自動生成された評価ルールを用いた リアルタイム筋力トレーニング指導システム

鈴木貴之[1]　岡部誠[2]

概要：健康意識の高まりに伴い自宅でのトレーニングが普及しているが，専門家の指導がない環境では，不適切なフォームによる怪我のリスクや運動効果の低下が課題となっている．本研究では，ユーザが提供した手本動画を学習する筋力トレーニングのためのリアルタイムフィードバックシステムを提案する．動画から抽出した「最適な 1 回」のデータを基準とし，LLM による評価ルールの生成と，動的時間伸縮法（DTW）による同期を組み合わせることでリアルタイムにフィードバックするシステムを実現した．これにより，従来のルールベース方式では困難であった「種目の多様性」と「リアルタイム性」に対応した，パーソナルトレーニング体験を提供する．

キーワード：大規模言語モデル，フォーム評価，姿勢推定

1. はじめに

近年，予防医学や健康寿命の延伸への関心の高まりを背景[1]に，基礎代謝の維持，生活習慣病の予防，およびメンタルヘルスの改善において重要な役割を果たす[2]フィットネスへの需要は増加しており，なかでも自宅環境でのトレーニングは広く普及している[3]．一方，専門家の監督がない状況では，不適切なフォームの反復によるケガのリスクの増大や，狙った筋群に負荷が入らないことによる運動効果の低下が問題となる[4]．したがって，家庭環境でも容易に導入でき，かつ実施中に改善点を提示できるフォーム支援技術が求められている[5].

従来のフォーム評価では，姿勢推定により得られる関節角度などの幾何特徴を用い，動作の誤りを検出してフィードバックを提示する枠組みが検討されてきた[6，7，8]．しかし，重要視すべき関節，許容誤差，注目すべき局面，およびフィードバック内容といった評価ルールの設計は種目ごとに異なる．そのため，開発者が設計した固定的な規則・閾値に依存し，種目追加のたびに再設計や再調整を要することがある[8，9，10].

また，近年のビデオ言語モデル（VLM）や大規模言語モデル（LLM）を用いて解析を行い，フィードバックを生成する方法は高い汎用性を持つ[11]一方，計算コストや遅延の観点から，動作直後に修正を促すことは容易でない[12，13].

そこで本研究では，ユーザが用意した任意の手本動画を学習し，その動作に特化した指導を即座に行うシステムを提案する．本システムは，準備フェーズとトレーニング実行フェーズを分離したハイブリッド構成をとる．具体的には，準備段階において LLM を用いて動作の意味を解釈し，種目ごとの評価ルールを自動生成する．そしてトレーニング実行時には，生成された評価ルールに基づき，動的時間伸縮法（DTW）[14]等の軽量なアルゴリズムを用いてリアルタイムに判定を行う．これにより，「多様な種目への適応」を実現しつつ，ユーザの動作に対して遅延のない「リアルタイム指導」を提供することを目指す.

2. 関連研究

2.1 姿勢推定に基づく運動フィードバック

単眼カメラや RGB 映像から姿勢推定を行い，誤差を検出して運動フィードバックへ結びつける研究は広く行われてきた．特に近年では，Deep Learning に基づく姿勢推定の発展により，OpenPose[15] や MediaPipe[16] のように単眼カメラのみでリアルタイムに姿勢推定を行うことが現実的になっている.

姿勢推定を用いた運動フィードバックの代表例として，Fieraru らは 3D 姿勢系列に基づいて反復運動の分割と誤り検出を行い，人間が理解可能なフィードバック提示を目指す枠組みを提案している[6]．さらに，Ko らは人物検出と MediaPipe を組み合わせて関節角度の逸脱を推定し，Web インタフェース上でリアルタイムにフォーム修正を支援する[7]．また，Chen らは OpenPose により姿勢推定を行い，幾何学的特徴に基づいて姿勢品質を評価し，改善提案を提示するアプローチを示している[8].

本研究では，ユーザの動作をリアルタイムに解析するため MediaPipe を用いて手本動画とユーザの姿勢推定を行い，角度差分を用いてリアルタイムにフィードバックを行う.

2.2 LLM を用いた運動への応用

近年は，スポーツ分野において大規模言語モデルを用いて「何が良くて何が悪いか」「どう直すべきか」を推論や応答させる研究が注目されている[17][18]. ASHUTOSH らは，映像と 3D 姿勢系列を入力に，改善点を含む自由形式の専門家のコメントを生成する枠組みを提案し，スコア提示に留まらないコーチング出力を目標としている．さらに同研究では，専門家コメントの要約や身体部位ラベル付けといった整形処理に LLM を用いることも明記されており，言

1 静岡大学
Shizuoka University

ⓒ2026 Information Processing Society of Japan

語モデルをパイプラインの一部として組み込む流れが示されている[11]. また DELMAS らは，姿勢差分を自然言語で表現・生成するタスクを扱い，家庭内フィットネスでのコーチング的フィードバックを，言語で扱える可能性を述べている[19]. ただし，これらの枠組みをそのままリアルタイムフィードバックへ適用すると，推論遅延や計算資源の制約がボトルネックになり得る.

