현대실용음악 제1권 제1호(2025년 8월), 1~3 ISSN 3091-9347
Journal of Contemporary Music , Vol.1, No.1(August, 2025),

마이크로타이밍 기반 인공지능 워킹베이스 생성 모델 연구

A Microtiming-Aware AI Model for Jazz Walking Bass Generation

박한솔(Park HanSol) *

본 연구는 인공지능(AI) 음악 생성 기술의 한계를 극복하고자, 재즈의 핵심 요소인 마이크로타이밍(Microtiming)에 주목하여 워킹베이스의 자동 생성 모델을 개발하였다. 워킹베이스는 리듬, 화성, 멜로디를 동시에 지탱하며 연주자의 리듬 뉘앙스를 가장 잘 반영하는 영역으로, 인간 연주자의 미세한 타이밍의 표현을 AI가 학습하도록 설계하는 것이 본 연구의 핵심이다. 우선, 폴 체임버스, 론 카터, 크리스천 맥브라이드의 연주 데이터를 Melodyne, Sonic Visualiser 등으로 분석하고, ±20ms 수준의 리듬 편차를 유지한 채 데이터 전처리를 진행했다. GPT-2 기반 모델 구조에 Rhythm Deviation Loss와 프롬프트 기반 학습을 결합하여 마이크로 리듬을 재현하도록 설계하였다. 모델의 성능은 정량적 지표(온셋의 밀도, 스윙 비율)와 FMM(Frequency Modulated Möbius) 기반 리듬 곡률의 시각화로 평가되었다. 실험 결과, AI 모델은 ±20ms 이내 리듬 편차를 일정 수준 모방했고, 일부 결과에서는 EGQ(Expressive Groove Quality) 점수에서도 인간 연주자와 유사한 표현력이 나타났다. 본 연구는 정량 리듬 중심의 기존 AI 음악 생성 방식에서 벗어나, 인간적인 표현력을 반영한 AI 창작 시스템으로의 전환을 시도했다는 점에서 학술적, 기술적인 의의가 있다. 향후, 다양한 리듬의 장르 확장과 실시간 협업 연주, 감성 피드백 기반 학습, 교육용 AI 튜터 시스템 등으로 발전 가능성을 제시하였다.

주제어: 인공지능(AI), 재즈 즉흥연주, 워킹베이스, 마이크로타이밍(Microtiming), 음악 생성 모델, GPT-2, 리듬 표현, 데이터 전처리, FMM 모델

This study aims to overcome the limitations of current AI music generation technologies by focusing on microtiming, a core element of jazz, to develop an automatic walking bass generation model. The walking bass plays a crucial role in supporting rhythm, harmony, and melody simultaneously, making it a prime area where a performer's subtle timing nuances are most evident. The core of this research is designing an AI model capable of learning and reproducing these human microtiming expressions.

To achieve this, recordings by renowned jazz bassists Paul Chambers, Ron Carter, and Christian McBride were analyzed using tools such as Melodyne and Sonic Visualiser. Data preprocessing was conducted while preserving rhythm deviations within a ±20ms range. A GPT-2-based model architecture was adopted, incorporating Rhythm Deviation Loss and prompt-based learning to enable the model to emulate micro-rhythmic variations.

The model's performance was evaluated through quantitative metrics (onset density, swing ratio) and rhythm curvature visualization based on the Frequency Modulated Möbius (FMM) model. Experimental results showed that the AI model successfully imitated rhythmic deviations within ±20ms, and in some cases, achieved an Expressive Groove Quality (EGQ) score comparable to human performance.

This research marks a shift from rigid, quantized rhythm generation toward AI-based systems that reflect human-like expressiveness in music creation. It also presents possibilities for future development in areas such as genre diversification, real-time collaborative performance, affective feedback learning, and AI-assisted music education tools.

Key words: Artificial Intelligence (AI), Jazz Improvisation, Walking Bass, Microtiming, Music Generation Model, GPT-2

논문투고일: 2025. 07. 13. 수정접수일: 2025. 08. 22. 게재확정일: 2025. 08. 28.
* 상명대학교 일반대학원 뉴미디어음악학과 Sangmyung University Graduate School, Department of New Media Music. E-mail: able-art@naver.com), 제1저자

I. 서론

재즈 즉흥 연주는 반복적인 학습과 연주 경험을 통해 축적된 패턴을 바탕으로 실시간 음악을 구성하는 고도의 창작 행위이다. 즉흥 연주는 완전히 새로운 음을 즉석에서 만들어내는 것이 아니라, 연주자가 오랜 시간 축적해온 음악적 패턴(프레이즈, 리듬, 코드 진행 등)을 변형, 조합하여 표현한다. 이 과정에서 연주자는 청각적·운동적 절차 기억에 의존하며, 이를 통해 음악적 흐름과 구조를 즉각적으로 형성한다. 이러한 즉흥성은 단순한 패턴 조합으로 설명되지 않는다. 재즈 연주에서는 기보에 없는 섬세한 리듬 조절이 빈번히 나타나며, 이를 마이크로타이밍(Microtiming)이라 한다. 마이크로타이밍은 정박에서 ±10~30ms 정도 앞당기거나 늦추는 미세한 시차로, 연주자의 개성과 음악적 표현력, 특히 그루브(groove)와 루바토(rubato) 감각을 형성하는 핵심 요소다. 재즈의 핵심 구성요소 중 하나인 워킹베이스(Walking Bass)는 리듬·화성·멜로디를 동시에 지탱하며, 곡 전체의 텐션과 흐름을 주도한다. 워킹베이스 연주에서는 각 음이 완벽히 정박에 맞지 않고, 의도적인 앞·뒤 밀기를 통해 자연스러운 흐름과 긴장감을 만들어낸다. 그러나 현재 인공지능(AI) 기반 음악 생성 모델은 주로 정확한 박자와 구조적 일관성에 초점을 맞추고 있어, 이러한 미세한 리듬 뉘앙스를 재현하는 데 한계가 있다. 현존 AI 음악 생성 기술의 대부분은 학습 데이터 전처리 과정에서 양자화(quantization)를 거치며, 이 과정에서 ±20ms 이내의 미세한 리듬 편차가 '불규칙성'으로 간주 되어 제거된다. 그 결과, 생성 음악은 기계적으로 정렬된 박자를 가지지만, 인간 연주자의 표현력과 개성은 반영되지 못한다. 이는 특히 루바토와 스윙 리듬이 중요한 재즈 장르에서 음악적 몰입감과 진정성을 떨어뜨린다. 따라서, 마이크로타이밍은 단순한 부수적 요소가 아닌 재즈 연주의 본질적인 표현 방식이며, 이를 AI 생성 모델에 반영하는 것은 단순한 기술적 개선을 넘어, 인간다운 음악 표현을 구현하는 핵심 과제가 된다. AI 기반 음악 생성 기술은 OpenAI의 MuseNet, Google Magenta 등 다양한 시도로 발전해 왔지만, 주로 음정과 박자의 정확성에 초점을 맞춰왔다. 반면, 인간 연주자의 리듬 표현력과 그루브 감각을 학습·재현하는 시도는 상대적으로 부족하다. 재즈는 리듬 뉘앙스가 음악적 감동을 결정짓는 장르이므로, 마이크로타이밍을 정확히 반영하지 못하는 AI 모델은 '재즈스러움'을 구현하는 데 근본적 한계를 가진다. 특히 워킹베이스는 연주자의 개성을 가장 잘 드러낼 수 있는 소재이자, AI가 리듬 표현력을 학습하기에 적합한 분석 대상이다. 또한, 인공지능의 예술 창작 도입은 단순히 기술적 시도가 아니라, 과거 녹음 기술의 도입처럼 예술의 개념과 경계를 변화시키는 역사적 전환점이 될 가능성이 있다. 녹음 기술이 대중음악의 부흥과 장르 확산을 가져왔듯, AI 음악 생성 모델의 발전은 교육, 공연, 창작 산업 전반에서 새로운 패러다임을 제시할 수 있다.

본 연구의 목적은 재즈 즉흥 연주의 핵심 요소인 워킹베이스를 대상으로, 마이크로타이밍 특성을 반영한 GPT-2 기반 AI 음악 생성 모델을 설계하고 그 성능을 실험적으로 검증하는 것이다. 이를 통해 다음을 달성하고자 한다.

첫째,, 실제 거장 베이시스트의 리듬 데이터 학습한다. Christian McBride, Paul Chambers, Ron Carter 등 거장의 라이브 연주 음원 30곡을 분석, 전처리하여 ±10~30ms 범위의 마이크로타이밍 정보를 유지한 학습 데이터 셋을 구축한다.

둘째,, 데이터 전처리, 정규화 방식의 고도화 한다. 템포, 조성, 마디 수 등을 표준화하면서도 리듬 편차를 보존하는 양자화 전처리 설계한다.

셋째, GPT-2 모델 구조 개선한다. 마이크로타이밍 민감도 기반 손실 함수(Rhythm Deviation Loss) 적용 및 프롬프트 설계로 특정 연주자 스타일 반영한다.

넷째, 실험을 통한 표현력 검증한다. 생성 결과와 실제 연주를 마이크로타이밍 분포, 스윙 비율, Onset 밀도 등 정량·정성 지표로 비교 분석한다.

본 연구는 기존의 양자화 중심 AI 음악 생성 방식이 가지는 기계적 한계를 극복하고, 인간 연주자의 리듬 뉘앙스를 재현할 수 있는 AI 모델 설계 가능성을 제시한다. 이를 통해 향후 재즈를 포함한 리듬 중심 장르에서 자연스럽고 몰입감 있는 AI 연주 구현의 기초를 마련하고자 한다.

II. 이론적 배경

1. 재즈 즉흥연주의 개념 및 음악적 특성

재즈 즉흥연주는 규칙과 창의성이 공존하는 음악적 표현 방식으로, 정해진 형식과 코드 진행 안에서 창의적으로 연주를 구성한다. 이는 단순한 감정 표현이 아니라 절차적 기억에 저장된 청각·운동 패턴을 기반으로 실시간 조합과 조작을 통해 이루어진다. 즉흥연주는 언어처럼 어휘와 문법이 존재하며, 반복 학습을 통해 익힌 패턴과 스케일을 바탕으로 맥락에 맞는 연주가 이루어진다. 스윙 리듬과 블루스 전통, 콜 앤 리스폰스와 같은 요소는 재즈 특유의 공동체성과 표현력을 강화한다. 이는 재즈가 단순한 음악 장르를 넘어 문화적이고 예술적인 가치를 지닌 창작 행위임을 보여준다.

2. 워킹베이스의 역할과 리듬 구조

워킹베이스는 재즈에서 리듬과 화성의 중심 역할을 하며, 일정한 박자 내에서 음을 연결해 음악의 흐름을 형성하는 연주 방식이다. 피치카토 주법을 통해 드럼과 유기적으로 반응하며 매 박자 음을 배치하고, 리듬의 뒷 박자를 강조해 그루브를 만든다. 표면상 단순해 보이나 코드 진행에 따라

즉흥적으로 라인을 생성해야 하므로 고도의 감각과 반복 훈련이 요구된다. 워킹베이스는 단순 반주를 넘어 리듬 해석, 미세한 시간차 조절, 개성 표현의 수단으로 작용하며, 밴드의 전체 흐름과 음악 정체성에 영향을 미친다. 특히 스윙과 블루스 기반 재즈 스타일에서 중요한 역할을 해왔으며, 최근에는 AI 기반 음악 생성 모델의 학습 대상으로도 주목받고 있다.

3. 마이크로타이밍의 개념 및 음악적 특성

마이크로타이밍(microtiming)은 정량화된 박자 구조(quantized metrical grid)를 기준으로, 실제 연주 시점이 기준 박자 선보다 앞서거나 뒤처지는 시간 오차를 의미한다. 본 연구에서는 120BPM 기준의 16분음표 간격으로 grid를 설정하고, DAW에서 추출한 MIDI 데이터의 note-on 이벤트 시점을 기준과 비교하여 Δt = t_actual − t_grid로 계산하였다. Δt는 tick 및 밀리초(ms) 단위로 환산하여 마이크로타이밍 분석의 기초 값으로 활용하였으며, 이는 일반적으로 ±10~30ms 내외의 미세한 시간 차이로 나타난다. 그래서 비트, 템포, 온셋 간의 차이를 통해 정의되며, 정박과 실제 음의 발생 시점 간의 차이를 나타낸다. 재즈 워킹베이스 등에서 마이크로타이밍은 리듬의 긴장과 해방, 스윙감과 그루브를 형성하는 핵심 요소로 작용한다. 특히 베이시스트는 리듬의 흐름을 안정시키면서도 밴드 전체의 유동성을 조율하며, 감정과 개성을 표현하는 중심 역할을 한다. 인공지능 음악 생성의 연구에서도 이 특성은 인간적인 표현을 모사하는 중요한 과제로 주목받고 있다.

4. Transfomer

Transformer는 2017년 Vaswani 등이 제안한 모델로, RNN이나 CNN 없이 오직 어텐션 메커니즘만으로 시퀀스를 처리한다. 핵심은 Self-Attention으로, 입력 시퀀스 내 모든 토큰 간의 관계를 병렬로 학습할 수 있어 학습 속도가 빠르고 장기 의존성 문제를 해결한다. 구조는 인코더와 디코더로 이루어지며, Positional Encoding을 통해 순서 정보를 반영한다. GPT, BERT, T5 등 최신 AI 모델은 모두 Transformer 기반이다. 단점으로는 높은 계산 비용과 긴 시퀀스를 처리 시 메모리 사용량 증가,하며 순차적인 정보 표현의 한계가 있다.

