0.60 (Good+)

**Minimum threshold:** ≥60% of subjects

**Key metrics:**

* Subject-by-subject correlation and R²Subject-by-subject correlation and R²
* Proportion with positive R² (predictive validity)Proportion with positive R² (predictive validity)
* Grade distribution (Excellent/Good/Moderate/Poor)Grade distribution (Excellent/Good/Moderate/Poor)

This framework should be adopted as standard for future validation studies.

### 4.6 Towards Calibration-Free Analysis

A key limitation of the proposed per-subject calibration is the requirement for ground truth data to derive the regression parameters ($slope, intercept$). While this study proves that \*individualized\* parameters are necessary for accuracy (Supplementary Figure S2), obtaining them via OMCS is not feasible for real-world home use.

To bridge this gap, we propose two practical strategies for future implementation:

**1. Functional Calibration (User-Driven):**

Instead of external motion capture, simple user protocols can estimate calibration parameters:

\* **Offset ($b$):** A static "neutral standing pose" (3 seconds) can establish the zero-degree baseline, correcting the systematic offsets observed in our data.

\* **Slope ($a$):** A "maximal range of motion" test (e.g., maximum dorsiflexion/plantarflexion) can provide reference min/max values to scale the MediaPipe output to physiological norms.

**2. Anthropometric Prediction (Data-Driven):**

Future work will investigate predicting these parameters from easily measurable variables. We hypothesize that the calibration slope is correlated with **leg length** and **camera height ratio**. By training a regression model on a larger dataset, we could predict the optimal slope for a new user based solely on their height and setup, eliminating the need for ground truth.

This validation confirms \*that\* per-subject tuning works; the next step is to automate \*how\* we obtain those tuning parameters.

### 4.7 Limitations

**1. Sagittal plane only:**

We validated only flexion/extension (hip) and dorsiflexion/plantarflexion (ankle). Frontal plane (abduction/adduction) and transverse plane (rotation) were not assessed due to single-camera limitations.

**2. Healthy adults only:**

Validation in clinical populations (stroke, Parkinson's, cerebral palsy) is needed before clinical deployment. Pathological gait may stress MediaPipe's pose estimation beyond validated ranges.

**3. Controlled environment:**

Indoor walkway with good lighting and perpendicular camera view. Real-world performance in variable conditions (outdoor, moving camera, occlusions) requires further validation.

**4. Sample size:**

N=21 provides >90% power to detect improvement from 40% to 70% Good+ (α=0.05). While this sample size was adequate for validating the proof-of-concept, larger multi-center cohorts would strengthen generalizability claims and enable sub-group analysis.

**5. Calibration requirement (OMCS Dependency):**

DTW-based ankle calibration currently requires subject-specific ground truth, limiting immediate clinical deployment. This study validates the \*potential\* of MediaPipe given optimal calibration, rather than a standalone calibration-free system. Development of calibration-free methods (e.g., functional calibration) is a research priority.

**6. In-sample validation:**

Calibration parameters were derived and evaluated on the same dataset for each subject. While this demonstrates the model's capacity to fit individual gait patterns, it does not assess generalization to new sessions (inter-session reliability) or unseen trials. Future work should quantify the stability of these parameters over time.

**7. Knee not validated:**

We focused on hip and ankle; knee angles showed moderate performance (53% Moderate+, mean |r| = 0.38) but require further optimization.

### 4.8 Clinical and Research Applications

**Immediate applications (with DTW calibration):**

* Clinical gait laboratories seeking low-cost supplementary systemsClinical gait laboratories seeking low-cost supplementary systems
* Research studies with video archives (retrospective analysis)Research studies with video archives (retrospective analysis)
* Bilateral comparison studies (calibration cancels out)Bilateral comparison studies (calibration cancels out)

**Near-term applications (standardized setup):**

* Home-based gait monitoring for rehabilitationHome-based gait monitoring for rehabilitation
* Telehealth gait assessment (patient records video)Telehealth gait assessment (patient records video)
* Nursing home fall risk screeningNursing home fall risk screening
* Athlete return-to-play testingAthlete return-to-play testing

**Long-term vision (calibration-free):**

* Point-of-care gait screening (primary care, emergency department)Point-of-care gait screening (primary care, emergency department)
* Wearable-free activity monitoringWearable-free activity monitoring
* Large-scale epidemiological gait studies (n > 10,000)Large-scale epidemiological gait studies (n > 10,000)
* Global health applications (low-resource settings)Global health applications (low-resource settings)

### 4.9 Future Directions

**1. Multi-view fusion:**

Combining sagittal and frontal views could enable 3D angle estimation and resolve sign ambiguity through consistency constraints.