本研究はこの点に対し，LLM は事前処理（評価ルール生成・総評生成）へ寄せ，実行時は軽量な比較処理でリアルタイム性を担保する.

3. ユーザインタフェース

本システムのユーザインタフェースは，「手本登録」「登録データの確認・更新」「トレーニング実行」の 3 つの画面から構成される．ユーザは，これらの画面を順に利用することで，手本動作の登録から，実際のトレーニングとフィードバックまでを直感的に行うことができる.

本システムでは，トレーニング種目を大きく 2 つのタイプに分類する．スクワットのように同じ動作を繰り返すものを単一反復動作，ランジのように左右の手足が交互に動くものを左右交互動作と定義する．ユーザは手本登録時にこれらを選択する. 本論文ではスクワットを例に説明する.

3.1 手本登録画面と登録データの確認画面

まず，ユーザは「手本登録」画面において，行いたいトレーニング種目名（例:スクワット）と，動作タイプ（例：単一反復動作），YouTube などで公開されているプロのスクワットの解説動画などの動画をシステムに入力する.

その動画から抽出された情報が「登録データの確認・更新」画面（図 1）に表示される．本画面では，「手本データ解析結果タブ」や「評価ルールタブ」などの機能ごとにタブが分かれている.

「手本データ解析結果タブ」では，画面左側に動作 1 回分の手本動作が表示され（図 1 左），右側には手本動作の解析結果が数値として一覧表示される（図 1 右）．ここでは，1 回の動作にかかる時間や，動作中に重要となる関節角度などを，ユーザが直感的に把握できる．例えばスクワットの場合，1 回の動作時間は 3．3 秒となっており，また，最も負荷がかかる姿勢における膝の角度は 90 度や腰の角度は 90 度などが表示されている．つまり，ユーザは「どのようなフォームが基準になっているのか」を視覚的に確認できる.

「評価ルールタブ」を押すことでトレーニング中に提示されるフィードバック内容が一覧として表示される（図 2）．例えば，「右股関節を曲げて！」といった指摘がユーザの角度が手本の角度よりも大きく差が正の場合にされることが確認できる．このように，どのような条件で提示されるのかを，事前に確認することが可能である.

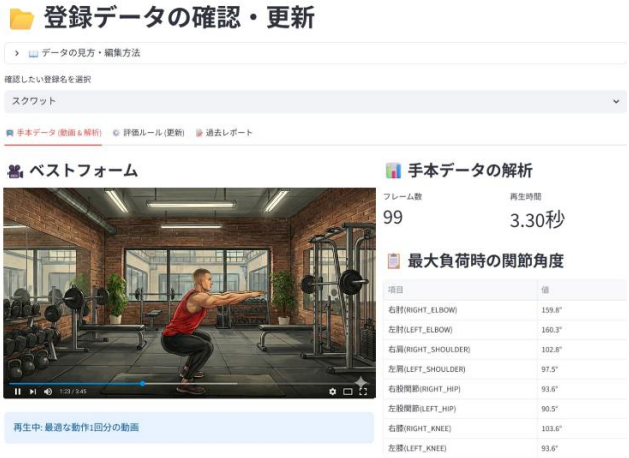

図 1：登録データの確認・更新画面

図 2：評価ルールタブ

3.2 トレーニング実行画面

「トレーニング実行」画面（図 3）では，ユーザがスタートボタンを押すことでトレーニングを開始する．画面左側には入力した手本動画が表示され，右側には Web カメラを通して取得したユーザ自身の映像がリアルタイムで表示される．これにより，ユーザは手本動作と自分の動作を同時に比較しながらトレーニングを行うことができる．ユーザの映像上には，手本動作の姿勢を示す紫色の半透明の骨格モデルが重ねて表示される（図 4（a））．ユーザが手本の姿勢と同じ姿勢になると，入力した動画が開始され，システムが音声で「スタート！」とユーザに伝える．ユーザが動作 1 回するごとに，「右膝伸ばして」といったフィードバックが画面下部に表示され，音声による読み上げも行われる. また，該当関節が次の動作の間赤く強調表示される（図 4（b））．これにより，ユーザは視覚と聴覚の両面から，直感的に指導を受けることができる.

トレーニング終了後には，全体の動作を振り返るための総評フィードバックが表示される．ここでは，フォームの崩れやすかった傾向や，次回に向けた改善点が提示され，ユーザは自身のトレーニング内容を客観的に把握できる．これらの結果は，登録データの確認画面にある「過去のレポート」からも，いつでも確認可能である.