III. 선행 논문과 차별성

1. 선행 논문

이기영 (2024) 음운론적 특성을 기반으로 인공지능 보컬 음성합성의 발음개선 연구 (상명대학교)

김영 (2023), 인공신경망 모델 구조 최적화를 통한 음악적 표현 개선 연구(상명대학교)

장민호, (2012) 라이브 어쿠스틱 뮤직의 디지털로의 에뮬레이션에 관한 연구 : 글렌 굴드의 바흐 골드베르그 변주곡 디지털 데이터를 바탕으로 (상명대학교,)

2. 논문의 차별성

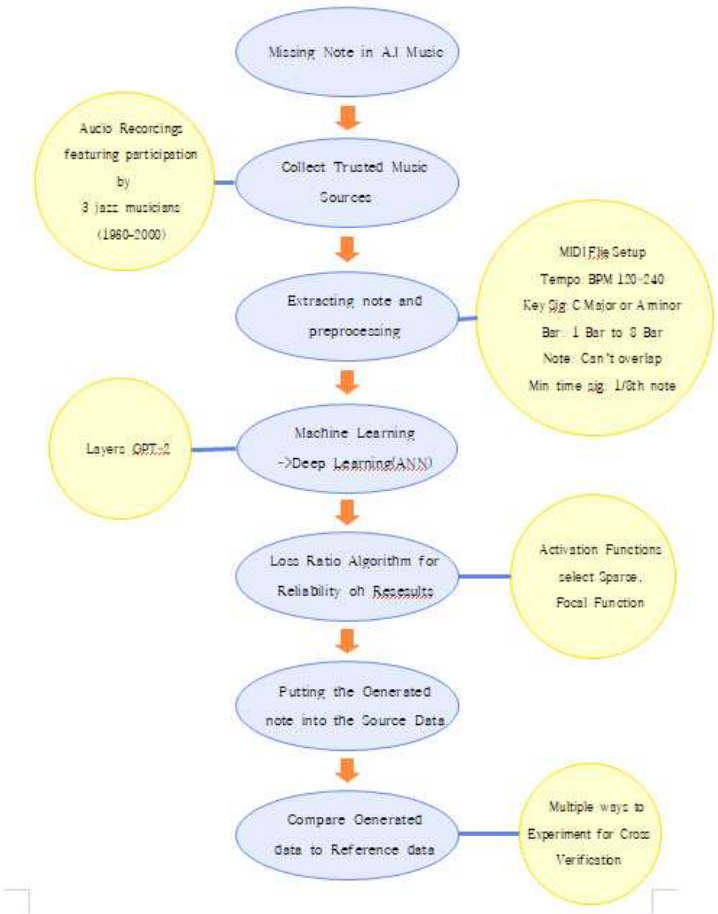
본 연구는 단순한 인공지능 음악 생성 기술을 넘어서, 실제 연주자의 리듬 뉘앙스를 보존하며 생성하는 시스템을 설계했다는 점에서 기존 연구와 뚜렷한 차별성을 갖는다. 기존의 MIR(Music Information Retrieval) 및 딥러닝 기반 음악 생성 연구들은 주로 음정·박자·구조적 일관성에 초점을 두었으며, 재즈 즉흥연주나 워킹베이스 생성을 다룬 연구들도 MuseNet, Magenta, DeepJazz, JazzGAN 등과 같이 멜로디·화성 진행 예측에 치중한 경우가 많았다. 그러나 이러한 모델들은 대부분 MIDI 기반의 양자화된 데이터를 사용해 ±10~30ms 수준의 마이크로 리듬 편차나 그루브 특성을 충분히 반영하지 못했다. 마이크로타이밍에 관한 선행연구(Fernando Benadon, Matthew Butterfield, Maarten Grachten 등)에서는 재즈와 펑크 그루브에서의 시차(time deviation)가 음악적 표현력과 몰입감에 미치는 영향이 반복적으로 강조되었으나, 이를 AI 생성 모델의 입력 데이터와 학습 구조에 직접 반영한 사례는 드물다. 특히 기존 베이스 연주 분석 연구들은 주로 리듬 패턴, 화성 기능, 프레이징 특성에 대한 분석에 머물렀고, 이를 기반으로 한 AI 생성 실험은 리듬의 정박 정확도 평가에 한정되는 경우가 많았다. 이에 본 연구는 기존 연구에서 다루지 못한 워킹베이스의 마이크로타이밍·그루브 특성을 AI 학습 데이터로 정량화·통합하였다는 점에서 차별성을 지닌다. Melodyne, Cubase, Spectralayers, VirtualRise 등 다양한 전문 도구를 활용하여 MIDI 기반이 아닌 오디오 기반의 리듬 데이터를 구축하고, ±10~30ms 수준의 마이크로 리듬 편차, 루바토, 밀고 당김, 뒤로 끌기(lay back) 등 인간적인 표현 요소를 데이터에 반영하였다. 이로써 AI가 단지 타이밍의 정확성을 넘어 감정적인 표현을 학습할 수 있도록 설계하였다. 모델 구조 또한 기존 GPT-2 기반 생성 모델을 개선하여 리듬 편차 민감도 기반 손실 함수와 마이크로타이밍 중심 입력 구조를 도입하였다. 생성된 결과는 악보 비교, 멜로다인 분석, 스펙트럼 시각화, 버추얼라이즈, 온셋 그래프, 엑셀 기반 수치 분석 등 다층적인 방법으로 정량, 정성 평가를 병행하였다. 특히

크리스천 맥브라이드, 론 카터, 폴 체임버스 등 거장 베이시스트의 실연 데이터를 기반으로 연주자별 스타일 편차를 수치화하여 반영함으로써, 단순한 humanize 기능 이상의 스타일 재현이 가능함을 입증하였다. 이러한 접근은 AI 생성 음악 분야에서 타이밍·그루브 요소를 데이터와 모델 구조에 동시에 반영한 사례로, 기존 MIR 및 딥러닝 기반 워킹베이스 생성 연구들과 차별화된다. 나아가 본 연구의 방법론은 다양한 장르와 연주자 스타일에 확장 적용 가능하며, 인간적인 리듬 표현을 구현하는 AI 음악 생성의 새로운 가능성을 제시한다.

IV. 시스템 설계 및 구현

1. 시스템 개발 목표

현재 인공지능 기반의 재즈 즉흥 연주는 인간 연주자의 창작에 비해 완성도에서 다소 부족한 면이 있다. 이에 본 연구는 워킹베이스 생성 모델의 생산성과 표현력을 향상시키고, 정량적 지표를 활용한 결과 비교 및 분석에 중점을 두었다. 연구 데이터는 재즈 아티스트 3명의 1960년대 발표곡 중 4 Feel 계열 곡 30곡을 선정하여 수집하였으며, 스튜디오 녹음을 기준으로 하되 음질이 양호한 라이브 음원도 일부 포함하였다. 최종 데이터는 총 400개의 MIDI 파일, 약 5,000마디 규모로 구성되었다. 모든 음원은 직접 검토하여 잡음을 제거하고, 신뢰도 높은 양자화 데이터를 확보하였다. 전처리 과정에서는 음원을 8마디 단위로 분할한 후, 8분음표 기준으로 MIDI로 변환하고, Melodyne 기반의 정량 분석 자동처리 시스템을 구축하였다. 모델은 GPT-2 구조를 기반으로 하며, 생성 결과는 Melodyne, 스펙트로그램, Virtualize, Excel 그래프 등 다양한 도구를 통해 정량적·정성적으로 평가하였다. 수치 기반 평가의 한계를 보완하기 위해 시각화 도구를 적극 활용하여, 생성된 베이스라인의 표현 특성을 직관적으로 비교 분석하였다. 본 연구에서는 특히 연주자의 리듬 뉘앙스를 정량적으로 분석하기 위해 마이크로타이밍 데이터를 수치화하였다. Swing ratio는 연속된 두 음표의 실제 연주 길이 비율로 계산하였으며, 예를 들어 첫 음이 320tick, 다음 음이 160tick인 경우 2:1 스윙 리듬으로 해석하였다. Δt 값이 평균적으로 +10tick 이상일 경우에는 기준보다 지속적으로 후행하는 laid-back 성향으로, Δt가 ±5tick 이내에서 반복되는 경우에는 groove 성향으로 분류하였다. 이 분석 기준은 Paul Chambers, Christian McBride 등 실존 연주자들의 리듬 특성을 반영하여 구성된 것이다.



<그림1> 시스템 개발 목표- 시스템 개발 다이어그램

2. 데이터 전처리

1) 음원 분리

세 명의 재즈 베이시스트가 참여한 앨범에서 선정한 30곡의 음원을 기반으로, 워킹베이스 라인을 표준화된 MIDI 파일로 변환하는 과정은 신호 잡음을 최소화하고 데이터 정확도를 높이는 데 핵심적인 절차이다 (Müller, 2021). 전체 음원에서 베이스 파트를 정확하게 분리하는 작업은 필수적이며, 이는 전체 연주를 통째로 MIDI로 변환할 경우 타 악기의 간섭으로 인한 오류 발생 가능성이 높기 때문이다. 최근에는 인공지능 기반 소스 분리의 기술이 발달함에 따라 특정 악기의 분리 성능이 크게 향상되었으며, 본 연구에서는 검증된 모델들을 조합하여 활용하였다. 구체적으로는 MDX-Net의 Kuielab_a_bass, Demucs v4(Htdemucs), Demucs v3(Repro_mdx_a_hybrid)를 혼합하여 사용하였다. Kuielab_a_bass는 저주파 대역을 강조해 베이스 영역을 뚜렷하게 추출하는 데 효과적이며, Demucs v4는 음색 보존 능력과 복잡한 대역 분리에 강점을 보인다. Demucs v3는 중음역대의 세밀한 음향 보완에 효과적이다. 이처럼 서로 다른 특성을 가진 모델들을 병합함으로써 단일 모델보다 향상된 베이스 분리 품질을 확보할 수 있었고, 다양한 악기가 혼합된 환경에서도 워킹베이스를 높은 정확도로 추출할 수 있었다. 그 결과, 이후 MIDI 변환 과정 또한 정량적인 분석에 적합한 신뢰도 높은 데이터로 구성되었다.

2) 음원 변환

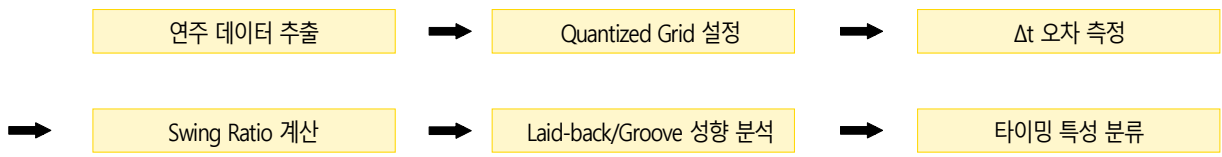
MIDI 기반 접근 방식은 재즈 즉흥 연주의 핵심 요소인 코드 진행, 멜로디, 리듬을 효과적으로 분석하고 학습하기 위해 사용된다. MIDI, MusicXML과 같은 MIDI 데이터를 시퀀스로 변환하여 RNN, LSTM, Transformer 계열 딥러닝 모델의 입력으로 활용하며, 대표적인 사례로는 Google Magenta의 Music Transformer와 OpenAI의 MuseNet이 있다. 이 방식은 노트 온/오프, velocity, duration 등을 시퀀스 화하여 자연어처럼 처리할 수 있다는 장점이 있다. MIDI 데이터의 주요 이점은 첫째, 전처리가 간단하고 파일 용량이 작아 대규모 데이터 셋의 처리에 적합하다는 점이다. 오디오에 비해 잡음 제거, 스펙트럼 변환 등의 복잡한 과정을 거치지 않아도 되며, 학습 리소스를 절약할 수 있다. 둘째, 코드 진행, 템포, 악기 구성 등 조건을 입력해 일관된 음악의 생성이 가능한 조건부 생성에도 유리하다. 셋째, 기존의 음악 소프트웨어와 연동이 쉬워 다양한 실험이 가능하며, 모델 성능 평가나 분석에도 효율적이다.

다만, MIDI 데이터는 연주자의 섬세한 뉘앙스나 배음, 톤 색감 등을 완전히 반영하기 어렵다는 한계를 지닌다. 이를 보완하기 위해 velocity, pitch bend 등의 컨트롤 데이터를 정밀하게 반영하는 방식이 활용된다. 또한, 미디 데이터는 멀티 모달 학습에도 확장이 가능하여, 텍스트 설명이나 비주얼 데이터와 결합해 악기의 스타일을 설명하거나 시각적으로 표현하는 데에도 유용하다. 본 연구는 이러한 MIDI 기반의 특성을 적극 활용하여 재즈 워킹베이스의 구조를 정량적이고 체계적으로 분석하고, 효율적인 학습 데이터로 전환하여 인공지능 모델의 즉흥 연주 생성 능력을 향상시키는 것을 목표로 한다.

학습 데이터는 총 30곡, 약 400개의 MIDI 시퀀스를 기반으로 5,000마디를 구성하였다.모든 데이터는 8마디 단위로 분절한 후, note-on, velocity, duration, Δt, 코드 정보, 템포를 추출하여 토큰화하였다.전처리 과정에서는 MIDI 노트 간 시간 편차를 tick(480ppq 기준)과 ms 단위로 병렬 기록하고, 오버랩, 노이즈 등 오류 데이터를 수동으로 제거하였다.또한 Melodyne과 Cubase, Spectralayers를 병행하여 연주 뉘앙스를 최대한 유지한 정제된 데이터를 구축하였다.

3) 마이크로타이밍 전처리 분석

전처리 과정에서는 마이크로타이밍으로 인해 다양한 변환 오류가 발생하였다. 예컨대, Paul Chambers는 코드 전환 직전 베이스 노트를 앞당겨 연주하는 경향이 있는데, Melodyne은 이를 올바르게 인식하지 못하고 중첩되거나 누락된 노트로 처리하는 오류가 발생하였다. Christian McBride의 빠른 템포 연주에서는 후행 리듬이 시스템상 지연된 입력으로 오해되었고, Ron Carter처럼 느린 템포에서도 섬세한 타이밍 조절이 MIDI 변환 과정에서 비정상적인 간격으로 왜곡되었다. 이러한 오류들은 단순한 감지 문제를 넘어, 정규화 과정에서 실제 연주자의 리듬 뉘앙스를 제거해버리는 한계를 드러낸다. 따라서 단순한 타이밍 정렬만으로는 실연 기반 음악의 리듬 특성을 정밀하게 분석하기 어렵고, 마이크로타이밍을 정량적으로 해석할 수 있는 체계적인 분석 틀이 요구된다.이에 따라 본 연구는 마이크로타이밍 기반 리듬 특성 분석을 위한 절차를 다음과 같은 단계로 정립하였다.



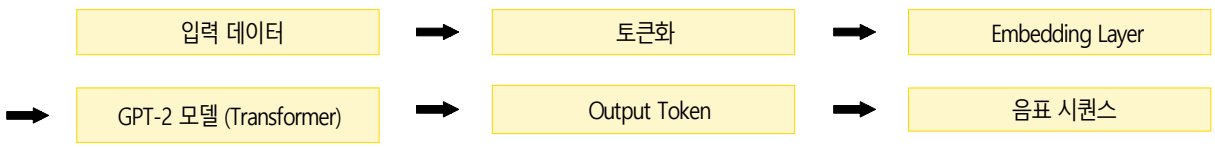
이 분석 흐름은 실연자의 리듬적 뉘앙스를 데이터 기반으로 계량화함으로써, 생성 모델이 단순한 타이밍 복원 이상으로 인간적인 표현력을 학습할 수 있도록 지원한다.

4) 악보 전처리와 정규화 프로세스 분석

본 연구는 원본 악보, Cubase 출력 악보, 조성과 편집을 거친 결과 악보를 비교 분석하여 자동 전사의 정확성과 한계를 밝혔다. 원본 악보는 청음을 통해 작업자가 직접 기록한 데이터를 직관적으로 참조할 점의 역할을 하였고, Cubase 출력의 악보는 재즈의 특성상 리듬을 더블 타임으로 인식하여 8분음표 중심으로 표기되었으며 Gb, Bb, Cb, C가 F#, Bb, B, C로 변형되는 등 오류가 발생하였다. 최종 악보는 조성의 통일(C장조, A단조), 템포를 180bpm으로 정규화하고 화성을 제외하고, 패턴 길이의 기준을 설정하는 등 데이터의 일관성을 확보하였다. 전처리 단계에서는 Spectralayers, Melodyne, Virtualize 등을 활용해서 음정의 오류, 노트의 누락, 지속의 시간과 불일치 등을 분석하고 정규화의 알고리즘을 개선하였다. 학습 단계에서는 음의 시작과 끝, 속도, 코드의 진행, 리듬의 패턴, 구조적인 문맥 등을 메타의 정보를 포함한 복합적인 데이터를 모델에 입력하였다. 또한 Logic Pro의 기능을 활용해서 Downbeat의 정렬을 수행하고 Melodyne Studio로 MIDI의 변환과 오류의 수정을 진행하였다. 결과 분석에서는 조건부 생성과 스타일의 트랜스퍼 역량을 평가하였고, AI 모델이 실제 연주와 유사한 리듬과 그루브를 반영할 수 있음을 확인하였다.

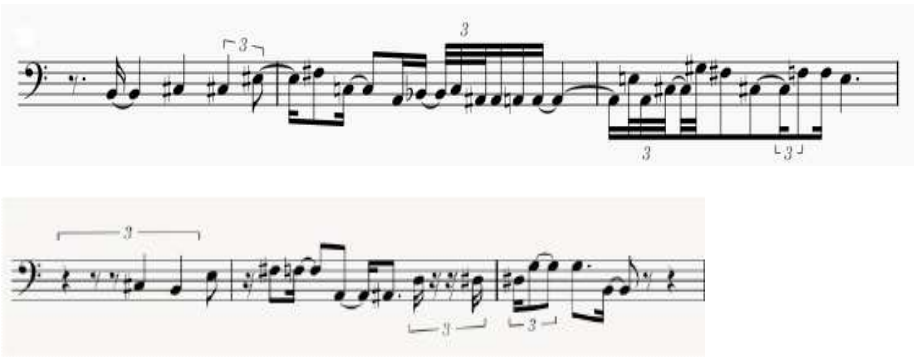
3. 모델 구조

본 연구는 OpenAI의 GPT-2 구조를 기반으로 워킹베이스 리듬 시퀀스를 생성하는 딥러닝 모델을 설계하였다. 기존의 자연어처리 구조를 음악 시퀀스에 응용한 것으로, 입력은 정규화된 마이크로타이밍 데이터, 코드 진행, 템포 등으로 구성되며, 출력은 마디 단위 워킹베이스 시퀀스이다. 특히, note-on 시간 간격과 Δt 값을 포함한 특성값을 토큰으로 변환하여, GPT-2의 attention 기반 구조로 리듬 패턴을 학습하도록 하였다. 아래는 본 모델의 구조 도식이다.

