THE UNIVERSITY OF BAMENDA

NATIONALHIGHER POLYTECHNIC INSTITUTE

DEPARTMENT OF COMPUTER ENGINEERING

REAL-TIME AI MUSIC ACCOMPANIMENT VIA PREDICTIVE LATENCY COMPENSATION

NCHINDE TANDJONG JOSUE	UBa25EP063

Instructor:

Dr. TCHAHOU T. ACHILLE

1. BACKGROUND

Generative Artificial Intelligence has revolutionized audio synthesis, with models like Google's MusicLM and OpenAI's Jukebox demonstrating the ability to generate complex musical compositions. These advancements utilize large-scale deep learning architectures to model long-term musical structure. However, most current research focuses on offline generation—creating music asynchronously after a prompt is received. This latency is acceptable for composition tools but catastrophic for live performance, where musicians rely on millisecond-level interaction loops to maintain synchronization.

2. PROBLEM DEFINITION

The central software engineering challenge in interactive music systems is the Synchronization Gap. In a live setting, the time taken for audio input processing, neural inference, and audio output creates a perceptible delay. Existing systems react after the musician plays, guaranteeing that the accompaniment is always slightly behind the beat. There is a lack of intelligent agents capable of quantifying this system latency and predictively generating audio for a future timestamp, effectively "fast-forwarding" the output to align perfectly with the live performer.

3. PROPOSED APPROACH

This project proposes a Predictive-Adaptive Architecture that treats music generation as a time-critical control system. Instead of minimizing latency, the system anticipates it. The approach consists of three core components:

1. Dynamic Latency Estimation: The system will continuously calculate the "Round-Trip Latency" of the hardware pipeline. This variable serves as a dynamic offset for the generation engine.

2. Predictive Sequence Generation: The Deep Learning model (optimized LSTM or Transformer) will be trained for Next-Step Anticipation. It analyzes the musician's

current trajectory to generate the musical sequence required for the future moment when the audio will actually be heard.

3. Tempo-Locked Synchronization: The system will generate symbolic music (MIDI) time-shifted by the exact duration of the computed latency. This ensures that despite processing delays, the audible output aligns perfectly with the live transient of the human performer.

4. EXPECTED OUTCOMES

The primary deliverable is a functional software prototype capable of accompanying a live musician without perceptible lag. Key success metrics include:

• Synchronization Accuracy: The phase difference between the human beat and the AI beat should be negligible (< 10ms), achieved via the predictive compensation algorithm.

• Adaptability: The system dynamically adjusts its "fast-forward" offset if CPU load changes.

• Musical Coherence: The generated accompaniment maintains harmonic consistency with the input melody.

5. REFERENCES

Jiang, J., et al. (2020). "RL-Duet: Online Music Accompaniment Generation using Reinforcement Learning." Proceedings of the AAAI Conference on Artificial Intelligence.

Huang, C. Z., et al. (2019). "Music Transformer: Generating Music with Long-Term Structure." International Conference on Learning Representations (ICLR).

Briot, J. P., Hadjeres, G., & Pachet, F. (2020). "Deep Learning Techniques for Music Generation – A Survey." Springer Computational Synthesis and Creative Systems.

Roberts, A., et al. (2018). "A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music." International Conference on Machine Learning (ICML).