ⓒ2026 Information Processing Society of Japan

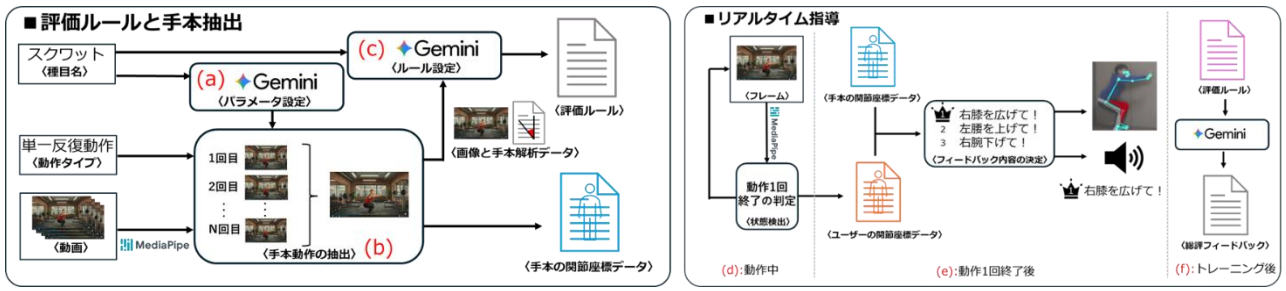

図 5：システムパイプライン．評価ルールと手本抽出において，ユーザは種目名，動作タイプ，動画を入力する．(a) システムは種目名から LLM を用いてパラメータ設定を行う．(b) 次に，出力されたパラメータ設定と動作タイプ，動画を用いて動画の中から手本となる動作 1 回の関節位置データを取得する．(c) 最後に，種目名と手本の動作 1 回の最負荷部の画像からトレーニング時に使用する評価ルールを生成する．(d) リアルタイム指導においては，リアルタイムで取得したフレームから動作状態を判定する．(e) ユーザの動作 1 回終了後，手本とユーザの関節位置データを比較し，ユーザに音声と画面表示でフィードバックを行う．(f) トレーニング終了後は，LLM によるフィードバック内容が出力される．

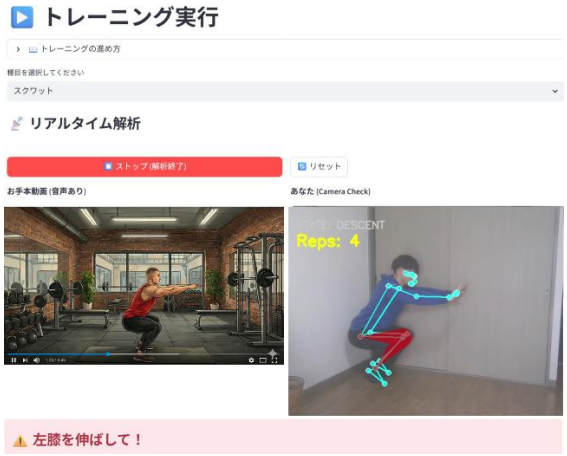

図 3：トレーニング実行画面

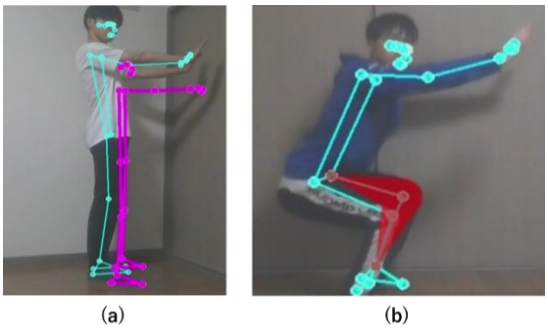

図 4：手本の骨格モデル（a）と関節の強調表示（b）

4. 分析と解析

本システムは，入力デバイスとして単眼 Web カメラのみを使用し，Streamlit 上で動作する．また，本システムで用いる LLM は Gemini API（gemini-2．5-flash）[20]である．図 5 にシステムパイプラインを示す．

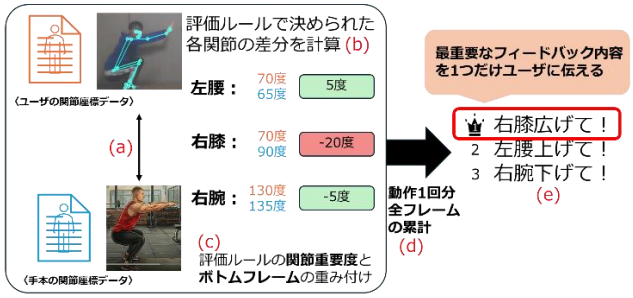

図 6：フィードバック内容の決定

図 6 は，本システムのフィードバック内容の決定の仕方を示す．　(a) DTW を用いて手本とユーザのフレーム対応を行う．(b) 対応づけられた各フレームに対して評価ルールによって定められた各関節の差分を計算する．例えば，右膝ではユーザが 70 度、手本が 90 度だとすると，ユーザの角度から手本の角度を引き，差が-20 度となる．同様に他の関節も計算を行う．(c) 角度差の絶対値に関節重要度とボトムフェーズに対して重みを付けスコアとする．(d) 動作 1 回分の全フレームに対して行いスコアを累計する．(e) スコアが最大なものをフィードバックする関節として選び，対応するフィードバック内容をユーザに伝える．

従来の手法では (a) で対応づけられたユーザと手本の関節位置データを LLM に入力する方法があるがリアルタイムでの出力が困難であるため，本研究では (b) や (c)、(e) で用いられる関節や関節の重要度、およびフィードバック内容をあらかじめ LLM を用いて出力させ．トレーニング実行時に評価ルールとして用いる．

4.1 骨格座標の取得と特徴量の選定

本システムでは，入力された動画および Web カメラから得たユーザのフレームから MediaPipe を用いて各関節の 2 次元座標（x，y）を取得する．MediaPipe は，入力フレームから各関節の 3 次元座標（x，y，z）を推定する機能を