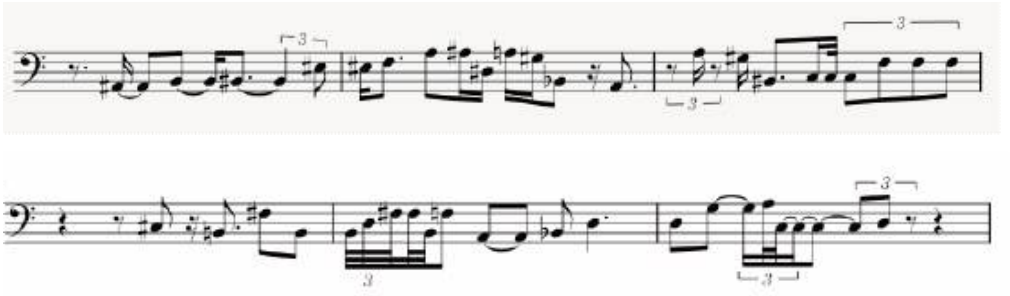


5. 마이크로타이밍 기반 전처리 분석

(1) < A-1> ‘Too Close for Comfort- Christian McBride Trio’ (대표 데이터)

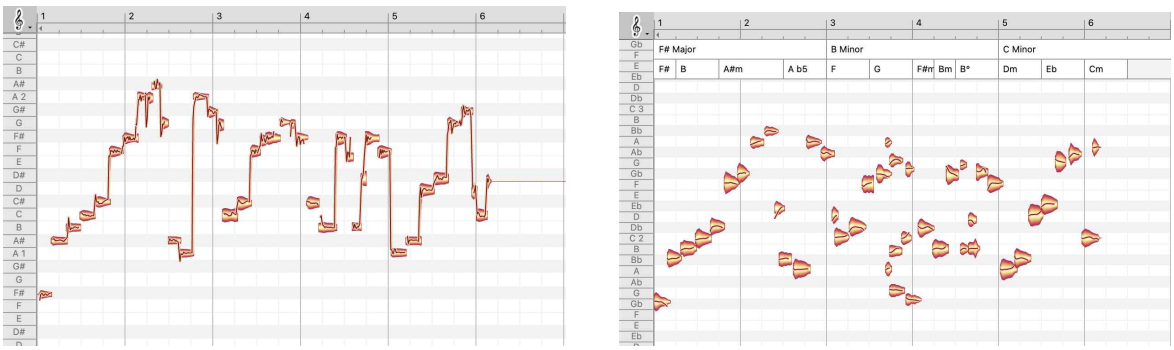


<그림 2> A-1 의 오리지널, 음원의 악보



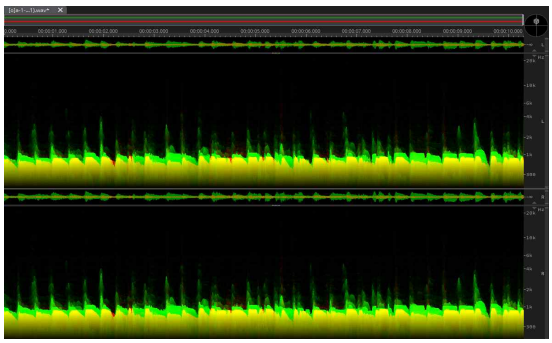
<그림 3> A-1 의 양자화된, 음원의 악보

첫 번째 악보 (오리지널 음원)는 셋잇단음표, 붓점, 당김음 등 복잡한 리듬의 구조와 마이크로타이밍이 포함되어 있어 연주자의 개성과 표현력이 뚜렷하게 드러난다. 반면, 두 번째 악보 (양자화된 음원)는 리듬이 정확히 정렬되어 일관성과 안정감은 있지만, 자연스러운 리듬의 뉘앙스와 표현력은 감소하였다. 결과적으로 양자화된 악보는 구조적인 정확성은 높지만, 오리지널 음원의 생동감과 리듬의 긴장감은 부족한 경향이 있다.



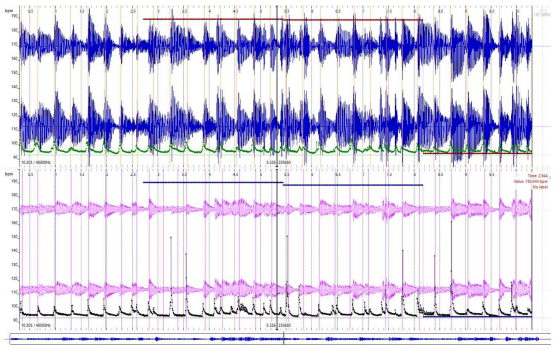
<그림 4,5> A-1의 오리지널, 양자화된 음원의 멜로다인

<그림4>의 왼쪽은 양자화된 오리지널의 음원으로, 리듬과 음의 시작점이 명확하게 정렬되어 정교한 마이크로타이밍을 보인다. 반면 오른쪽의 생성된 음원은 슬라이드 등 자연스러운 음정의 변화를 보이며, 박자 간격이 불규칙하고 뉘앙스가 풍부하게 표현된다. 특히 저음역대에서 Melodyne은 140ms 수준의 시점 오류를 반복적으로 발생시켜, ±30ms 이내가 요구되는 재즈 워킹베이스의 리듬 표현에는 부정확한 결과를 보였다. 이를 보완하기 위해 버추얼라이즈를 활용해 오디오와 MIDI 간 시간 차이를 정량적으로 분석하였고, Melodyne의 오류가 선행과 지연 형태로 혼재되어 있음이 확인하는 것이다. 결과적으로, 향후 AI 학습에는 Melodyne 외 다중 분석 툴의 병행 사용이 필요하며, 고해상도 마이크로타이밍 분석이 AI의 음악적인 자연스러움을 높이는 것이 핵심이다.



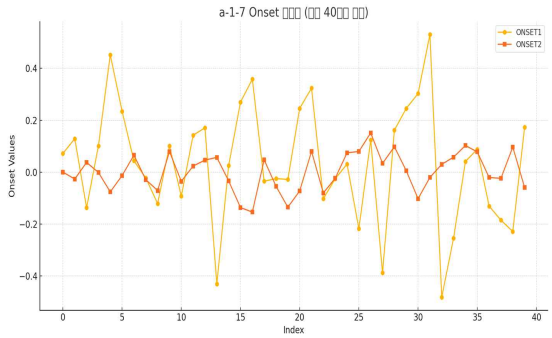
<그림6> A-1의 오리지널, 양자화된 음원의 레이어드한 스펙드로그램

오리지널 음원 (녹색)은 각 음의 시작점이 명확히 드러나며, 일정한 박자의 간격 속에서도 미세한 타이밍의 편차가 존재한다. 이는 연주자의 의도적인 타이밍의 조절로, 스윙이나 싱코페이션과 같은 리듬 요소를 형성하여 음악의 그루브와 표현력을 강화한다. 반면, 양자화된 음원(빨간색)은 규칙적이고 정밀한 타이밍의 분포를 보이며, 자연스러운 리듬의 긴장감이나 해소는 상대적으로 감소한다.

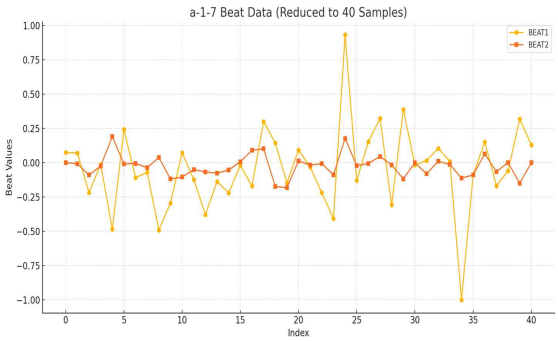


<그림7> A-1의 오리지널, 양자화된 음원의 레이어드한 버츄얼라이즈 온셋, 비트, 템포 비교 파형

오리지널 음원 (파란색)은 음의 시작점이 박자의 기준선에서 미세하게 앞뒤로 벗어나며, 연주자의 개성과 표현력이 반영된 자연스러운 마이크로타이밍을 보여준다. 음의 길이와 강도도 일정하지 않아 음악적인 그루브와 생동감을 형성한다. 반면, 양자화된 음원 (보라색)은 일정한 간격과 길이로 정밀하게 정리되어 있으며, 마이크로타이밍과 다이나믹의 변동이 거의 없어서 기계적이고 규칙적인 연주 스타일을 나타낸다.



<그림 8> A-1의 오리지널, 양자화된 음원의 온셋 비교 그래프



<그림 9> A-1의 오리지널, 양자화된 음원의 비트 비교 그래프

온셋 그래프에서 오리지널 음원 (노란색)은 시작 타이밍의 변동 폭이 크고, 감정적 표현이 강하게 드러난다. 양자화된 음원 (주황색)은 일정한 타이밍을 유지하며 정밀성과 일관성을 강조한다. 비트 그래프에서도 오리지널 음원은 자연스럽고 유기적인 리듬의 변화를 보여주며 긴장감을 형성하지만, 양자화된 음원은 일정하고 안정된 리듬의 패턴으로 예측 가능성을 보여준다.

<그림 10> A-1의 오리지널, 양자화된 음원의 마이크로타이밍 비교 그래프

마이크로타이밍 그래프에서 오리지널 음원은 박자가 앞뒤로 크게 흔들리는 변동 폭을 보여 자연스럽고 표현력 있는 연주를 나타낸다. 반면, 양자화된 음원은 거의 직선에 가까운 형태로 일정한 타이밍을 유지하며, 기계적인 정확성과 일관성을 강조한 리듬의 특성을 보인다.

<표1> Microtiming 평균, 표준편차

<Microtiming 1> 평균	<Microtiming1> 표준편차	<Microtiming 2> 평균	<Microtiming2> 표준편차
0.000125	0.100034	-0.0000000128	0.000763

Microtiming1은 평균값이 약간 늦은 양수이며, 표준편차도 커서 연주자가 의도적으로 느리게 또는 자유롭게 리듬을 조절한 흔적이 뚜렷하다. 이는 레이백 스타일이 즉흥적이고 표현적인 연주를 반영한다. 반면, Microtiming2는 평균값이 거의 0에 가까운 음수이고, 표준편차도 매우 작아, 기준 타이밍에 정확히 맞춘 정밀하고 일관된 연주 성향을 보여준다.

	A	B	C	D	E	F
1	ONSET1	BEAT1	MICROTIMING1	ONSET2	BEAT2	MICROTIMING2
2	0.0723274000	0.0734890000	-0.0011616000	0.0003577470	0.0003395080	0.0000182390
3	0.0709871000	0.0721699000	-0.0011828000	0.0004802940	0.0004621740	0.0000181200
4	0.0791907000	0.0803905000	-0.0011998000	0.0007389780	0.0007194280	0.0000195500
5	0.0786574000	0.0793688000	-0.0007114000	0.0009019370	0.0008845330	0.0000174040
6	0.0867683000	0.0868220000	-0.0000537000	0.0010011200	0.0009834770	0.0000176430
7	0.0862745000	0.0861336000	0.0001409000	0.0010263900	0.0010099400	0.0000164500
8	0.0917928000	0.0916145000	0.0001783000	0.0011073400	0.0010913600	0.0000159800
9	0.0919538000	0.0927016000	-0.0007478000	0.0010902900	0.0010767000	0.0000135900
10	0.0954499000	0.0971425000	-0.0016926000	0.0010007600	0.0009881260	0.0000126340
11	0.0960995000	0.0981951000	-0.0020956000	0.0009410380	0.0009294750	0.0000115630
12	0.0986571000	0.1008280000	-0.0021709000	0.0010486800	0.0010378400	0.0000108400
13	0.1011260000	0.1029240000	-0.0017980000	0.0012707700	0.0012605200	0.0000102500
14	0.1025930000	0.1042580000	-0.0016650000	0.0015124100	0.0015033500	0.0000090600
15	0.1056900000	0.1076100000	-0.0019200000	0.0017064800	0.0016980200	0.0000084600
16	0.1057640000	0.1084100000	-0.0026460000	0.0017759800	0.0017676400	0.0000083400
17	0.1083400000	0.1119080000	-0.0035680000	0.0017511800	0.0017416500	0.0000095300
18	0.1087530000	0.1125140000	-0.0037610000	0.0016365100	0.0016267300	0.0000097800
19	0.1113470000	0.1154070000	-0.0040600000	0.0015556800	0.0015451900	0.0000104900
20	0.1113660000	0.1161540000	-0.0047880000	0.0015982400	0.0015876300	0.0000106100

<표2> A-1의 오리지널, 양자화된 음원의 온셋, 비트, 마이크로타이밍 비교 데이터

오리지널 음원 (Onset1, BEAT1, Microtiming1)은 기준 타이밍보다 앞서거나 늦게 연주되는 경향이 뚜렷하며, 전반적으로 변동 폭이 크다. 반면 양자화된 음원(Onset2, BEAT2, Microtiming2)은 기준에 매우 근접한 정밀한 타이밍을 보여주며, 일정하고 정확한 리듬의 특성을 지닌다.

6. 마이크로타이밍 기반 전처리 결과

본 연구는 인공지능 재즈 워킹베이스 생성 모델의 성능 저하의 원인을 데이터 전처리의 부정확성에서 찾고, 이를 개선하기 위해 마이크로타이밍 기반의 정밀 전처리를 도입하였다. 오리지널 오디오, Melodyne, Cubase 악보, 양자화된 MIDI 등 다양한 데이터를 비교, 분석하고, Virtualrise, Spectralayers 등 도구를 활용해 Onset 오류, 타이밍 왜곡, 음표 누락 현상을 계량화하였다. 특히 Melodyne은 저주파 영역에서 Onset 인식 오류를 반복적으로 보였으며, MIDI 양자화는 재즈 특유의 리듬 표현을 제거해 모델의 표현력이 저하되었다. 그러나 마이크로타이밍 전처리를 적용한 후 평균 편차와 표준편차가 크게 감소하였다 음표 누락률과 MIDI-Onset 편차도 현저히 줄어드는 개선이 나타났다. 이는 리듬의 자연스러움과 음악성 향상에 기여 되었으며, GPT-2 기반 모델은 코드 일관성, 템포의 적응력, 즉흥성, 협업 안정성에서 뚜렷한 발전을 보였다. 본 연구는 고정밀 리듬 기반의 전처리가 AI 음악의 생성 품질 향상의 핵심임을 입증하였다.

V. 1차 시스템 구축 및 실험

1. 1차 모델의 설계

본 연구에서는 인공지능(AI)이 재즈 베이스 워킹 라인을 생성할 수 있도록 다양한 실험을 설계하고 수행하였다. 실험은 총 5회에 걸쳐 진행되었고, 인공지능 모델의 성능을 평가하기 위해 학습 데이터, 프롬프트 설정, 학습의 횟수, 샘플링의 온도 등의 변수를 체계적으로 조정하였다. 각 실험에서는 AI가 특정 연주자의 스타일을 얼마나 충실히 반영하는지, 그리고 음악적으로 얼마나 자연스럽게 재현할 수 있는지를 검토하였다.