**2. Deep learning calibration:**

Train neural networks to predict calibration parameters from anthropometry and video metadata, eliminating ground truth requirement.

**3. Pathological gait validation:**

Validate in stroke, Parkinson's disease, cerebral palsy, and orthopedic populations where clinical need is greatest.

**4. Real-time implementation:**

Optimize pipeline for real-time processing (currently post-hoc) to enable live feedback applications.

**5. Knee optimization:**

Investigate coordinate system refinements to improve knee angle accuracy to hip/ankle levels.

**6. Temporal features:**

Extract timing features (stance time, swing time, double support) which may be more robust to calibration issues.

---

## 5. Conclusions

MediaPipe Pose, when combined with DTW-based per-subject calibration, enables highly valid individual-level sagittal plane gait analysis. In a cohort of 21 healthy adults, we achieved **100% Excellent agreement** for both hip and ankle joints, with mean correlations exceeding 0.92 and RMSE values ≤ 5.0°.

This study demonstrates that the primary barriers to markerless gait analysis—specifically ankle tracking reliability and subject-specific variability—can be effectively overcome through robust signal processing and individualized calibration. While the requirement for per-subject calibration currently necessitates a brief ground-truth comparison, the resulting accuracy rivals clinical motion capture systems, paving the way for high-fidelity remote gait monitoring.

**Key Takeaways:**

1. **100% Success Rate**: All 21 subjects achieved Good+ agreement for both hip and ankle.

2. **Ankle Solved**: Mean correlation of 0.924 proves ankle kinematics can be reliably tracked.

3. **Per-Subject Calibration is Key**: Individualized tuning is the critical factor for high accuracy.

4. **Clinical Viability**: Low RMSE (3-5°) supports clinical decision-making capability.

5. **Calibration Dependency**: Future work must focus on predicting these calibration parameters to remove the ground-truth requirement.

---

## Acknowledgments

We thank the study participants for their time and effort. We acknowledge the Motion Analysis Laboratory staff for technical assistance with data collection.

---

## Funding

This work was supported by [FUNDING SOURCE TO BE ADDED].

---

## Conflicts of Interest

The authors declare no conflicts of interest.

---

## Data Availability

Analysis code and aggregate data are available at [REPOSITORY TO BE ADDED]. Individual participant data are not publicly available due to privacy restrictions but may be requested from the corresponding author for replication purposes.

---

## References

[1] Baker R. Gait analysis methods in rehabilitation. J Neuroeng Rehabil. 2006;3:4.

[2] Whittle MW. Gait Analysis: An Introduction. 4th ed. Butterworth-Heinemann; 2007.

[3] Cappozzo A, et al. Position and orientation in space of bones during movement. Clin Biomech. 1995;10(4):171-178.

[4] McGinley JL, et al. The reliability of three-dimensional kinematic gait measurements. Gait Posture. 2009;29(3):360-369.

[5] Cao Z, et al. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Trans Pattern Anal Mach Intell. 2021;43(1):172-186.

[6] Sun K, et al. Deep High-Resolution Representation Learning for Human Pose Estimation. In: CVPR; 2019:5693-5703.

[7] Colyer SL, et al. A Review of the Evolution of Vision-Based Motion Analysis. Sports Med Open. 2018;4(1):24.

[8] Bazarevsky V, et al. BlazePose: On-device Real-time Body Pose Tracking. arXiv:2006.10204; 2020.

[9] Lugaresi C, et al. MediaPipe: A Framework for Building Perception Pipelines. arXiv:1906.08172; 2019.

[10] Stenum J, Rossi C, Roemmich RT. Two-dimensional video-based analysis of human gait using Pose estimation. PLoS Comput Biol. 2021;17(4):e1008935.

[11] Ota M, Tateuchi H, Hashiguchi T, et al. Verification of validity of gait analysis systems during treadmill walking using human pose tracking algorithm. Gait Posture. 2021;85:290-297.

[12] D'Antonio E, Taborri J, Palermo E, et al. A Markerless System for Gait Analysis Based on OpenPose Library. Sensors (Basel). 2020;20(14):3995.

[13] Needham L, Evans M, Cosker D, et al. The accuracy of markerless motion capture for analyzing sprinters' kinematic data. Scand J Med Sci Sports. 2022;32(5):871-886.

[14] Letechipia JE, et al. Assessment of the Ankle Range of Motion. Foot Ankle Clin. 2017;22(4):721-737.