ⓒ2026 Information Processing Society of Japan

有しているが，単眼映像からの深度推定は，ユーザの服装，体格，あるいは照明条件によって推定値に大きなノイズが含まれる傾向がある．

また，入力される手本動画は，一般に流通している 2 次元の動画ファイルである．ユーザの実環境（Web カメラ）と手本動画の撮影環境が異なる場合，両者の推定深度のスケールや精度を厳密に一致させることは困難である．

よって，本研究では単眼の Web カメラのみを入力デバイスとしており，2 次元座標（x， y）のみを解析対象として採用する．

4.2 手本動画の解析と最適動作の抽出

ユーザが入力した動画全体をそのまま手本とするのではなく，動画内に含まれる複数回の反復動作の中から，動作 1 回のみを抽出して「手本動作」として採用する．一般に，ユーザが入力した動画には，動作開始前の予備動作やセット間の休憩，あるいは疲労に起因するフォームの乱れといったノイズが含まれることが多い．これらを排除せずに手本として採用すれば，システムは不適切な動作までも「模倣すべき正解」として誤学習するリスクがある．加えて，長時間の時系列データに対する DTW の常時探索は計算コストの増大を招き，リアルタイム指導における致命的な遅延要因となる．したがって，複数の反復動作の中から最も標準的で安定した動作 1 回を抽出する．

4.2.1 LLM による解析パラメータの推定

手本動作の抽出には，どの関節の動きを基準関節とするか，およびノイズ除去量（平滑化）やピーク検出条件などの解析パラメータを適切に設定する必要がある．従来は開発者が種目ごとに手動調整するか固定値を用いることが多いが，多様な種目に対応する本システムでは個別調整が拡張性を阻害する．そこで本研究では，LLM の知識を活用し，種目名のみから解析パラメータをゼロショットで推論して出力させる．プロンプトでは LLM に「スポーツバイオメカニクスの専門家かつデータアナリスト」の役割を付与し，出力形式を JSON に制約した．

本システムでは，ユーザ入力の種目名（例：スクワット）から，主に次を出力させる．

■**手本抽出の基準関節**：周期性が顕著な関節名．

■**信号処理パラメータ**：平滑化窓幅や最小ピーク距離など，誤検知を抑えつつピーク検出精度を高める設定．

■**DTW 評価用の特徴量**：「関節 3 点の組」のリスト（例：右膝角は（右腰，右膝，右足首）と記載される）．

4.2.2 動作 1 回ごとの切り分け

ユーザが入力した動画を解析し，動作の開始姿勢から次の開始姿勢までの動作 1 回ごとに切り分ける処理を行う．

本システムでは，単一反復動作には基準関節の Y 座標に基づくピーク検出を，左右交互動作には主成分分析（PCA: Principal Component Analysis）[21]による左右分離を適用する．左右交互動作では，単一の時系列データ（例：腰の Y 座標の変化）だけでは右動作と左動作の波形が同じになり，トレーニング実行時にユーザの動作と逆側の手本動作を誤って対応付けるリスクがある．

■**単一反復動作**

単一反復動作では，動作の周期性が顕著に表れる基準関節の Y 座標時系データから動作 1 回を切り分ける．

まず，解析パラメータにより指定された基準関節（例：右腰）について，各フレームの Y 座標を取得し時系列データ y(t)を構成する．MediaPipe の座標は画像左上を原点（0, 0）とする正規化座標であり，下方向ほど Y が増加するため，沈み込み（下降局面）は Y の増加として観測される．

y(t)には Web カメラの解像度限界や微細な身体の揺れに起因するノイズが含まれるため，データの変化を正しく捉えつつ，動作の重要な特徴であるピーク形状を維持することのできる Savitzky-Golay フィルタを適用する．平滑化された y(t)に対し，前述の座標系特性により物理的な「最大負荷点（ボトムポジション）」に対応する極大値（ピーク）を検出する．検出された全ピークのフレーム間隔の中央値を「基準テンポ T_ref」とし，これから大きく外れる区間（準備動作・停止など）を外れ値として除外することで，安定に反復されているトレーニング区間を特定する．

安定区間内で検出したピークの前後にある極小値を探索する．Y 座標が最小となる極小値は身体が最も高い開始・終了姿勢に対応するため，直前の極小値を開始点，直後の極小値を終了点として 1 レップ区間を確定する．ただし先頭・末尾の動作で片側の谷が得られない場合は，基準テンポ T_ref を用いて開始点または終了点を補完し，ボトムを含む完全な動作 1 回ごとの切り分けを実現する．

■**左右交互動作**

左右交互動作では，上下動だけでは左右の区別がつかないため，本システムでは左右の関節の座標差分を特徴量として主成分分析（PCA）を適用する．PCA とは，高次元データの分散が最大となる方向（主成分軸）を探索し，新たな基底空間へ射影する次元圧縮手法である．

各フレームで股関節を原点とし，胴体サイズでスケーリングして関節座標を正規化した後，肩・肘・手首・股・膝・足首の左右 6 ペアについて座標差分を計算し，12 次元の特徴量ベクトルを構成する．