1차 실험에서는 AI가 특정 연주자의 스타일을 학습하고 반영할 수 있는지를 분석하였다. 1-1 실험에서는 크리스천 맥브라이드(Christian McBride) 스타일의 워킹베이스 라인을 생성하도록 설정하였다. 이를 위해 "G7 walking line in the style of Christian McBride, tempo between 100 and 120, note duration over 17 seconds"라는 프롬프트를 활용하여 A-1 데이터 셋을 학습시켰다. 분석 결과, AI는 기본적인 워킹베이스 리듬 패턴을 형성하는 데에 성공하였으며, 맥브라이드 특유의 강한 어택과 크로매틱 진행을 일부 반영하였다. 그러나 스윙 리듬의 변형과 마이크로타이밍의 조절이 부족하여 실제 연주의 유기적인 흐름을 완전히 재현하지는 못하는 한계가 드러났다.

이어진 1-2 및 1-3 실험에서는 폴 체임버스(Paul Chambers) 스타일의 G메이저 워킹베이스의 라인을 생성하도록 하였다. "Generate a walking bass line in the key of G major, in the style of Paul Chambers."라는 프롬프트를 기반으로, MIDI 기반 데이터 셋을 활용하여 AI가 해당 스타일을 학습할 수 있도록 하였다. 그 결과, AI는 기본적인 워킹베이스 패턴을 형성하였으나, 기존 학습 데이터에서 반복적으로 나타난 패턴을 그대로 재현하는 경향이 강하게 나타났다. 이는 특정 연주자의 음악적인 특성을 모방하기 위해서는 단순한 코드 예측 이상의 고차원적 패턴 인식의 능력이 필요함을 시사한다.

1-5 실험에서는 론 카터 (Ron Carter) 스타일의 베이스라인을 생성하도록 하였고, 프롬프트에는 "Generate a walking bass line in the style of Ron Carter, using chromatic approach notes and syncopation."이라는 문장을 사용하였다. 분석 결과, AI는 크로매틱 어프로치 노트와 싱코페이션 기법을 일정 부분 반영하였으나, 론 카터 특유의 유려한 프레이징과 불규칙적 리듬의 패턴까지는 재현하지 못했다. 이는 AI가 인간 연주자의 즉흥성을 완전히 모방하기에는 여전히 한계가 존재함을 의미한다.

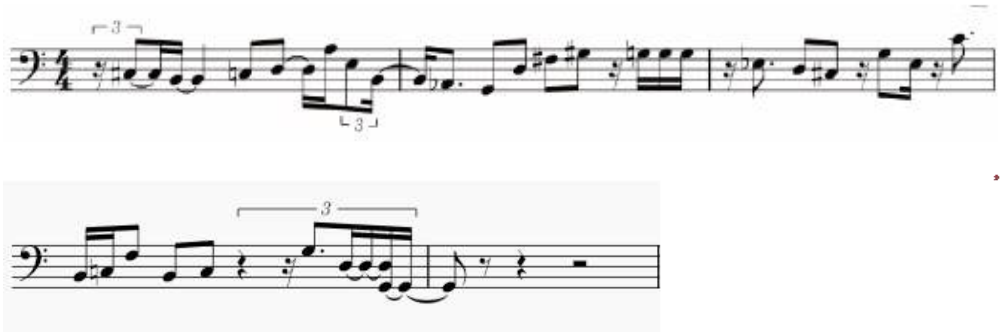
1-6 실험에서는 샘플링 온도(Sampling Temperature)를 조절하며 AI가 생성하는 결과물의 즉흥성과 예측 가능성을 비교하였다. 샘플링 온도를 낮추었을 때에는 정형화된 패턴이 증가하고 예측 가능성이 높아졌고, 즉흥성이 감소하는 경향이 나타났다. 이는 자연스러운 즉흥 연주를 생성하기 위해 일정 수준 이상의 샘플링 온도를 설정할 필요가 있음을 시사한다.

1-7 및 1-8 실험에서는 학습 횟수를 500회로 설정하고, 각각 크리스천 맥브라이드와 론 카터 스타일의 연주 생성을 시도하였다. 결과적으로 학습 횟수가 증가함에 따라 스타일 반영 능력은 향상되었지만, 특정 스타일에 과도하게 치우치는 과적합 (overfitting) 현상도 관찰되었다. 이는 다양한 코드 진행과 템포 변화에 유연하게 대응하는 능력이 제한될 수 있음을 보여준다.

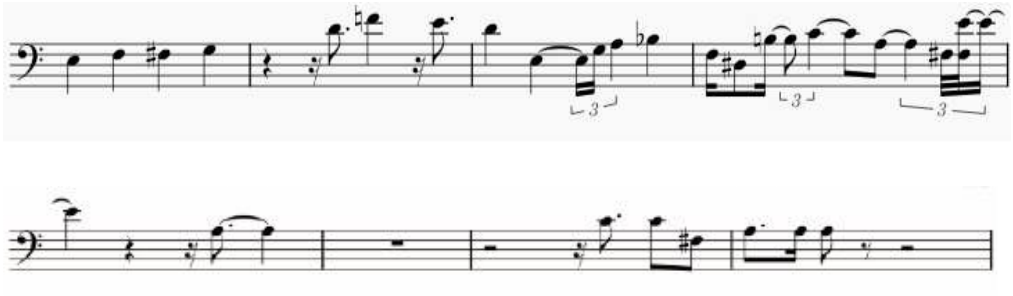
또한 초기 생성 모델에서는 전처리의 부정확성으로 인해 리듬 표현이 제한되었으며, Melodyne 기반 MIDI 변환 과정에서 어택 인식 오류 및 타이밍 왜곡 현상이 발생하였다. 이러한 오류는 이후 마이크로타이밍 기반의 정밀 보정을 통해 일부 개선되었고, 데이터 양자화와 정제의 중요성이 강조되었다. 이 과정을 통해 AI 생성 음악의 리듬 자연스러움과 즉흥성의 표현이 향상할 수 있었다.

2. 마이크로타이밍 분석

(1) <1-7> (대표 데이터)

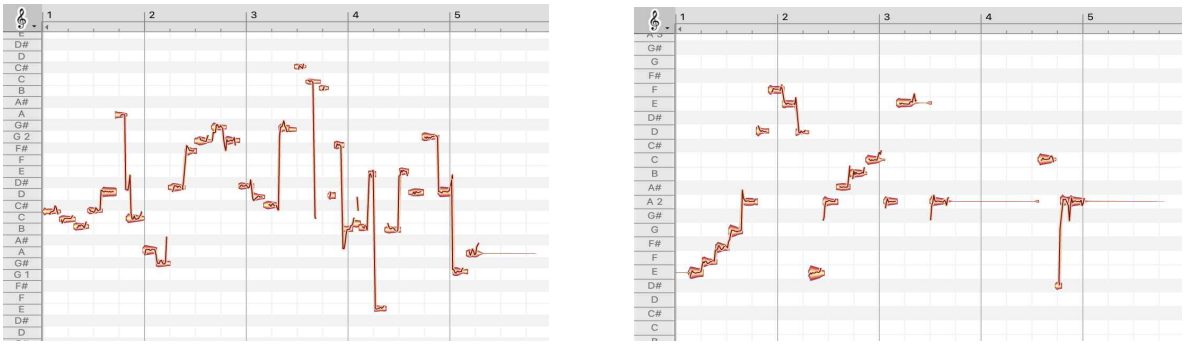


<그림11> 1-7의 오리지널 음원의 악보



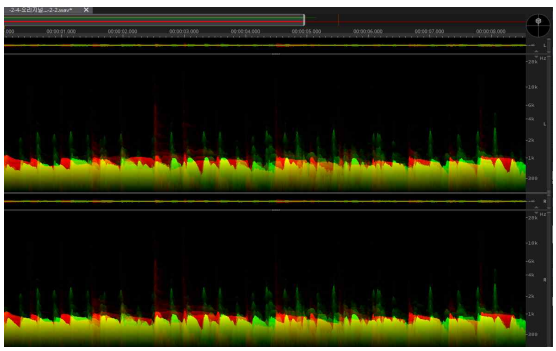
<그림12> 1-7의 생성된 음원의 악보

첫 번째 악보(오리지널 음원)는 셋잇단음표, 스타카토, 레가토 등 복잡하고 유동적인 리듬의 표현이 특징이며, 마이크로타이밍의 변동성이 크다. 반면 두 번째 악보(생성 음원)는 리듬과 음의 이동이 단순화되어 있고, 음표 길이가 균등하며 박자 변동이 적어 자연스러운 연주의 느낌이 감소하는 경향이 있다.



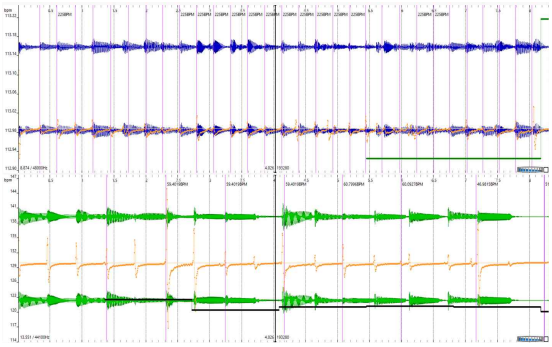
<그림13,14> 1-7의 오리지널, 생성된 음원의 멜로다인

<그림13> 왼쪽의 오리지널 음원 데이터는 박자의 변동성이 커 보이며, 빠른 음들의 지속 시간이 다양하게 나타난다. <그림14> 오른쪽은 생성된 음원 데이터에서는 노트 길이가 일정한 경향을 보이며, 박자의 흔들림이 적어 보인다. 오리지널 음원 데이터는 루바토가 포함되며, 연주자의 감각적인 표현이 포함된다. 반면, 생성된 음원은 비교적 정량화된 박자의 패턴을 유지한다.



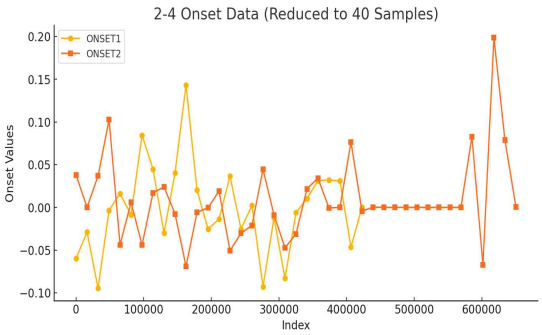
<그림15> 1-7의 오리지널, 생성된 음원의 레이어드한 스펙드로그램

오리지널 음원은 중, 저음역에서 에너지가 고르게 분포되고 강, 약 조절이 뚜렷하며, 스윙감과 리듬의 변동성이 포함되어 있다. 반면 생성된 음원은 특정 대역에서 일정한 에너지를 유지하고, 박자와 어택이 기계적으로 일정해 자연스러움이 떨어지는 경향이 있다

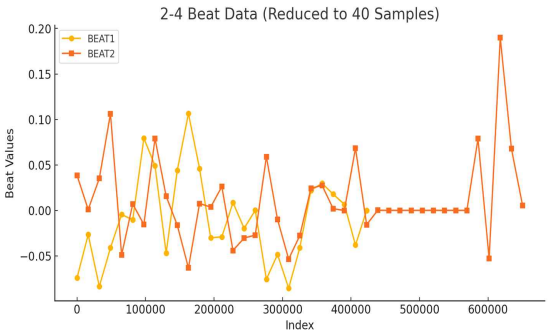


<그림16> 1-7의 오리지널, 생성된 음원의 버츄얼라이즈 온셋, 비트, 템포 비교 파형

오리지널 음원은 박자의 미세한 흔들림과 자연스러운 강세의 변화를 통해 스윙감과 음악적인 표현력이 강조된다. 반면, 생성된 음원은 리듬과 강세가 균일하게 유지되어 정확성은 높지만 자연스러움과 다이내믹 표현은 부족한 경향이 있다.

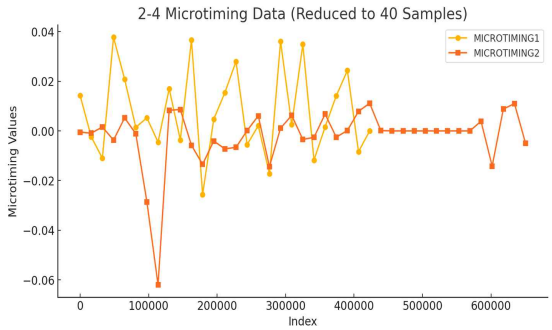


<그림17> 1-7의 오리지널, 생성된 음원의 온셋 비교 그래프



<그림18> 1-7의 오리지널, 생성된 음원의 비트 비교 그래프

오리지널 음원은 Onset과 비트에서 변동성과 유연성이 커서 자연스러운 리듬 감성과 흐름을 보여준다. 반면 생성된 음원은 Onset과 박자가 일정하게 정렬되어 기계적이고 안정적인 리듬 특성을 가진다.



<그림19> 1-7의 오리지널, 생성된 음원의 마이크로타이밍 비교 그래프

오리지널 음원은 박자의 변동성이 크고, 피크 값이 여러 곳에서 발생한다. 자연스러운 연주의 흔들림이 반영된다. 생성된 음원의 값은 변화가 상대적으로 작고 일정한 패턴을 유지한다. 박자가 안정적이며 기계적으로 정렬된 형태를 보인다. 중간 구간에서는 거의 일정한 값으로 유지되었다.

<표 3> Microtiming 평균, 표준편차

<MICROTIMING 1 평균>	<MICROTIMING 1 표준편차>	<MICROTIMING 2 평균>	<MICROTIMING 2 표준편차>
0.00002709	0.01962	0.0000001292	0.00910

오리지널 데이터 평균값이 0.00002709로 생성된 데이터보다는 상대적으로 크다. 표준편차 0.01962로 변동성이 크며, 미세한 리듬의 흔들림이 포함된다.

생성된 데이터 평균값이 0.0000001292로 거의 0에 가깝고, 리듬이 정렬된다. 표준편차 0.00910으로 원본보다 변화가 적다. 박자가 일정하게 정돈되었으며, 기계적으로 보정된 패턴을 보인다.

<표 4> 1-7의 오리지널, 생성된 음원의 온셋, 비트, 마이크로타이밍 비교 데이터

	A	B	C	D	E	F
1	ONSET1	BEAT1	MICROTIN	ONSET2	BEAT1	MICROTIMING1
2	-0.05981	-0.07407	0.014252	0.037926	0.038425	-0.0004995
3	-0.05997	-0.0741	0.01413	0.038044	0.038584	-0.00054
4	-0.06006	-0.07416	0.014099	0.038162	0.038717	-0.0005548
5	-0.06012	-0.07419	0.014069	0.038284	0.038842	-0.0005578
6	-0.06018	-0.07422	0.014038	0.038485	0.039062	-0.0005764
7	-0.06012	-0.07419	0.014069	0.038675	0.039302	-0.0006272
8	-0.06012	-0.07425	0.01413	0.03875	0.039424	-0.0006748
9	-0.06015	-0.07425	0.014099	0.038803	0.039472	-0.000669
10	-0.06015	-0.07416	0.014008	0.038854	0.03949	-0.0006356
11	-0.06021	-0.07407	0.013855	0.038827	0.039467	-0.0006393
12	-0.06021	-0.07394	0.013733	0.038775	0.039426	-0.0006503
13	-0.06024	-0.07391	0.013672	0.038722	0.039316	-0.000594
14	-0.06018	-0.07385	0.013672	0.038671	0.039143	-0.0004719
15	-0.06009	-0.07379	0.013702	0.038631	0.03899	-0.0003594
16	-0.06006	-0.07379	0.013733	0.038471	0.038781	-0.00031
17	-0.06	-0.07364	0.013641	0.038226	0.038518	-0.0002921
18	-0.05991	-0.07352	0.013611	0.038019	0.038284	-0.0002651
19	-0.05981	-0.07339	0.01358	0.037778	0.038021	-0.0002428
20	-0.05963	-0.07318	0.01355	0.037454	0.037704	-0.0002505

Onset2 (생성된 음원)은 정렬된 박자 패턴을 유지한다. 반면, 오리지널 데이터 (Onset1)는 박자가 유동적으로 변화하며, 자연스러운 타이밍의 특성을 보인다.

