
1. CHAPTER 1 - INTRODUCTION

1.1. Background

The process of learning to dance has always been dependent on a process of watching a teacher and instant corrections. But to most students, working at home implies not getting that professional feedback. Although YouTube instructions and social media have made dancing easier, it is a one-way traffic, the student can watch the teacher, but nobody is keeping an eye on the student to correct any technical errors. This usually results in the novices making the same posture and timing mistakes without even knowing it.

Now we can follow body movements with the help of a smartphone thanks to the emergence of computer vision technologies such as MediaPipe, which is developed by Google. But simply getting a list of coordinates is not enough for a dancer. The actual challenge is in contrasting two different videos which may begin at different times or may have a different tempo. The project is aimed at developing a system that can address the following practical problems. The system can be used to give objective grading by aligning with Fast Dynamic Time Warping (FastDTW) and fixing timing offsets with the help of Audio Cross-Correlation. Better still, it converts that information to the language of the dancers such as straighten your elbow and it becomes more of a personal trainer than a computer program.

To ground the evaluation framework in established dance pedagogy, this research adopts two dimensions from the Indonesian traditional dance philosophy: Wiraga and Wirama. Wiraga, derived from the word raga (body), refers to the correctness of physical form: the precision, control, and execution quality of each movement. Wirama, derived from irama (rhythm), refers to the dancer's temporal synchronization with the musical beat. A third dimension, Wirasa (rasa, meaning feeling), encompasses emotional depth and artistic expression; however, as this dimension cannot be reliably quantified through 2D pose landmark analysis, it falls outside the computational scope of this system and is identified as a direction for future work. Wiraga and Wirama together form the two-pillar scoring framework of this system, each weighted equally at 50%, and both validated by an expert dance coach with 20 years of professional experience.

1.2. Problem Identification

Even with the current technology available, several specific issues make it hard to automate dance training:

- The "Starting Time" Problem: Students rarely start dancing at the exact same millisecond as the teacher in a recording, which makes simple frame-by-frame comparison impossible.
- Complex Data vs. Simple Feedback: Most AI systems give results in joint angles or coordinates, but dancers need clear, actionable advice they can understand.
- Visibility and Occlusion: In home environments, limbs often get hidden behind the body or go out of frame, which usually causes the tracking data to break or become unreliable.
- Metric Accuracy: There isn't a widely used standard that proves whether a computer-generated score actually matches the grade a professional dance judge would give.

1.3. Research Objectives

This research aims to implement and evaluate a lightweight AI-assisted self-assessment prototype for basic hip-hop dance training. The objectives are as follows:

1. To implement a kinematic analysis pipeline that extracts 13 analytical joint-angle features from standard dance video recordings using MediaPipe BlazePose.
2. To implement temporal alignment using FFT-based audio cross-correlation for global start-time correction and FastDTW for local movement-sequence comparison.
3. To implement movement-specific Wiraga scoring for Bouncing, Stepping, and Sliding, together with Wirama scoring for timing synchronization.
4. To implement a rule-based semantic feedback module that translates detected movement deviations into learner-friendly corrective messages.
5. To evaluate the agreement between system-generated scores and expert evaluation scores.

6. To evaluate the usability of the proposed system for beginner users using the System Usability Scale.

1.4. Research Questions

- RQ1: How can FFT-based audio cross-correlation and FastDTW be implemented to align student and reference dance videos?
- RQ2: How can kinematic deviations from pose-tracking data be translated into rule-based semantic natural-language feedback for beginner hip-hop learners?
- RQ3: To what extent do system-generated dance scores correlate with professional expert evaluation scores?
- RQ4: How usable is the proposed dance self-assessment system for beginner users?

1.5. Hypothesis

- H1: FFT-based audio cross-correlation and FastDTW can be implemented as a temporal alignment mechanism for student and reference dance videos.
- H2: Kinematic deviations can be translated into rule-based semantic natural-language feedback for beginner hip-hop learners.
- H3: System-generated dance scores will strongly correlate with expert evaluation scores, with Pearson correlation r greater than or equal to 0.70.
- H4: The proposed dance self-assessment system will achieve acceptable usability, indicated by a System Usability Scale score of at least 68.

1.6. Significance of the Study

This study contributes to the development of accessible AI-assisted movement-learning tools for beginner dance practice.

Technically, the study demonstrates an integrated pipeline that combines pose estimation, landmark visibility filtering, joint-angle extraction, audio-based temporal alignment, FastDTW sequence comparison, movement-specific scoring, and semantic feedback in a lightweight web-based system.

Pedagogically, the study provides beginner learners with structured self-assessment support when a dance instructor is not immediately available. Instead of

receiving only numerical values, learners receive feedback that identifies relevant movement areas such as knee bending, foot placement, posture, and timing.

Practically, the study evaluates the system against professional expert scoring and usability feedback. The evaluation identifies the strengths and limitations of the current prototype, particularly the need for score calibration and improved movement-specific features before expert-level reliability can be claimed.

1.7. Scope and Limitations

The system is limited to three foundational hip-hop movements: Bouncing, Stepping, and Sliding. It is designed for single-person recordings in an indoor environment using a stationary smartphone camera or similar device. The student and reference video are expected to use the same music track so that audio-based temporal alignment can be performed. The system derives 13 analytical joint-angle features from selected BlazePose landmarks. It does not use all 33 landmarks as direct scoring inputs because the selected features are sufficient for the current movements. The system supports temporary landmark occlusion through visibility filtering and temporal fallback, but it is not designed for multi-person videos, severe camera-angle changes, floor work, breakdancing, or expressive dance assessment. The current system uses a user-selected movement category. Although an internal movement-pattern classifier is used for contextual joint weighting and timeline display, it does not yet validate, reject, or automatically correct an incorrectly selected upload category. The three evaluated movements were selected and validated by Coach Ambrosius Robby as suitable foundational exercises for beginner hip-hop learners.