多次元の差分の特徴量ベクトルを動作の周期性を表す単一の時系列データに変換するために PCA を適用する．左右交互動作において，時系列データ上で最も分散が大きくなる要因は，左右の脚が入れ替わる動きであるため，第 1 主成分（PC1）スコアを抽出して 1 次元時系列データへ圧縮する．このデータは左右の入れ替わりに応じて正負が変わるため左右を分けて表現できる．

ⓒ2026 Information Processing Society of Japan

得られる PC1 スコアの絶対値 |PC1|を周期検出用の 1 次元データとして用い，単一反復動作と同様に平滑化，極値検出および基準テンポ T_ref に基づく安定区間抽出を行うことで動作 1 回ごとに切り分けを行う．

その後，ピーク時点の PC1 符号により 2 つの動作グループに分離する．この動作グループの右動作か左動作かを決めるため，各動作で左右の足首の座標分散を比較し，移動量が大きい側を踏み込み側とみなして左右を割り当てる．具体的には，（右足首分散−左足首分散）を動作グループごとに集計し, 正であれば右動作, 負であれば左動作とする．

4.2.3 DTW を用いた手本動作の選定

前節までの処理により切り出された N 回の動作の中から, フィードバックの基準となる「手本動作」を選出する．抽出された動作$R = \{r_1,\ \ldots,\ r_N\}$の組み合わせに対して，動作ごとの速度差（時間軸の伸縮）を吸収し，純粋な軌道形状の違いのみを定量化できる動的時間伸縮法（DTW）を用いた総当たり比較を行い，相互の距離を計算する．そして，以下の式に基づき，他のすべての動作との距離の総和が最小となる動作r_{best}，すなわちデータセットの統計的中心に位置する動作を特定し，これをシステムの手本動作として採用する．

$$r_{best} = \underset{r_i \in R}{\operatorname{argmin}} \sum_{j=1}^{N} DTW\ (r_i,\ r_j)$$

複数の時系列データから代表波形を求める手法としては平均化が一般的であるが，身体動作において単純な平均を行うと，波形のピークが鈍り，ボトムのような切り返しの情報が失われる可能性がある．対して，本システムでは実在するデータの中から一つを選ぶアプローチを用いるため, 人間が行った動作の微細なニュアンスや物理的な整合性を完全に保ったまま，手本動作のデータを得られる．

また左右交互動作では，右動作と左動作それぞれについて同様に両側の手本動作を選ぶ．

4.3 LLM による評価ルールの自動生成

抽出された手本動作は動作 1 回分の関節位置データとして保存されるが，関節位置データだけでは「膝を曲げるべき」「腰を落とすべき」といった種目固有の評価基準を直接定義できない．そこで本システムでは，LLM を用いて手本データから評価ルールを自動生成する．評価ルールとは，リアルタイムフィードバックを行うための重要視する関節やフィードバック内容などの種目特有の設定である．

入力として，種目名に加え，最大負荷時のフレームであるボトム画像と，その関節角度や動作所要時間など手本動作の解析結果を LLM に与える．これにより, 骨格座標のみでは不足しやすい身体の向きや動作局面の情報を補完し，より適切な評価ルール生成を可能にする. プロンプトでは, LLM に「スポーツバイオメカニクスの専門家」という役割を与えた．

出力は JSON 形式とし，フィードバックの内容に加え，動作回数のカウントや DTW の特徴量として用いる関節などのアルゴリズム制御パラメータを含めて定義し，フィードバック内容は種目において重要度の高い関節のみに限定することで，重要な修正点に絞った提示を実現する．

4.4 DTW を用いたリアルタイム指導アルゴリズム

生成された評価ルールを用いて，Web カメラの前で動くユーザに対し動作 1 回ごとにフィードバックを行う．リアルタイムフィードバックは，開始姿勢の確定および動作 1 回の検出，手本とユーザのフレームの同期，重み付きスコアに基づくフィードバック内容の決定，音声・視覚提示から構成される．

4.4.1 開始判定と状態遷移による動作 1 回の検出

まずユーザが手本の開始姿勢に十分近い状態を一定フレーム数維持したときにトレーニングを開始する．開始姿勢の近さは，手本開始姿勢とユーザ姿勢のコサイン類似度で評価し，閾値を超える状態が連続 15 フレーム程度続いた時点で開始とする．開始が確定すると，開始時の基準関節の Y 座標を保存し，以降の動作 1 回の検出を行う．

動作 1 回の検出は，基準関節 Y 座標の変化に基づく 3 状態のステートマシンで行う．

- **WAIT（待機）**：開始姿勢付近で待機し，基準関節 Y が開始位置から一定量以上（例：+0．05）増加したとき下降開始と判定し DESCENT へ遷移する．
- **DESCENT（下降）**：下降中はボトム候補（Y 最大）を更新し続ける．直近 5 フレームの Y 平均との差よりも現在値が十分小さくなった場合（例：平均との差−0．01）に上昇開始とみなし，ASCENT へ遷移する．
- **ASCENT（上昇）**：基準関節 Y が開始位置付近（例：開始+0. 02 以内）へ戻った時点で動作 1 回完了とし, WAIT に戻る．