BEAT1 (오리지널 음원)은 비트 간격이 일정하지 않으며 유연하다. BEAT2 (생성된 음원)은 일정한 간격으로 정리된다. Microtiming1 (오리지널 음원)은 연주자의 리듬 유연성이 살아있어 자연스러운 타이밍의 변화를 포함한다. Microtiming2 (생성된 음원)은 박자가 균일하게 정렬되었으며, 리듬의 변화가 적다.

총 500 epoch에 걸친 학습을 수행하였으며, 훈련 손실은 학습 초반 10 epoch 이내에서 급격히 감소하여 0.76 이하로 하락하였고, 이후에도 점진적인 하락을 유지하며 학습 후반에는 약 0.01 수준으로 수렴하였다. 이는 모델이 훈련 데이터에 적합 되었음을 시사한다. 검증 손실은 2 epoch부터 측정되었으며, 약 50 epoch 이후부터는 완만한 상승세로 전환되었다.

2. 실험 분석 및 결과

1차 생성 모델의 음악적인 흐름은 유지를 했지만 마이크로타이밍과 즉흥성에서 인간의 연주와는 큰 차이를 보였다. 이는 전처리 과정에서 템포의 변화가 반영되지 않아 Onset과 지속적인 시간이 불균형하게 처리된 것이 원인 이다. 특히 BPM 정규화가 되지 않아 리듬의 표현에 왜곡이 발생했고, 향후에 템포가 독립적으로 전처리와 오류가 난 데이터에 제거가 필요하다.

<표 5> 1-7 Epoch별 손실 지표 및 학습 시간 요약표

	A	B	C	D
1	epoch	avg_loss	eval_loss	training_time (s)
2	1	2.02075		2.535243273
3	2	1.0058875	1.218222499	3.491038799
4	3	0.8312125	0.8756722212	3.308761835
5	4	0.779925	0.8581299782	3.463993788
6	5	0.765575	0.8569941521	3.33911562
7	6	0.7349875	0.8515064716	3.529251337
8	7	0.7183625	0.8491945267	3.350130558
9	8	0.706825	0.8658666015	3.582276344
10	9	0.680825	0.8684809208	3.381706238
11	10	0.6764555556	0.8845412731	3.627158165
12	11	0.6386375	0.8845412731	3.375264406
13	12	0.63105	0.9392433763	4.34300828
14	13	0.5919625	0.9446037412	3.418727875
15	14	0.5550625	1.032917142	3.668245554
16	15	0.5007375	1.07478416	3.479063749
17	16	0.4525625	1.163337588	3.528882742
18	17	0.396575	1.271915674	3.27446413
19	18	0.3434625	1.328704476	3.614350796
20	19	0.2764875	1.420096755	3.370171309

VI. 2차 시스템 구축 및 실험

1. 2차 모델의 실험 설계

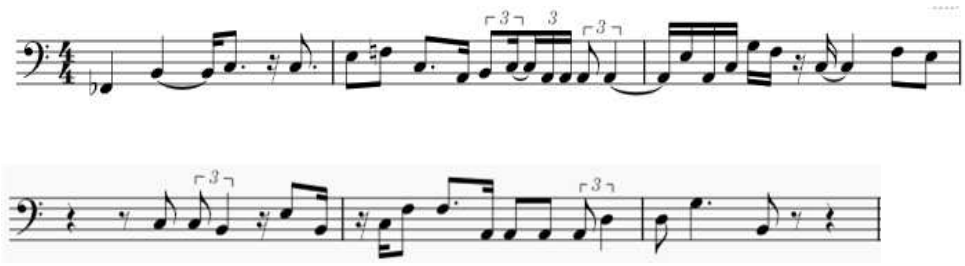
본 연구의 1차 인공지능 생성 모델은 정량적 평가와 정성적 청취 분석을 통해, 재즈 워킹베이스의 자연스러운 리듬과 즉흥성을 재현하는 데 한계를 보였다. 특히 Melodyne 기반 전처리 과정에서 시간 정보가 절대값(ms)으로 고정 처리되어 각 곡의 템포의 차이를 고려하지 못했으며, 이로 인해 모델 학습 시 리듬의 편차가 왜곡되거나 과도하게 보정되는 현상이 발생했다. 이러한 구조적인 오류는 재즈 연주에서 표현의 일부로 작용하는 마이크로타이밍의 특성을 반영하지 못한 전처리 방식에 기인하는 것으로 분석된다. 이에 따라 2차 생성 모델은 전처리 방식 개선의 핵심 전략으로 BPM 정규화 기준(168~176 범위)을 명시적으로 프롬프트에 포함하고, 학습 및 생성 전 과정에 이를 적용하였다. 이 방식은 전체 데이터 셋의 템포 스케일을 균일하게 정렬하고, 음표 간 상대적 타이밍(Onset Distance)이 템포 변화의 영향을 받지 않도록 조정함으로써 모델이 구조적 리듬 단위를 안정적으로 학습할 수 있도록 돕는다. 이로써 마이크로타이밍을 포함한 미세 리듬 표현의 왜곡 없이 자연스러운 생성이 가능하게 되었다. 또한 학습 데이터 구성 시 오리지널 음원의 마이크로타이밍의 평균 및 표준편차의 분석 결과를 반영하여, ±30ms 이내의 리듬 편차를 유지하는 방식으로 시간 축의 정렬을 수행하였다. 이러한 방식은 인간 연주자의 리듬 해석에 근접한 자연스러운 표현을 유지하면서도, 기준 BPM을 중심으로 모델 학습의 일관성을 높이는 기반이 되었다.

2차 모델은 총 500회의 학습을 수행하였으며, 2-1차 모델은 1000회, 2-2차는 500회로 반복 학습을 진행하였다. 각 모델의 프롬프트는 다음과 같다. 2-1차: "Jazz Bass Walking Line in G major, Resembling Christian McBride, Lasting for 60 seconds", 2-2차: "Jazz Bass Walking line in G major, Resembling Paul Chambers, Lasting for 60 Seconds", 2-3차: "Jazz Bass Walking line in G major, Resembling Ron Carter, Lasting for 60 Seconds",

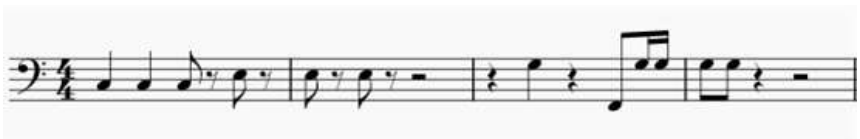
2-4차는 위 세 베이시스트의 스타일을 통합하여 학습하였다. 2-5차는 재학습 기반 모델로 동일한 조건(코드 진행, 템포) 하에서 1차 모델 대비 개선된 리듬의 안정성, 템포의 일관성, 즉흥성의 측면에서 긍정적인 평가를 받았다. 특히 2-5차 모델은 BPM에 따른 상대적 시간의 간격을 명확히 학습함으로써, 특정 연주자의 타이밍 스타일을 더욱 정밀하게 반영할 수 있음을 확인하였다. 이는 단순히 데이터의 양을 늘리는 접근보다, 구조적인 전처리와 프롬프트 설계가 모델의 표현력에 결정적인 영향을 미침을 보여준다. 향후에 모델 구조와 전처리 알고리즘 간 상관성을 심층 분석하고, 연주자 스타일의 재현 정확도 및 실연 가능성에 대한 평가를 통해 인공지능 기반 재즈 생성의 실용성과 예술성을 동시에 확보하는 방향으로 연구를 확장할 수 있을 것이다.

2. 마이크로타이밍 분석

(1) 2-1 (대표 데이터)



<그림20> 2-1의 오리지널 음원의 악보



<그림21> 2-1의 생성된 음원의 악보

오리지널 악보는 셋잇단음표와 16분음표의 미세한 시간 차이를 통해 스윙감, 밀고 당기기, 레이백 등 실제 연주의 리듬적인 표현이 살아 있다. 반면 생성된 악보는 모든 음표가 정확히 박자에 맞춰져 있어 리듬의 흔들림이나 표현력이 부족하다.

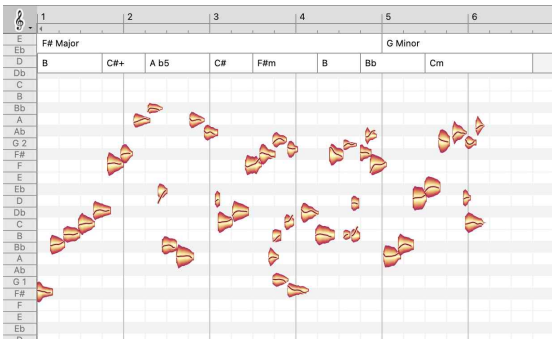


그림 <22> 2-1의 오리지널

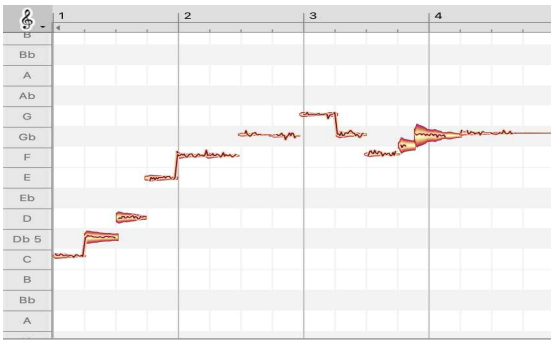
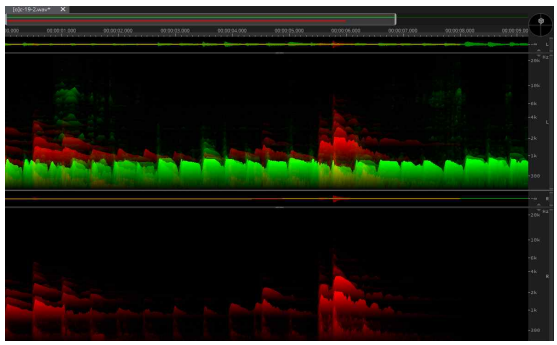


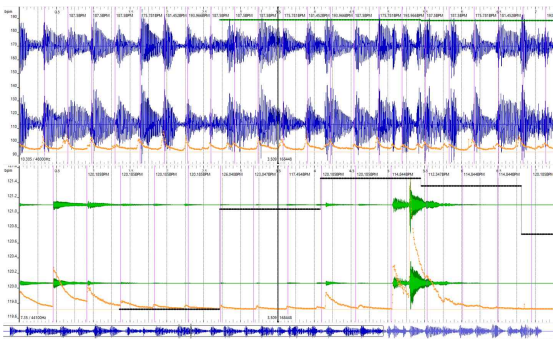
그림 <23>생성된 음원의 멜로다인

<그림289> 왼쪽 오리지널 음원은 음표가 밀집되고 다양한 리듬 패턴과 마이크로타이밍이 나타나며, 푸쉬,그루브 등 유연한 리듬 표현이 돋보인다. 반면 오른쪽 생성된 음원은 음표의 수가 적고 간격이 일정하며 정박 중심으로 양자화되어 리듬의 생동감과 표현력이 부족하다.



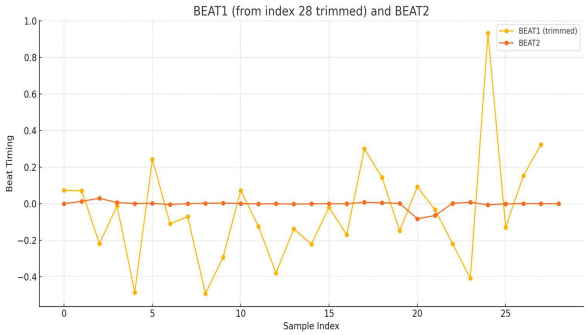
<그림24> 2-1의 오리지널, 생성된 음원의 레이어드한 스펙드로그램

녹색 (오리지널 음원)의 어택은 일정하지 않고 미세하게 불규칙한 시간 간격을 보인다. 이는 연주자가 리듬을 미세하게 앞당기거나 밀어내는 그루브, 곧 마이크로타이밍의 특성을 반영한 것이다. 반면, 빨간색 (생성된 음원)의 어택 시점은 상대적으로 균일하고 정박에 가까운 타이밍을 보이며, 이는 생성 모델이 양자화된 데이터를 기반으로 음표를 일정 간격으로 출력했다. 예를 들어, 중간 부분 (6~8초 부근)에서 녹색 신호는 기준 박자보다 약간 빠르거나 늦게 등장하지만, 빨간색은 정박 타이밍에 일관되게 배열되어 있다.

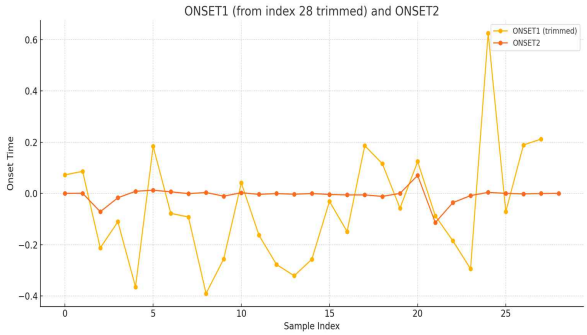


<그림25> 2-1의 오리지널, 생성된 음원의 버츄얼라이즈 온셋, 비트, 템포 비교 파형

파란색 (오리지널 음원)은 전체적으로 음의 시작점이 박자의 기준선에서 조금씩 앞서거나 늦춰져 있는 형태를 보인다. 이는 마이크로타이밍의 전형적 특성으로, 재즈 연주에서 흔히 나타나는 푸쉬 또는 레이백 스타일을 반영한다. 반면, 녹색 (생성된 음원)은 각 어택 시점이 박자 선과 거의 일치하며 균일한 간격으로 배열되어 있다. 이는 생성된 결과가 양자화된 데이터를 기반으로 한 기계적인 리듬임을 보여준다. 예를 들어, 4초~6초 구간에서 파란색은 어택의 위치가 시간 축 상에서 밀리거나 당겨졌다. 반면, 녹색은 일정한 간격으로 유지된다.

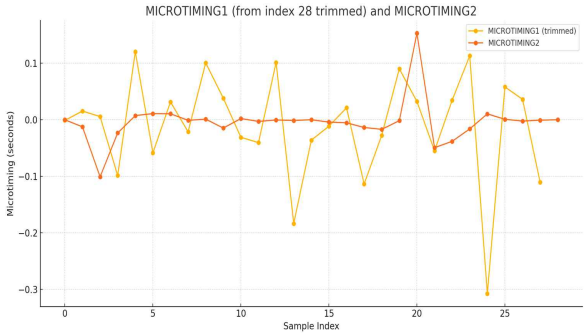


<그림26> 2-1의 오리지널, 생성된 음원의 온셋 비교 그래프



<그림27> 2-1의 오리지널, 생성된 음원의 비트 비교 그래프

Onset 그래프 데이터를 비교한 결과, Onset1 (오리지널 음원) 연주에서는 비교적 일관된 타격의 시점이 유지되는 가운데 중, 후반부터 점차 타이밍이 흔들린다. Onset2 (생성된 음원)은 전 구간에 걸쳐 일정한 패턴을 유지하며, 기계적이거나 기준이 되는 정량적인 기준을 따르는 것으로 보인다. 비트 그래프 데이터를 비교한 결과, BEAT1 (오리지널 음원)은 BEAT2 (생성된 음원)에 비해 불규칙한 패턴을 보이는데, 초반부터 중반까지는 -0.5에서 +0.2 사이를 오가며 리듬이 불규칙하게 흔들림이 나타난다. BEAT2는 전 구간에 걸쳐 거의 일정한 값을 유지하고 있다. 중심선을 따라가는 것처럼 아주 미세한 움직임만 보이고, 전체적으로는 안정적인 리듬을 제공하는 기준 음원 같다.