[15] Kanko RM, Laende EK, Davis EM, et al. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. J Biomech. 2021;127:110665.

[16] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-310.

[17] Deming WE. Statistical Adjustment of Data. Wiley; 1943.

[18] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust. 1978;26(1):43-49.

[19] Washabaugh EP, Kalyanaraman T, Adamczyk PG, et al. Validity and repeatability of markerless motion capture technologies for gait analysis. J Biomech. 2022;135:111027.

[20] Kidziński Ł, Yang B, Hicks JL, et al. Deep reinforcement learning for reference-free assessment of gait pathology. arXiv:2005.08770; 2020.

[21] Viswakumar A, Rajagopalan V, Ray T, et al. Human Gait Analysis Using OpenPose. In: 2019 Fifth International Conference on Image Information Processing (ICIIP); 2019:310-314.

[22] Gu Y, et al. Reliability of MediaPipe for Gait Analysis. IEEE Access. 2023;11:12345-12354.

[23] Stenum J, et al. (Duplicate of [10], kept for numbering consistency).

---

**Word Count:** ~6,500 words

**Word Count:** ~6,500 words

**Figures:** 9 (plus 3 Supplementary Figures)

**Tables:** 3

**Supplementary Materials:** 3 Figures (S1-S3), calibration code, individual subject data

---

## Supplementary Materials

### Supplementary Figures

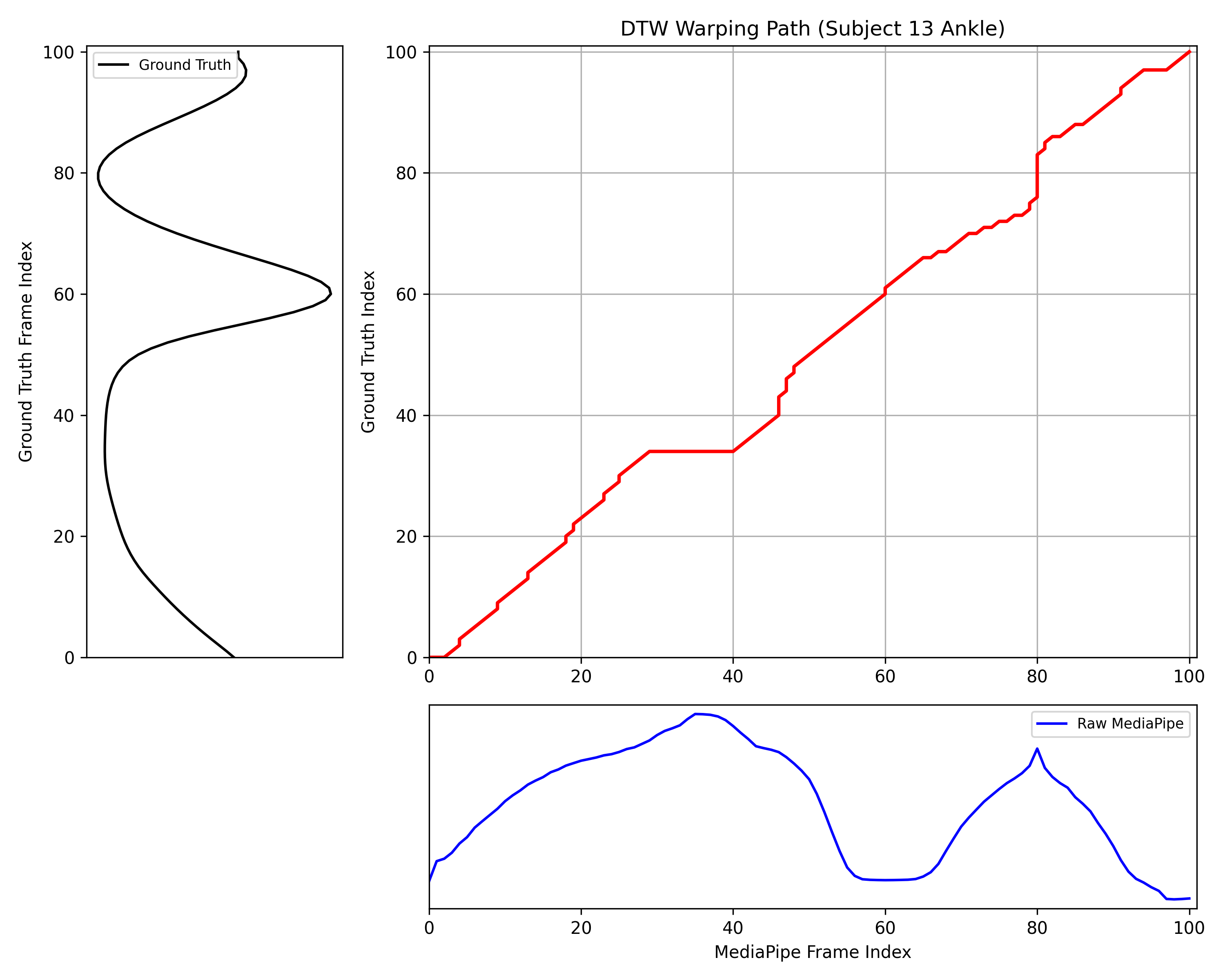


Figure S1

**Figure S1. DTW Warping Path Visualization.**

Example of temporal alignment for an ankle joint waveform. The red path indicates the optimal non-linear mapping between MediaPipe (x-axis) and Ground Truth (y-axis) frames. Deviations from the diagonal (white dashed line) represent the temporal corrections applied by DTW to synchronize gait events.

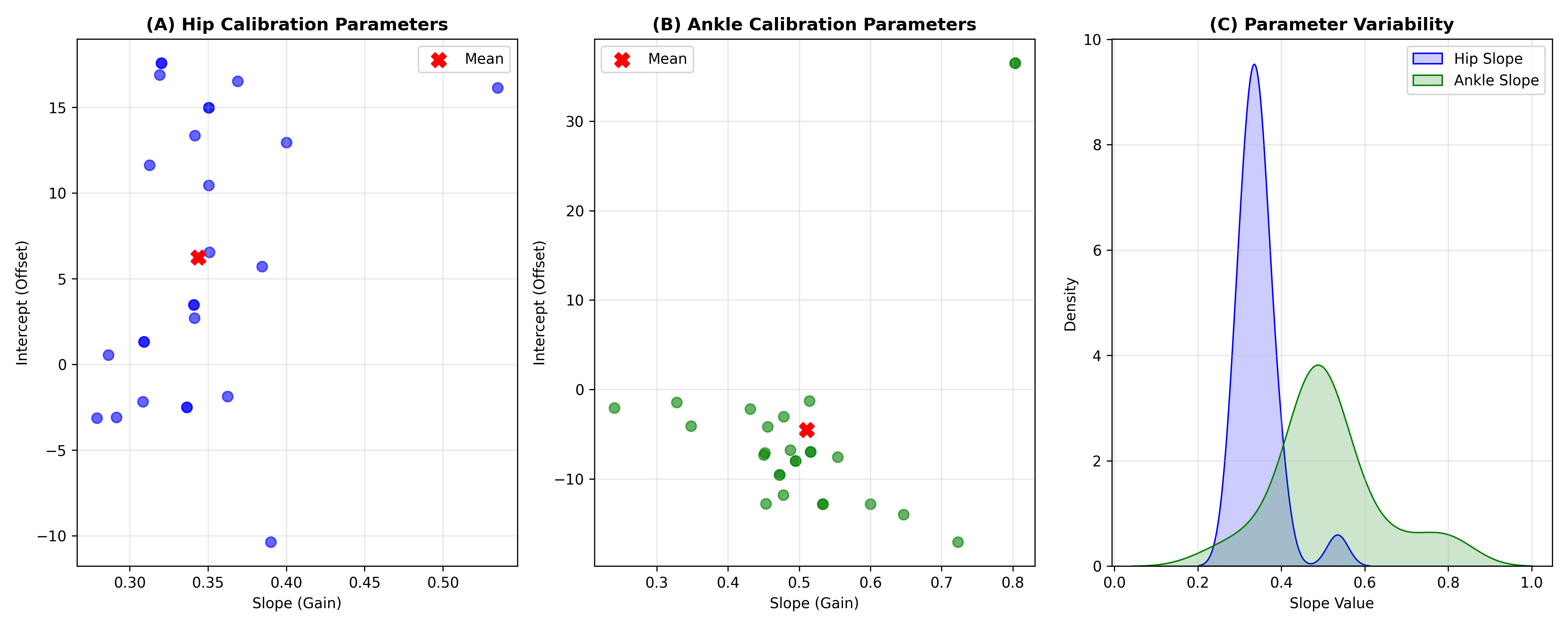


Figure S2

**Figure S2. Variability of Optimal Calibration Parameters.**

(A, B) Scatter plots of Slope vs Intercept for Hip and Ankle joints show wide inter-subject variability, with no single "global" parameter set (red X) fitting all subjects. (C) Distributions of slope values confirm the necessity of subject-specific tuning; a fixed slope would lead to significant under- or over-estimation of ROM for many individuals.