左右交互動作の場合，システムは手本解析結果に基づき，第 1 動作を左側の動作として初期化する．実行時には，動作 1 回の完了を検知するたびに，判定対象とする手本データを右・左へと入れ替える．この制御により，DTW 適用時においてユーザと手本動作の左右の整合性を常に担保し，リアルタイムでの誤一致を防止している．

4.4.2 DTW を用いたタイミング同期

ユーザと手本では動作速度が異なるため，単純なフレーム対応では比較できない．そこで本システムでは，計算コストを抑えつつ同期精度を担保するため，フレーム毎の逐次計算ではなく動作 1 回が完了した時点で，その動作中に蓄積したユーザの関節位置データと手本動作の関節位置データに対して DTW を適用し，フレーム対応を行う．実装では計算量を抑えるため，高速 DTW を用いている．DTW

ⓒ2026 Information Processing Society of Japan

によるフレームの同期には，評価ルールで指定された DTW に用いる関節角度を特徴量として用いる．

4.4.3 フィードバック内容の決定

動作 1 回の中で最重要な改善点を決めるために DTW で対応付いた全フレームの大きさを T，同期された各フレーム t として，評価項目 i（例：「左膝の角度」）のスコア S_i は以下の式で算出される．

$$S_i = \sum_{t=1}^{T} \max(0,\ |\theta_{user}(t) - \theta_{ref}(t)| - \theta_{thresh}) * W_{joint}^{i} * W_{phase}(t)$$

ここで，各変数は以下を表す．

- $|\theta_{user}(t) - \theta_{ref}(t)|$: 対応するフレーム間の角度差分．
- θ_{thresh}: LLM が設定した許容誤差（例：15 度）．この閾値を超えた分のみをペナルティとして加算することで，Web カメラの計測ノイズによる過剰指摘を抑制する．
- W_{joint}^{i}: 評価ルールに基づく，その種目における当該関節の重要度（例：スクワットなら膝や股関節を高く設定）．
- $W_{phase}(t)$: 動作フェーズ（下降・ボトム・上昇）に応じた重み係数．最大負荷がかかり，怪我のリスクが高い「ボトムフェーズ」等の重みを設定する．本システムでは，ボトムの重みを 3．0，それ以外のフェーズを 1．0 としている．

この累積スコアS_iが最も高い項目を最重要な改善点として特定する．この累積加算方式により，一瞬の計測ノイズや微細なズレに過剰反応することなく，動作全体を通してユーザが最も修正すべきフォームの崩れを的確に抽出することが可能となる．

4.4.4 音声および表示によるフィードバックの提示

フィードバック内容は，動作 1 回終了後の音声，および次の動作の間，画面上の関節の強調表示で提示する．左右交互動作では，左右を分けて管理し，「右足の動作に対する指摘」は「次回の右足動作時」に提示される．

■音声フィードバック:評価ルールに基づくフィードバック内容（例：「右膝を伸ばして」）を，Microsoft Speech API（SAPI） を用いて低遅延で音声合成を行いフィードバックする．

■視覚フィードバック: 次の動作を行う間，修正すべき関節が赤く強調表示される．ユーザは鏡を見る感覚で「赤くなっている場所」を意識するだけで，直感的にフォーム修正を行うことができる．

4.5 LLM による総評フィードバック

トレーニング終了後，各動作の所要時間や動作の深さ，各レップで発生したエラー内容およびその回数を LLM に入力し，次回のトレーニングに向けた包括的なフィードバックを自動生成する．

プロンプトにおいて，LLM に対し「プロのパーソナルトレーナー」という役割を与え，単なる回数表示に留まらず，図 7 に示すように「回数を重ねるごとにスクワットの深さが少しずつ改善している」や「1 回当たりの動作の所要時間も安定している」などの全体の傾向や「次回は左膝を強く踏み込む意識をしましょう」などの次回に向けたフィードバックを生成する．

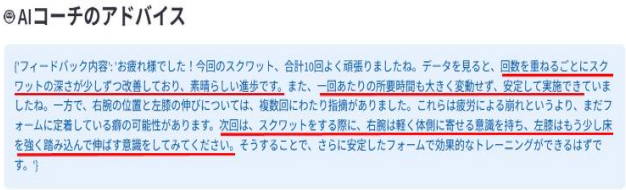

図 7：スクワットにおける総評フィードバック

5. 実験

本章では，実装したシステムの有効性を検証するために行った生成された評価ルールとその評価ルールに基づいたリアルタイム指導の妥当性，およびユーザテストについて述べる．

5.1 評価ルール生成

■スクワット

「スクワット」という種目名と「単一反復動作」の動作タイプ，右向きにスクワットを行っている手本動画を入力した際に，LLM が生成した評価ルールの一部を図 8 に示す．結果として，基準関節（rep_trigger_joint）は右股関節が設定された．MediaPipe を用いて取得した座標の信頼度が右側の方が高いため，適した左右の選択していることが分かる．関節ごとの重みに関して，スクワットで重要とされる腰や膝への重みが高く設定されていることが分かる．また，フィードバック内容として「左股関節の深さ」や「右膝の深さ」といったスクワットにおいて重要な指標が抽出され，閾値には Web カメラの誤差を考慮した 20．0 度が設定されたことを確認した．