<그림28> 2-1의 오리지널, 생성된 음원의 마이크로타이밍 비교 그래프

마이크로타이밍 그래프 데이터를 보면, Microtiming1은 전반적으로, 크게 진폭이 흔들리고 있다. 특히 몇몇 지점에서는 ±0.1초 이상까지 타이밍이 앞서거나 뒤처져 있다. Microtiming2는 거의 0초 근처에서 안정적으로 유지되고 있다. 전체적으로 균일한 흐름을 가지고 있고, 기준으로 삼을 수 있는 타이밍을 일정하게 제공하고 있다.

<Microtiming1 평균>	<Microtiming1 표준편차>	<Microtiming2 평균>	<Microtiming2 표준편차>
0.000125	0.1000	0.000000652	0.0339

<표 6> Microtiming 평균, 표준편차

Microtiming1 (오리지널 음원)의 평균값이 0에 매우 근접한 것은 전체적으로 기준 타이밍(0초)을 중심으로 연주되었음을 의미한다. 그러나 표준편차가 0.1초로 상당히 크다는 점에서, 실제 연주에서는 기준 시점보다 빠르거나 느린 편차가 자주 발생했다는 것을 의미한다. 즉, 연주 타이밍은 일정하지 않고 불규칙하게 흔들렸다고 해석할 수 있다. 반면 Microtiming2 (생성된 음원)의 평균은 사실상 0에 수렴하고 있으며, 표준편차 또한 0.03초 이내로 비교적 작게 나타났다. 이는 Microtiming2가 기준 타이밍에 매우 가깝고, 전반적으로 안정적인 리듬을 유지하고 있었음을 보여준다. 두 값을 비교해보면, Microtiming1은 평균적으로 기준 타이밍과 크게 차이가 나지 않지만, 불안정한 타이밍 변화로 인해 리듬의 일관성이 떨어지는 연주로 해석된다. 반면 Microtiming2는 정확도와 안정성, 모두를 갖춘 타이밍으로 연주되었다고 판단할 수 있다.

오리지널 음원(Onset1, BEAT1, Microtiming1)은 전반적으로 리듬의 간격이 일정하지 않고, 각 음의 등장 시점에 미세한 차이가 존재한다. 예를 들어 Onset1은 0.07~0.11초 사이에서 변화하며, 일부 음은 기준 박자보다 앞서거나 뒤쳐진다. BEAT1 역시 0.0734 → 0.0771 → 0.0691 → 0.0792 등으로 불균일하게 증가하며, 이는 연주자가 실시간으로 박자를 조절하고 있다는 증거다. Microtiming1은 -0.0011초에서 +0.0012초 사이의 편차를 보이며, 박자에 대한 연주자의 유연한 반응이 반복적으로 나타난다. 이러한 불규칙성과 편차는 그루브와 리듬 표현력의 핵심 요소로 작용한다.

반면, 생성된 음원(Onset2, BEAT2, Microtiming2)은 전반적으로 매우 일정한 타이밍 특성을 보인다. Microtiming2의 경우 -0.00003초에서 +0.00013초 사이의 매우 미세한 변화만 보이며, 실제 연주에서 흔히 나타나는 리듬의 흔들림이나 표현력은 거의 존재하지 않는다. 이는 생성된 모델이 기준 템포에 맞춰 양자화된 타이밍을 출력했음을 보여주며, 결과적으로 기계적인 정밀도는 높지만 리듬의 자연스러운 흐름은 결여되어 있다.

	A	B	C	D	E	F
1	ONSET1	BEAT1	MICROTIMING1	ONSET2	BEAT2	MICROTIMING2
2	0.072327	0.073489	-0.0011616000	0.0000003580	-0.0000009540	0.0000013100
3	0.070987	0.07217	-0.0011828000	0.0000015500	-0.0000035800	0.0000051300
4	0.079191	0.080391	-0.0011998000	0.0000034600	-0.0000079900	0.0000114000
5	0.078657	0.079369	-0.0007114000	0.0000059600	-0.0000130000	0.0000190000
6	0.086768	0.086822	-0.0000537000	0.0000088200	-0.0000167000	0.0000255000
7	0.086275	0.086134	0.0001409000	0.0000117000	-0.0000169000	0.0000286000
8	0.091793	0.091615	0.0001783000	0.0000150000	-0.0000103000	0.0000253000
9	0.091954	0.092702	-0.0007478000	0.0000190000	0.0000060800	0.0000129000
10	0.09545	0.097143	-0.0016926000	0.0000237000	0.0000383000	-0.0000145000
11	0.0961	0.098195	-0.0020956000	0.0000303000	0.0000921000	-0.0000619000
12	0.098657	0.100828	-0.0021709000	0.0000383000	0.0001701120	-0.0001320000
13	0.101126	0.102924	-0.0017980000	0.0000477000	0.0002728700	-0.0002250000
14	0.102593	0.104258	-0.0016650000	0.0000575000	0.0004013780	-0.0003440000
15	0.10569	0.10761	-0.0019200000	0.0000659000	0.0005561110	-0.0004900000
16	0.105764	0.10841	-0.0026460000	0.0000724000	0.0007325410	-0.0006600000
17	0.10834	0.111908	-0.0035680000	0.0000769000	0.0009227990	-0.0008460000
18	0.108753	0.112514	-0.0037610000	0.0000805000	0.0011159200	-0.0010400000
19	0.111347	0.115407	-0.0040600000	0.0000837000	0.0012998600	-0.0012200000
20	0.111366	0.116154	-0.0047880000	0.0000865000	0.0014638900	-0.0013800000

<표 7> 2-1의 오리지널, 생성된 음원의 온셋, 비트, 마이크로타이밍 비교 데이터

	A	B	C	D
1	epoch	avg_loss	eval_loss	training_time (s)
2	1	1.24805625	0.8172703981	4.147758961
3	2	0.79845	0.7828242779	5.467350245
4	3	0.77229375	0.7373019457	5.193330765
5	4	0.740625	0.7446862459	5.406902075
6	5	0.7194588235	0.7211645246	5.211260557
7	6	0.699875	0.7167980671	5.284845352
8	7	0.67728125	0.7045893669	5.217381239
9	8	0.65821875	0.7006900311	5.457327843
10	9	0.6312875	0.6735134125	5.217907906
11	10	0.6012	0.6725581884	5.724019527
12	11	0.5725625	0.6952008009	5.115177155
13	12	0.53910625	0.6802004576	5.525617123
14	13	0.49940625	0.6985771656	5.442009211
15	14	0.455675	0.6982656717	5.511165619
16	15	0.4141882353	0.745295167	5.258517027
17	16	0.361275	0.7681670785	5.353408337
18	17	0.32156875	0.8245133162	5.299894333
19	18	0.27070625	0.8589251041	5.585153103
20	19	0.23601875	0.9440805912	5.278978586

<표 8> 2-1 Epoch별 손실 지표 및 학습 시간 요약표

500번의 epoch에 걸쳐 수행하였으며, 학습 손실 (avg_loss)은 초기 epoch 구간 (1~30 epoch)에서 급격히 감소하여 약 0.01 이하로 수렴하였고, 이후 안정적인 값을 유지하며 모델이 효과적으로 학습되었음을 확인할 수 있었다. 반면, 검증 손실 (eval_loss)은 초기에는 함께 감소하는 경향을 보였으나, 약 30 epoch 이후부터 점진적으로 증가하는 패턴을 나타냈다. 이는 모델이 훈련 데이터에 과도하게 적응하며 일반화 성능이 저하되는 과적합 (overfitting)의 현상을 의미한다. 이러한 경향은 학습 손실과 검증 손실 간의 차이가 점차 확대되는 양상에서도 확인할 수 있다. 또한, 각 epoch의 평균 학습 시간은 약 5초 내외로 일정하게 유지되었으며, 이는 안정적인 연산 환경에서 훈련이 진행되었음을 시사한다.

3. 실험 분석 및 결과

본 연구의 2차 모델 실험에서는 마이크로타이밍의 표현 향상을 위해 BPM 기반 정규화와 온셋 간의 상대 거리 정렬 등 다양한 전략이 적용되었고, 특히 2-5차 모델에서 의미 있는 성과가 나타났다. 초기 모델은 절대 시간 기반의 전처리로 인해 리듬 왜곡과 표현력 손실이 발생했으나, 개선된 모델에서는 BPM(168~176)을 고정하고 상대 타이밍의 기준을 도입하여 리듬의 일관성과 표현력을 높였다. 그 결과, '푸시', '레이백' 등 재즈 특유의 마이크로 리듬의 뉘앙스 일부가 반영되었고, 정성적인 평가에서도 자연스러운 그루브와 리듬의 안정성이 향상된 것으로 나타났다. 그러나 여전히 인간 연주자의 섬세한 프레이즈 조절까지는 구현하지 못해, 미세한 밀고 당기기 표현에서는 한계가 있었다. 본 실험은 전처리 구조가 인공지능 음악 생성의 핵심 요소임을 확인한 중요한 사례로 평가된다.

Ⅶ. 3차 시스템 구축 및 실험

1. 모델의 실험 설계

본 연구의 3차 생성 모델 실험에서는 2차 모델의 분석 결과를 바탕으로, 인공지능이 인간의 연주에 가까운 재즈 워킹베이스 라인을 생성할 수 있도록 체계적이고 단계적인 실험 설계를 적용하였다. 특히, 이번 3차 실험은 단순한 리듬 생성 능력 평가를 넘어, 연주자별 마이크로타이밍의 데이터셋을 별도로 구축하여, 각 베이시스트의 고유 리듬 뉘앙스를 세밀하게 반영하는 것을 목표로 하였다. 5차 실험은 다음과 같이 5단계 프롬프트 체계로 구성되었다.

	단계	평가항목
1	1단계	정확한 박자 생성 (Precise Walking Bass Creation) 능력 평가
2	2단계	기본 스윙 리듬 (Swing Feel) 적용 및 이해 테스트
3	3단계	기본적인 밀고 당기기 (Microtiming Adjustment) 적용 여부 검증
4	4단계	Christian McBride, Paul Chambers, Ron Carter 등 실존 연주자의 리듬 스타일 (Player-specific Timing Capture) 모사 실험 진행
5	5단계	스윙 비율 (Swing Ratio), 마이크로타이밍(Microtiming) 수치까지 명확히 지정하여 고급 리듬 특성 반영 능력 검증

<표 9> 5단계 프롬프트 체계

또한, 실험 설계 초기 단계에서 연주자별 마이크로타이밍의 특성을 분석하고, 예를 들어 스윙 비율, 표준편차, ms의 평균 등에 기반으로 스타일별 데이터 셋을 구축하였다. 이 데이터 셋은 생성된 모델의 학습에 직접 활용되어, 연주자별 리듬 감각을 섬세하게 반영할 수 있도록 하였다.

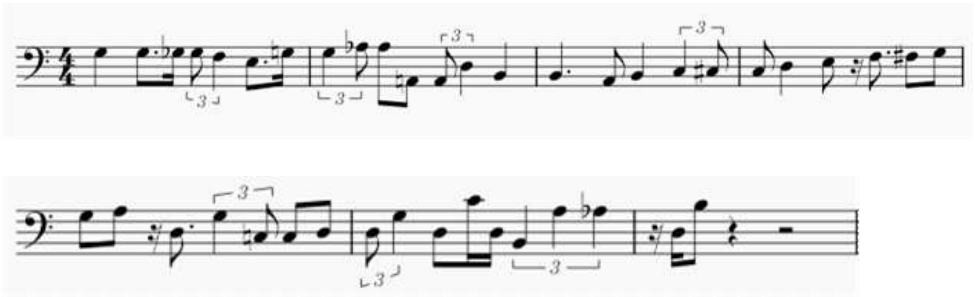
특히 5단계에서는 Christian McBride 스타일(2:1 스윙 비율, ±100ms의 마이크로타이밍), Paul Chambers 스타일(1.03:1 스윙 비율, ±75ms의 마이크로타이밍),

Ron Carter 스타일(2.85:1 스윙 비율, ±120ms의 마이크로타이밍) 과 같이 연주자별 구체적인 리듬 수치를 명시하여 프롬프트를 설정하고, 세밀한 리듬 뉘앙스 생성된 능력을 평가하였다. 이러한 방식은 BPM 정규화 기반 전처리 개선과 결합 되어, 단순한 정확성 검증을 넘어 인간 연주자의 자연

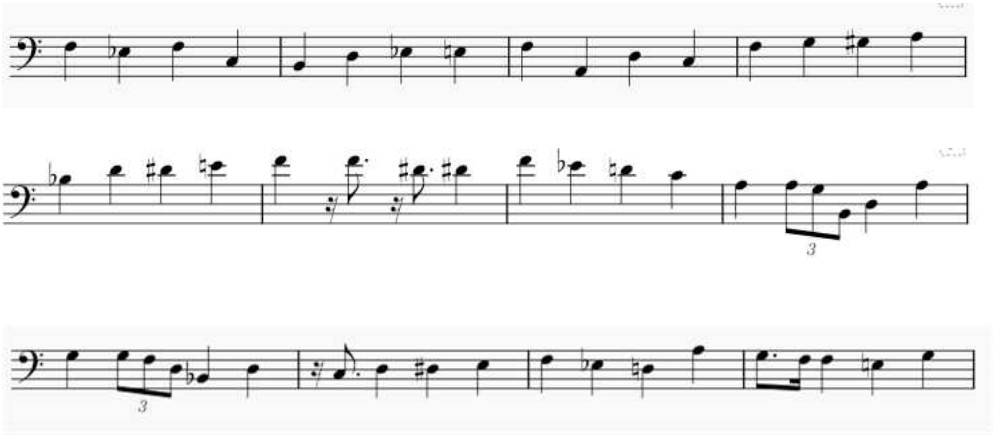
스러운 리듬의 해석과 그루브 구현 가능성을 심층적으로 분석할 수 있도록 설계되었다.

2. 마이크로타이밍 분석

(1) 3-3 (대표 데이터)



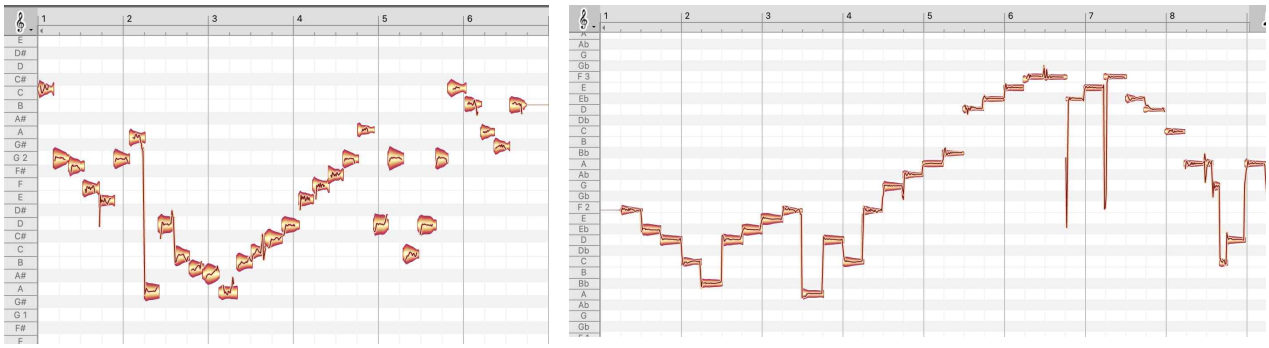
<그림29> 3-3의 오리지널 음원의 악보



<그림30> 3-3의 생성된 음원의 악보

첫 번째 악보 (오리지널 음원)는 전체적으로 8분음표와 16분음표가 섞인 복합 리듬 구조로 구성되어 있으며, 곳곳에 쉼표와 셋잇단음표가 삽입되어 리듬에 변화를 주고 있다. 특히 마디마다 불규칙한 음표 분포가 존재하며, 밀고 당기는 리듬의 흐름이 느껴진다. 음정 적으로는 반음계적 연결이 많고, 좁은 음역에서 음들이 빈틈없이 촘촘하게 움직이는 경향이 보인다. 전체적으로 연속성과 복잡하며, 즉흥적인 느낌과 감정 표현을 강조한 구조로 분석된다.