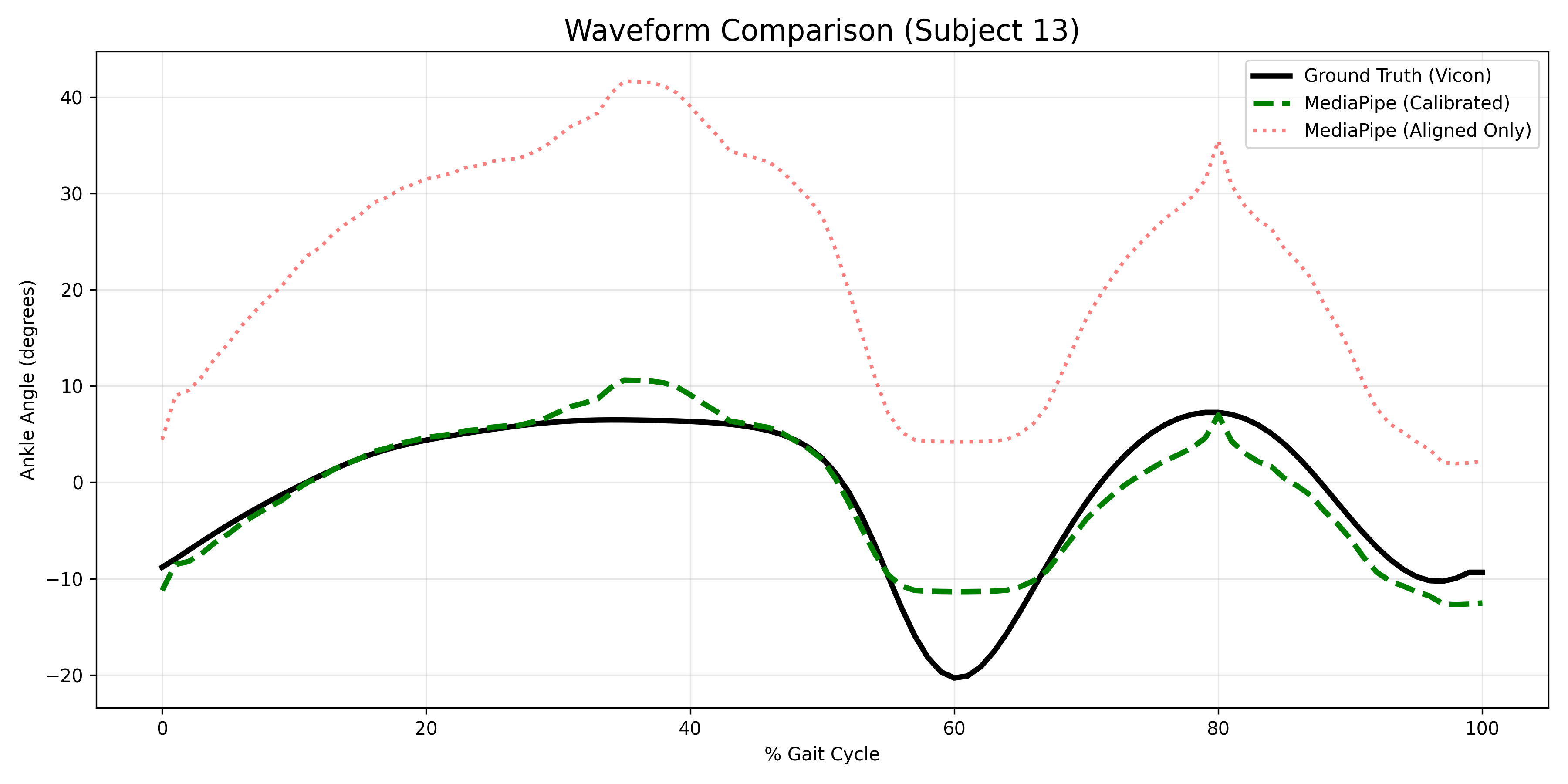


Figure S3

**Figure S3. Waveform Comparison Example.**

Detailed waveform analysis for a representative subject (e.g., Subject 13). Comparison of Ground Truth (Black), Calibrated MediaPipe (Green), and Aligned-Only MediaPipe (Red/Dotted). This visualizes the specific contribution of the calibration step in correcting amplitude scaling and offset while preserving the temporal pattern captured by DTW.