図 8：スクワットにおける評価ルールの一部

■プッシュアップ

プッシュアップにおいても，同様にして評価ルールの生成を行った．図 9 に示すように基準関節としてプッシュアップにおいてよく動く関節である右肩が選ばれている．ま

ⓒ2026 Information Processing Society of Japan

た，重要な関節は肘，肩，腰，膝とされ，特に肘や肩の重みが高く設定されているため正しく設定出来ていると分かる．また，フィードバック内容に関して，スクワットと同様に腰についてのフィードバック内容が設定されているが「腰を下げて」や「腰を上げて」というように種目ごとの特性を反映したフィードバック内容が生成されていることが分かる．

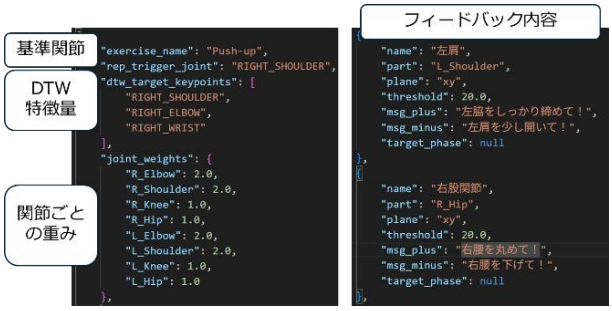

図 9：プッシュアップにおける評価ルールの一部

5.2 リアルタイム指導

5.1 で生成された評価ルールを用いてスクワットでは手本動作よりも膝を深く曲げるフォームでの動作を行い，プッシュアップでは手本動作よりも腰を高く上げたまま動作を行った．その結果を図 10 に示す．スクワットにおいて動作終了直後，同時に音声によって「左膝を伸ばして！」という指示が提示され，次の動作では画面上の左膝部分に強調エフェクトが表示された．また，プッシュアップにおいても「左腰を下げて」と音声によるフィードバックがなされ，次の動作では左腰が赤く強調表示された．

これらは，手本動作データとの DTW 距離計算において，膝関節の角度差分が設定された閾値を超過したことを正しく検知した結果である．これにより，提案手法が種目ごとに幾何学的なズレを言語的な指導へと正しく変換できていることが示された．

図 10：フォームのフィードバック

5.3 ユーザテスト

提案システムの有用性および使用感を評価するためにユーザテストを行った．

健康な大学生 9 名を対象に，スクワットでの評価実験を行った．被験者は手本動画と Web カメラ映像のみを見て条件 1，提案手法を用いた条件 2 の 2 つの実施条件としてスクワットをそれぞれ 10 回行った．順序効果を抑えるためクロスオーバー試験とし，被験者を条件 1，条件 2 の順に実施するグループ A（5 名）と，条件 2，条件 1 の順に実施するグループ B（4 名）に分けた．実験では，まず各被験者に手本動作 1 回分の動画とフィードバック内容を確認させ，実施画面の UI を説明した．次に，全身が映るよう Web カメラの角度を調整したうえで，各グループの最初の条件を実施し，1 分間の休憩を挟んで次の条件を実施した．全条件終了後，被験者はアンケートに回答した．

アンケートは，5 段階リッカート尺度（1＝全く思わない～5＝強くそう思う）による質問項目と，自由記述（本システムの良かった点，改善点，気づいたこと等）から構成される．アンケートの質問の結果を表 1 に示す．

表 1：アンケートの質問および回答の平均結果

	質問	平均
Q1	自分の映像だけ見るより、システムがあった方が正しいフォームを確認出来た.	4.7
Q2	画面上の赤い強調表示を見て、直すべき箇所が直感的に分かった.	3.9
Q3	音声フィードバックのタイミングや内容は適切だった.	4.1
Q4	トレーニング終了後のLLMによるアドバイスは、具体的で分かりやすかった.	4.7
Q5	このシステムを実際のトレーニングでも使いたいと思う.	4.4
Q6	このシステムを使用して、「次も頑張ろう」というやる気（モチベーション）が出た.	4.1
Q7	操作が直感的で分かりやすかった.	4.6
Q8	システムの反応速度(遅延)は気にならなかった.	4.2

アンケート結果（5 段階平均）では，全 8 項目が 3.9～4.7 と高評価であり，特に Q1 の「フォーム確認に役立った」と Q4 の「LLM の助言が具体的で分かりやすい」が高かったことから，本システムが自己フォームの理解と改善点把握に有効である可能性が示された．また，Q7 の操作性と Q4 の利用意向も良好であった．

また，自由記述において以下のような意見を得た．

- リアルタイムで解説が入ると，どこを直せばいいのかすぐに分かるのでよかった．「膝を伸ばして」「股関節を伸ばして」と言われてもあまりピンと来なかった．全身を画面内に入れるのが大変だった．
- スクワットとは余り関係のなさそうな右腕への指摘が多かったのが気になりました．
- どこを直せばいいかその都度分かったのでとても使いやすかった．