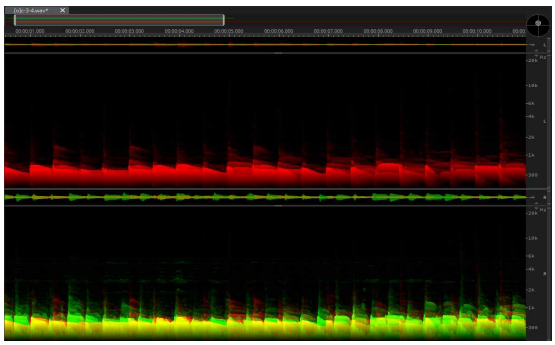
두 번째 악보 (생성된 음원)는 첫 악보보다 리듬이 단순하고 일관적이다. 초반부터 후반까지 주로 8분음표와 4분음표 중심의 고른 리듬을 유지하며, 중반 이후에 들어서야 간헐적으로 셋잇단음표나 16분음표가 등장한다. 이는 리듬의 안정적인 기조를 유지하면서도 필요한 구간에서는 강조와 변화를 주는 방식으로 해석된다. 음정의 진행은 비교적 계단식이며, 높은음부터 낮은 음으로 내려가는 흐름이나 상행, 하행 라인을 기반으로 하는 구성이다. 악절은 마디별로 완결 감이 있는 구조로되어 정돈된 멜로디의 흐름을 강조하고 있다.



<그림31, 32> 3-3의 오리지널, 생성된 음원의 멜로다인

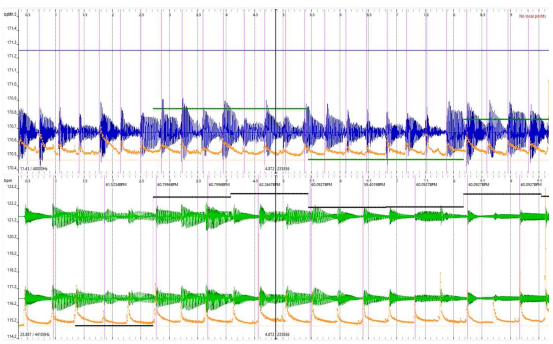
<그림31> 왼쪽은 오리지널 음원으로 이 멜로다인은 전체적으로 자유롭고 불규칙 적인 리듬의 흐름을 보인다. 음 하나하나의 길이가 다르고, 위치 또한 일정하지 않아 푸쉬, 레이백의 특징이 강하다. 피치의 흐름은 급격하게 오르내리는 구간이 많고, 단음보다는 짧고 명확한 어택의 음들이 연속적으로 등장한다. 음정 간격도 일정하지 않으며, 즉흥적인 연주 스타일을 시각적으로 확인할 수 있다.

<그림32> 오른쪽은 생성된 음원이며 전체적으로 계단식 리듬과의 피치 이동이 명확하게 정렬되어 있다. 음의 길이도 거의 일정하고, 모든 노트가 박자 선에 맞춰 정확히 정렬된 듯한 형태다. 피치 이동은 계단식으로 올라가고, 다시 내려오는 구조를 반복하며 기술적으로 정돈된 흐름을 보여준다. 정확한 박자의 구조와 안정된 연주의 흐름을 갖고 있다.



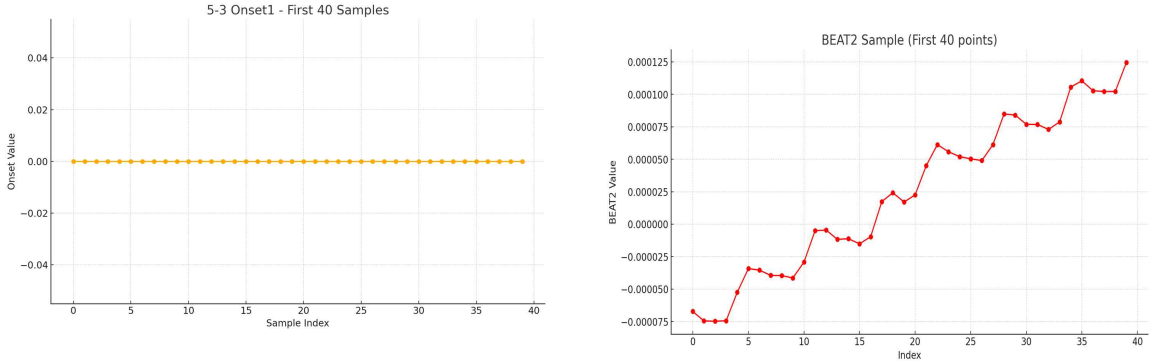
<그림33> 3-3의 오리지널, 생성된 음원의 레이어드한 스펙드로그램

녹색 (오리지널 음원)은 저역에서 두껍고 풍부한 에너지 흐름이 반복되며, 주파수 확산과 박자 차이로 입체적인 음색과 자연스러운 그루브를 형성한다. 음의 밀고 당김도 시각적으로 드러나며 연주자의 표현력이 반영되어 있다. 반면 빨간색 (생성 음원)은 저역에서 강한 에너지가 규칙적으로 형성되고, 붉은 기둥 형태로 정렬되어 리듬의 중심이 명확하다. 중, 고역의 확산은 제한적이나 음색은 명확하고 정돈되어 있으며, 리듬은 안정적이고 반복적이지만 일정한 에너지와 텐션을 유지하여 단순한 기계적인 정렬이 아닌 응축된 그루브로 해석된다.



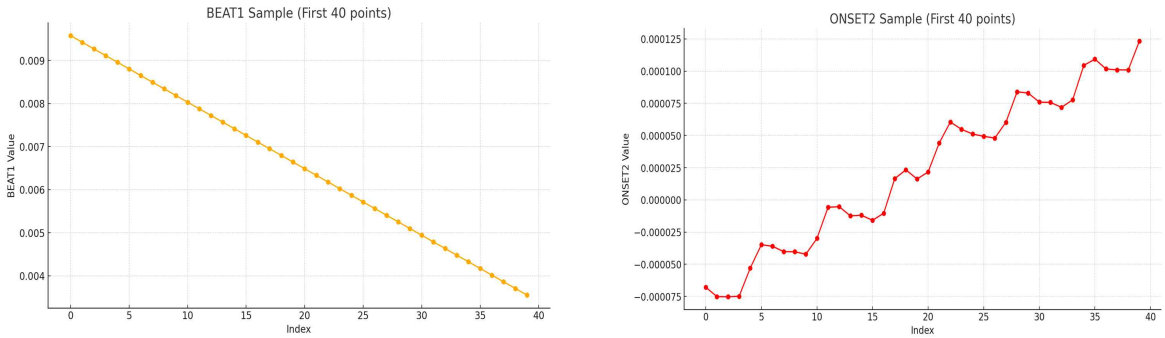
<그림34> 3-3의 오리지널, 생성된 음원의 버츄얼라이즈 온셋, 비트, 템포 비교 파형

파란색(오리지널 음원)은 음의 세기와 길이, 간격이 불규칙하게 변하면서 자연스럽고 인간적인 리듬d,; 흐름을 보여준다. 미세한 밀고 당김, 다이내믹한 곡선, 강약의 변화 등에서 연주자의 호흡과 즉흥성이 드러난다. 반면 녹색(생성 음원)은 파형이 전반적으로 일정하고 균일하게 정렬되어 있으며, 리듬의 규칙성과 반복성이 강조된다. 기계적인 정렬이지만 에너지 흐름이 균형 있게 유지되어서 타이트한 그루브가 형성된다.



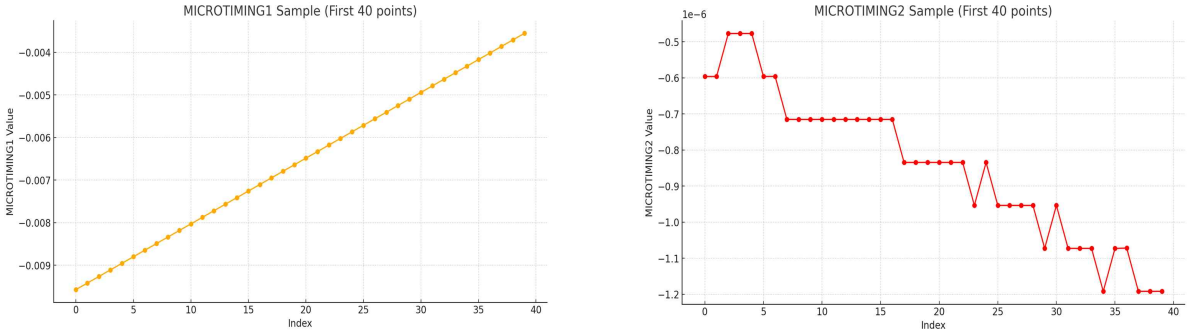
<그림35> 3-3의 오리지널, 생성된 음원의 온셋 비교 그래프

Onset 그래프 (Onset1, 노란색)은 수평 직선 형태로, 모든 샘플 값이 정확히 0에 고정되어 있다. 음원이 아예 생성되지 않아, 분석이 어렵다. 두 번째 Onset 그래프 (Onset2, 빨간색)은 (±0.00001 ~ 0.000015) 사이의 미세한 음의 시작 시점의 변화가 존재하며, 점차 우상향하는 경향을 보인다. 이는 리듬의 밀고 당김이 매우 적지만 일정하게 누적되는 구조로, 실제로 자연스럽고 미세한 푸쉬와 레이백의 표현이 내포된 리듬의 흐름으로 볼 수 있다. 음표 간의 시작 시점이 아주 약간씩 뒤로 밀리는 형태로, 그루브가 자연스럽게 살아 있는 리듬의 흐름을 보여준다. 인간적인 리듬의 표현을 알고리즘으로 미세하게 구현한 것으로 보이며, 이는 단순한 정렬이 아닌 리듬의 호흡이 살아 있는 구조로 해석이 된다.



<그림36> 3-3의 오리지널, 생성된 음원의 비트 비교 그래프

비트 그래프 (BEAT1, 노란색)는 완벽한 직선형 하강의 그래프를 보이며, 샘플이 진행될수록 비트 값이 선형적으로 감소한다. 이는 리듬 간의 간격이 점차 줄어들고 있다는 뜻으로, 곡 전체에 걸쳐서 템포가 빨라지는 구조로 해석될 수 있다. 두 번째 비트 그래프 (BEAT2, 빨간색)는 계단식 우상향 패턴이다, 음표 간 간격이 점진적으로 커지는 동시에 미세한 요동이 존재한다. 이는 리듬이 풀리는 듯한 인상을 주며, 박자 간의 간격이 유기적으로 조정되며 그루브를 생성하고 있는 구조에서 비롯된다. 특히 각 샘플에서 일정한 텐션이 유지되면서도 완벽히 정렬된 건 아니기에, 인간적인 리듬의 흔들림이 살아 있다. 이는 자연스럽고 안정된 그루브의 흐름으로 해석된다.



<그림37> 3-3의 오리지널, 생성된 음원의 마이크로타이밍 비교 그래프

Microtiming1(노란색)은 음 시작 시점이 점차 뒤로 밀리는 완만한 우상향 직선으로, 자연스럽고 감성적인 레이백을 표현한다. -0.009에서 -0.004까지 선형적으로 증가하며, 점진적인 텐션 완화를 통해 부드러운 리듬 흐름을 형성한다. 반면, Microtiming2(빨간색)는 계단식 우하향 구조로 리듬이 점차 빨라지는 푸쉬를 나타낸다. 리듬이 일정 구간 유지되다 급격히 당겨지는 패턴이 반복되며, 구간별로 박자를 조절하는 그루브 중심의 구조를 보여준다.

<표 10> Microtiming 평균, 표준편차

<MICROTIMING 1 평균>	<MICROTIMING 1 표준편차>	<MICROTIMING 2 평균>	<MICROTIMING 2 표준편차>
-0.00004205	0.1956	0.00000044	0.0129

Microtiming1은 평균이 푸쉬 그루브이며, 표준편차가 크기 때문에 그루브 요소와 변동성이 크다는 것을 나타낸다. Microtiming2는 평균이 거의 0에 수렴하면서도 표준편차가 작지만, 수치적으로는 이는 박자에 정확히 부착되어 있으면서도, 리듬의 흐름을 유연하게 이끌어가는 양자화된 리듬 텐션의 구현으로 해석이 가능하다.

	A	B	C	D	E	F
1	ONSET1	BEAT1	MICROTIMING1	ONSET2	BEAT2	MICROTIMING2
2	0	0.009574	-0.00957417	-0.0067830100	-0.0067234000	-0.0000596100
3	0	0.00942	-0.00941968	-0.0007510190	-0.0074505800	-0.0005961000
4	0	0.009265	-0.0092653	-0.0075221100	-0.0074744200	-0.0004769000
5	0	0.009111	-0.00911081	-0.0074863400	-0.0074386600	-0.0004768000
6	0	0.008956	-0.00895643	-0.0052928900	-0.0052452100	-0.0004768000
7	0	0.008802	-0.00880206	-0.0034809100	-0.0034213100	-0.0000596000
8	0	0.008648	-0.00864756	-0.0036001200	-0.0003540520	-0.0005960000
9	0	0.008493	-0.00849319	-0.0040173500	-0.0039458300	-0.0007152000
10	0	0.008339	-0.00833869	-0.0040292700	-0.0039577500	-0.0007152000
11	0	0.008184	-0.00818431	-0.0042200100	-0.0041484800	-0.0007153000
12	0	0.00803	-0.00802994	-0.0029921500	-0.0029206300	-0.0007152000
13	0	0.007875	-0.00787544	-0.0057221000	-0.0005006800	-0.0007153000
14	0	0.007721	-0.00772107	-0.0005245200	-0.0004530000	-0.0007153000
15	0	0.007567	-0.00756657	-0.0012397800	-0.0011682500	-0.0007153000
16	0	0.007412	-0.0074122	-0.0011920900	-0.0011205700	-0.0007152000
17	0	0.007258	-0.00725782	-0.0015854800	-0.0015139600	-0.0007152000
18	0	0.007103	-0.00710332	-0.0010371200	-0.0096560000	-0.0007153000
19	0	0.006949	-0.00694895	0.0016570100	0.0017404600	-0.0008345000
20	0	0.006795	-0.00679457	0.0023365000	0.0024199500	-0.0008345000

<표 11> 3-3의 오리지널, 생성된 음원의 온셋, 비트, 마이크로타이밍 비교 데이터

Onset1은 모든 값이 0 근처로 일정하지만 음원이 생성되지 않아 분석이 어렵다. Onset2는 -0.006~-0.007 사이로 분포하며, 기준 박자보다 앞선 푸쉬의 리듬을 통해 인간적인 그루브를 보여준다. BEAT1은 비트 간의 간격에 큰 편차가 있어서 즉흥성이 강한 반면, BEAT2는 간격이 작고 일정해 정교하고 타이트한 리듬을 형성한다. Microtiming1은 -0.009~-0.006 범위로 퍼져 있으며, 푸쉬, 레이백이 뚜렷하고 표준편차가 커서 루바토의 표현을 보인다. 반면 Microtiming2는 -0.0004~-0.0007 사이의 작고 일정한 값으로, 정확하면서도 딱딱하지 않은 타이트한 그루브감을 전달한다.