自由記述では，リアルタイムで修正点が分かり集中できる点が評価された反面，指示が不明確で誤解が生じること，種目に対して重要度の低い部位（腕）への指摘，初期姿勢合わせや撮影（全身が画角に入る）の難しさが課題として挙げられた．これらは表 1 の Q2 や Q3 の結果が他の項目に比べて低いことにも示されている．

ⓒ2026 Information Processing Society of Japan

6. まとめ

本研究では，自宅での筋力トレーニングにおける怪我のリスク軽減と運動効果の最大化を目指し，LLM による評価ルールの自動生成と DTW を用いたリアルタイムフィードバックシステムを提案した．

従来のフォーム指導システムでは，種目ごとに個別の判定ロジックを実装する必要があり，多種多様なトレーニング種目への対応が困難であった．これに対し，本システムではユーザが提示した動画から身体の関節角度や動きの特徴を抽出し，LLM（Gemini API）を活用して評価ルールを動的に生成する手法を導入した．これにより，特定の種目に依存しない汎用的なフォーム評価基盤を構築することが可能となった．

また，実行時のフィードバックにおいては，DTW を用いることで，ユーザの動作速度の変化に柔軟に対応しながら手本データとの時間的同期を実現した．これにより，トレーニングの動作中にリアルタイムで具体的な改善指示を提示することが可能となった．

今後の課題としては，カメラの設置角度や照明条件による姿勢推定精度の変動への対策，およびより多種目に適応できる手本動作の抽出などが挙げられる．さらに，LLM が生成するルールの精度をより高めるために，専門的な解剖学知識を用いたプロンプトエンジニアリングの最適化についても検討を進める必要がある．

7. 謝辞

ユーザテストにご助力頂いた皆様に深く感謝致します．

8. 参考文献

[1] 厚生労働省．健康日本 21（第三次）．
[オンライン]入手先:
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/kenkounippon21_00006.html（参照 2026-02-08）

[2] 厚生労働省：健康づくりのための身体活動・運動ガイド．［オンライン］入手先:
https://www.mhlw.go.jp/content/001194020.pdf （参照 2026-02-08）

[3] A'NAJA, M. Newsome, et al. 2025 ACSM worldwide fitness trends: future directions of the health and fitness industry. ACSM's Health & Fitness Journal, 2024, 28.6: 11-25.

[4] KERR, Zachary Y.; COLLINS, Christy L.; DAWN COMSTOCK, R. Epidemiology of weight training-related injuries presenting to United States emergency departments, 1990 to 2007. The American Journal of Sports Medicine, 2010, 38.4: 765-771.

[5]SIGRIST, Roland, et al. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. Psychonomic bulletin & review, 2013, 20.1: 21-53.

[6]FIERARU, Mihai, et al. Aifit: Automatic 3d human-interpretable feedback models for fitness training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 9919-9928.

[7]KO, Yeong-Min; NASRIDINOV, Aziz; PARK, So-Hyun. Real-time AI posture correction for powerlifting exercises using YOLOv5 and MediaPipe. IEEE Access, 2024.

[8]CHEN, Steven; YANG, Richard R. Pose trainer: correcting exercise posture using pose estimation. arXiv preprint arXiv:2006.11718, 2020.

[9]ZHAO, Ziyi, et al. 3d pose based feedback for physical exercises. In: Proceedings of the Asian Conference on Computer Vision. 2022. p. 1316-1332.

[10]KANASE, Rahul Ravikant, et al. Pose estimation and correcting exercise posture. In: ITM Web of Conferences. EDP Sciences, 2021. p. 03031.

[11]ASHUTOSH, Kumar, et al. ExpertAF: Expert actionable feedback from video. In: Proceedings of the Computer Vision and Pattern Recognition Conference. 2025. p. 13582-13594.

[12]ZHANG, Haoji, et al. Flash-VStream: Efficient Real-Time Understanding for Long Video Streams. arXiv preprint arXiv:2506.23825, 2025.

[13]SCHNEIDER, Benjamin, et al. QuickVideo: Real-Time Long Video Understanding with System Algorithm Co-Design. arXiv preprint arXiv:2505.16175, 2025.

[14]S. Sempena, N. U. Maulidevi, and P. R. Aryan. Human action recognition using dynamic time warping. In Proc. ICEEI. 2011, pp.1-5, IEEE.

[15]CAO, Zhe, et al. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 7291-7299.

[16]LUGARESI, Camillo, et al. Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172, 2019.

[17]髙橋洸汰， 岡部誠． 後方伸身宙返りの段階的学習をサポートする練習支援システム. WISS 2025, 2025, p.142-149.

[18]LI, Haopeng, et al. Sports-qa: A large-scale video question answering benchmark for complex and professional sports. arXiv preprint arXiv:2401.01505, 2024.

[19]DELMAS, Ginger, et al. Posefix: correcting 3d human poses with natural language. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. p. 15018-15028.

[20]Google Gemini. ［オンライン］入手先：https://gemini.google/about/（参照 2026-01-29）．

[21]WOLD, Svante; ESBENSEN, Kim; GELADI, Paul. Principal component analysis. Chemometrics and intelligent laboratory systems, 1987, 2.1-3: 37-52.

ⓒ2026 Information Processing Society of Japan