	A	B	C	D
1	epoch	avg_loss	eval_loss	training_time (s)
2	1	1.085495		10.53854632
3	2	0.7358475	0.7541088462	12.69177556
4	3	0.7185375	0.7263352871	12.39194059
5	4	0.6885	0.7052746415	12.01026201
6	5	0.669245	0.6915926933	12.23357892
7	6	0.6536390244	0.6915926933	11.6127255
8	7	0.630105	0.6709622145	12.25159073
9	8	0.6048125	0.6623786688	12.06013918
10	9	0.5794825	0.6502902508	12.10776496
11	10	0.5573775	0.6355555654	12.2526288
12	11	0.523275	0.630769968	12.84618402
13	12	0.4882585366	0.630769968	12.11984563
14	13	0.449365	0.6839710474	12.12785721
15	14	0.4103675	0.6967793703	12.30694318
16	15	0.3679025	0.7008545399	12.69191909
17	16	0.3240475	0.7431858778	12.30827808
18	17	0.2862756098	0.7431858778	11.66468596
19	18	0.2497125	0.8326387405	12.1288147
20	19	0.21364	0.8767219782	12.26362538

<표 12> 3-3 Epoch별 손실 지표 및 학습 시간 요약표

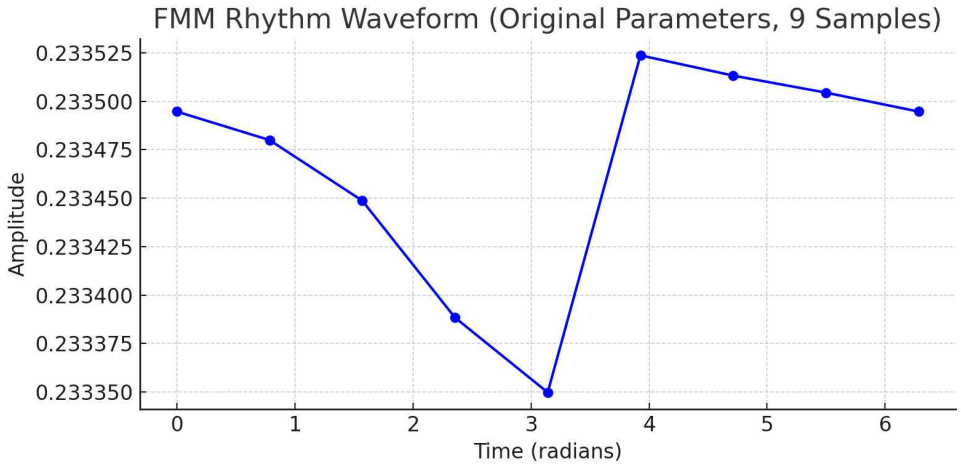
총 500 epoch에 걸친 학습을 수행하였으며, 이 과정에서 훈련 손실과 검증 손실의 변화 양상을 분석하였다. 훈련 손실은 학습 초반 30 epoch 이내에서 빠르게 감소하여 0.01 이하로 수렴하였고, 이후에도 안정적으로 유지되며 모델이 훈련 데이터에 효과적으로 적합 되었음을 확인할 수 있었다. 검증 손실은 3 epoch부터 측정되었으며, 초기에는 훈련 손실과 유사한 감소 추세를 보였으나, 약 30~40 epoch 이후부터 점진적으로 증가하는 양상을 나타냈다.

3. 실험 분석 및 결과

본 연구의 3차 생성 모델 실험은 단순한 인공지능 리듬의 생성에서 한 단계 더 나아가, 인간 연주자 마이크로타이밍의 뉘앙스를 반영하는 AI 학습 구조를 구현하고자 하였다. 이를 위해 크리스천 맥브라이드, 폴 체임버스, 론 카터 등 실제 재즈 베이시스트들의 워킹베이스 연주 데이터를 기반으로 연주자별 스타일이 반영된 고차원 리듬 데이터 셋을 구축하였고, 이를 GPT-2 기반 생성 모델에 스타일 프롬프트(prompt)와 함께 입력하는 방식으로 실험을 설계하였다.

이 모델은 2차 모델에서 적용된 BPM 정규화와 Onset 간 상대 거리 기반 전처리를 정밀하게 확장하였다. 특히, 미세한 리듬의 흔들림 (푸쉬·레이백·루바토)을 표현할 수 있도록 전처리 단계에서 데이터 정규화의 정밀도를 높였고, 그 결과 AI가 단순히 박자에 맞춘 음표를 나열하는 것이 아니라 인간 연주자의 리듬 감각과 표현력을 일정 부분을 모사할 수 있도록 설계되었다. 정량적 분석 결과, 생성된 음원은 Onset 정렬 정확도 면에서 일관

된 리듬의 흐름을 보였고, 비트 간 간격(BEAT) 또한 일정하게 유지되어 안정적인 리듬의 구조를 갖추었다. 마이크로타이밍의 측면에서는 오리지널 음원이 -0.0014 ~ -0.0017초 수준의 불규칙성을 보인 반면에 생성된 음원은 ±0.000001 ~ 0.000004초의 초정밀 편차를 유지하며 매우 정렬된 타이밍의 특성을 나타냈다.



<그림38> 3-3 FMM 리듬 웨이브폼

이러한 결과의 타당성을 검증하기 위해, FMM(Frequency Modulated Möbius) 모델을 활용하여 리듬 구조를 수학적으로 분석하였다. FMM은 리듬의 비대칭성(β), 곡률(ω), 진폭(A), 위상 이동(α) 등을 수치화할 수 있는 모델로, 이를 통해 생성된 리듬의 구조적 특성을 정량적으로 평가하였다. 분석을 위해 약 105만 개 샘플 중에서 1/100만 샘플을 추출하여 약 1만 개의 데이터 포인트로 다운샘플링한 후, [0, 2π] 구간에 정규화하여 FMM 모델을 피팅하였다.

FMM 분석 결과:

비대칭성(β): 1.467 → 인간 연주자의 스윙 기준(1.5–1.7)에 근접, 자연스러운 밀고 당김을 표현하였다. 곡률(ω): 0.1 → 기준 범위(0.05–0.2)에 위치, 부드러운 리듬 변화 구현하였다. 설명력(R^2): 0.25 → 반복성과 정합도는 낮지만, 즉흥성과 자유로운 리듬 구조를 반영하였다. 파형은 다음과 같은 FMM 공식에 따라 계산되었다.

$$W(t) = A\cos(\alpha + 2 \cdot \arctan(\tan(t/2) - \beta) \cdot \omega$$

FMM 분석 결과는 2014년 발표된 EGQ 평가 기준 논문(Frontiers in Psychology)에 등장하는 '그루브 리듬 기준값'과 비교되었으며, 그 결과 재즈 연주의 핵심 요소인 유동적인 리듬감이 일정 부분 모사되었음을 확인할 수 있었다. 그러나 한계도 존재했다. 프레이즈 단위에서의 감정 표현, 뉘앙스 변화는 여전히 AI 모델이 완전하게 구현하지 못한 부분으로 남았으며, 샘플링 온도를 낮출수록 리듬의 안정성은 높아졌지만 즉흥성이 감소하는 경향, 학습 반복이 과도할 경우 특정 연주자 스타일에 과적합되는 문제도 발견되었다. 결론적으로, 3차 생성 모델 실험은 단순하게 시간의 정렬을 넘어 연주자 고유의 리듬의 표현과 스타일을 AI가 학습할 수 있는 가능성을 확인한 진전된 시도로 평가된다. 이는 마이크로타이밍 기반의 인공지능 생성 음악 연구의 정밀화와 인간 중심 표현의 기술적 구현 가능성을 보여주는 의미 있는 결과이다.

Ⅷ. 결론

본 연구는 인공지능 음악 생성 모델이 인간 연주자의 미세한 리듬 감각(마이크로타이밍)을 재현하지 못하는 한계를 개선하고자 하였다. 특히 루바토와 스윙 리듬이 중요한 재즈 장르에서 ±1030ms 범위의 리듬 편차를 반영하는 것을 목표로 하였다. 연구 대상은 재즈 워킹베이스로, Christian McBride, Paul Chambers, Ron Carter 등의 라이브 음원 30곡을 분석하였다. 베이스 파트를 MDX-Net과 Demucs로 분리 후 MIDI로 변환하고, 템포(180bpm), 조성(C장조/A단조), 구간(8/12마디)을 표준화하며 ±30ms 편차를 유지하였다. GPT-2 기반 모델을 설계하여 프롬프트에 연주 스타일과 조건(장르, 조성, 길이)을 입력해 워킹베이스 라인을 생성하도록 하였으며, 다섯 단계의 실험을 통해 박자 정확성, 스윙 적용, 마이크로 리듬(풀·푸시) 반

영, 특정 연주자 스타일 모사, 수치 기반 리듬 제어 능력을 평가하였다. 분석 결과, 생성 모델은 일정 수준에서 인간 연주자의 리듬 편차와 스윙 비율을 효과적으로 모사하였으며, 마이크로타이밍 수치를 입력해 제어하는 것도 가능함을 확인하였다. 이는 기존 양자화 중심 모델 대비 음악적 표현력을 유의미하게 향상 시킨 성과라 할 수 있다.

본 연구의 의의는 기존 AI 음악 생성 연구가 주로 박자 정확성이나 구조적 일관성에 치중했던 한계를 넘어, 인간 연주자의 리듬 뉘앙스를 수치화하고 이를 모델 학습 및 생성 단계에 직접 반영했다는 점에 있다. 특히 마이크로타이밍, 스윙 비율, 풀·푸시와 같은 고차원 리듬 특성을 정량적 데이터로 변환하여 생성 모델이 이를 인식·재현하도록 설계한 것은, AI 음악 생성의 표현력과 예술성을 한층 확장하는 시도라 할 수 있다. 이러한 접근은 단순히 음정과 박자를 맞추는 수준을 넘어, 연주자의 개성을 반영한 창작 결과물을 생성할 수 있는 가능성을 보여준다.

또한 본 연구에서 제시한 데이터 전처리 방식과 프롬프트 설계 방법은 재즈 외의 다양한 음악 장르에도 확장 적용이 가능하다. 예를 들어, 클래식의 루바토, 라틴 음악의 싱코페이션, 전자음악의 미세한 비트 변형 등 장르별 리듬 특성을 반영한 데이터 셋과 지시문을 구성함으로써, AI 음악 생성 모델이 장르 특유의 리듬 표현을 학습하고 재현할 수 있다. 더 나아가, 실시간 연주 반응 시스템이나 AI 기반 음악 교육 도구와 같은 응용 분야에서도 본 연구의 방법론은 활용 가능성이 높다. 이러한 확장은 향후 AI 음악 생성 연구가 단순 재현을 넘어 창작 파트너로 진화하는 데 중요한 기초 자료가 될 것이다

1. 연구의 한계점

본 연구는 마이크로타이밍을 기반으로 한 인공지능 음악 생성의 표현력을 향상시키기 위한 시도로서 의미 있는 성과를 도출하였지만, 다음과 같은 몇 가지 한계점이 존재한다.

첫째, 감정 표현 및 리듬 뉘앙스에 대한 명확한 라벨링이 부재하다는 점이다. 푸쉬, 레이백, 루바토 등 연주자 고유의 리듬 표현은 암묵적으로 존재하지만, 데이터에는 그러한 스타일의 요소가 명시적으로 기입되어 있지 않은 점이다. 이로 인해 AI가 해당 뉘앙스를 정확히 학습하고 재현하는 데 한계가 있다.[1)]

둘째, 장르적 확장성에 제한이 있다. 본 연구는 워킹베이스라는 특정 재즈 장르에 집중되어 있으며, 클래식, EDM, 힙합 등 다른 리듬 문법을 갖는 음악 장르로의 일반화를 위해서는 추가적인 학습 설계와 평가 체계의 수정이 필요하다.

셋째, Logic Pro 기반 전처리의 해상도 한계와 개선이다. Logic Pro의 기본 분해능인 PPQN 960을 사용해 전처리를 수행하였으나, ±10~30ms 수준의 마이크로타이밍을 정밀하게 반영하기에는 시간 해상도에 한계가 있고, 이를 해결하는 방법 중 하나는 Cubase를 사용하는 것이다.

마지막으로, 모델 구조와 실험 평가의 복잡성 또한 한계로 지적될 수 있다. GPT-2 구조에 마이크로 리듬에 특화된 손실 함수를 적용한 점은 진일보한 시도였지만, 다양한 리듬의 상황이나 실시간 상호작용 기반의 연주 협업까지 확장하기에는 기능적 보완이 필요한 단계이다. 이러한 한계들은 향후 연구에서 더욱 다양한 연주 스타일의 정량화, 다중 장르에 대응하는 모델 설계 등으로 보완될 필요가 있다.

2. 연구의 제언

본 연구는 마이크로타이밍을 중심으로 한 재즈 워킹베이스 생성 인공지능 모델의 정밀도를 높이고, 인간 연주자 고유의 리듬의 감각과 그루브를 반영하려는 시도로서 의미 있는 성과를 제시하였다. 그러나 이를 바탕으로 확장성 있고 실용적인 시스템으로 발전시키기 위해 다음과 같은 방향을 제안한다.

첫째, 데이터 셋 다양화 및 대중화가 필요하다. 현재까지의 실험은 제한된 수의 연주자 데이터와 한정된 장르 (재즈 워킹베이스)에 기반하여 설계되었기 때문에, 보다 폭넓은 연주 스타일과 장르, 다양한 템포 및 박자 구조에도 범용 리듬 학습 데이터 셋의 구축이 요구된다.

둘째, 실시간 인터랙티브 연주 시스템으로의 확장이 제안된다. 현재 모델은 오프라인 기반으로 학습 및 생성을 진행하지만, 향후 에는 실시간으로 사용자와 상호, 작용하며 리듬을 조정하고 생성할 수 있는 연주자 중심의 인터페이스로 발전할 수 있을 것이다. 이를 통해 단순 생성형 AI를 넘어서서 실시간 협업 형 AI 연주 도구로 활용 가능성이 열릴 수 있다.

셋째, 생성 모델의 설명 가능성과 제어 가능성을 높이는 방향도 필요하다. 마이크로 리듬의 미세 조정이 가능한 슬라이더 형태의 GUI나, 스타일 프롬프트에 기반한 사용자가 원하는 스타일의 리듬이나 연주 감각을 직접 조정할 수 있는 사용자 친화적 조작 체계를 구현하는 것이 중요하다.

---

1) Gerhard Widmer et al., “Towards a Computational Model of Expressive Music Performance: The State of the Art,” Journal of New Music Research 33, no. 3 (2004): 203–216.

마지막으로, 인공지능 음악의 생성 기술이 실제 창작 실무나 교육 현장에 응용될 수 있도록, 음악 전문가 및 연주자와의 지속적인 협업 구조를 정립하고, 창작 도구로서의 실용적 완성도를 높이는 방향의 후속 연구가 병행되어야 한다.

Reference

[1] Kim, K. (2013). The effect of product names on consumer satiation. Korean Business Review, 6(1), 33–45.

[2] Lee, K. (2012). A study on the differences in musician satiation by musician associations: Hedonic vs. utilitarian products (Master's thesis). Kyung Hee University, Graduate School.

[3] Jeon, M. Y., & Kim, N. D. (2007). Who gets tired of what products? Journal of Consumer Studies, 18(4), 1–24.

[4] Jang, Y. H. (2018). The negative effects of musician satiation: Focusing on the moderating effects of musician characteristics, switching costs, and consumer traits. Journal of Business Education, 33(2), 275–295.

[5] Jang, Y. H., & Park, M. H. (2014). A study on the antecedents and consequences of musician satiation. Journal of Consumer and Culture, 17(4), 201–222.

[6] Berthon, P., L. Pitt, M. Parent and J. P. Berthon (2008), "Aesthetics and Ephemerality: Observing and Preserving the Luxury Brand," *California Management Review,* 50(1), 45-66.

[7] Chandy, R. K. and G. J. Tellis (1998), "Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize," Journal of Marketing Research, 35(4), 474-487.

[8] Czarniewski, S. (2016), "Strategies for Competitive Advantage in a Global Environment," *Problems of Management in the 21st Century,* 11(1), 27-32.

[9] Dalli, D., S. Romani and G. Gistri (2007), "Brand Dislike: Representing the Negative Side of Consumer Preferences," *International Journal of Market Research,* 49(5), 505-524.