

특집논문 (Special Paper)

방송공학회논문지 제31권 제5호, 2026년 9월 (JBE Vol.31, No.5, September 2026)

<https://doi.org/10.5909/JBE.2026.31.5.873>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

BPM 및 Target Timing 기반 목표 반응 보상을 이용한 강화학습 기반 주기적 동작 변형 방법

박 성 실^{a)†}

Reinforcement Learning-Based Method for Periodic Human Motion Transformation Using BPM-Based Cycle Alignment and Target-Timing Response Rewards

Soungsill Park^{a)†}

요 약

본 연구는 BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 이용하여 주기적 인간 동작을 변형하는 강화학습 기반 제어 방법을 제안한다. 제안 방법은 실제 동작 주기와 BPM으로부터 계산된 Beat Interval 간의 오차를 줄여 시간적 정렬성을 유도하고 Target Timing에서 손 관절이 목표 위치 방향으로 반응하도록 학습한다. 이를 위해 물리 기반 시뮬레이션 환경에서 조건 정보를 포함한 정책을 학습하고 AMP 기반 모션 모방 보상과 주기 정렬 보상, 목표 반응 보상을 결합하였다. 실험 결과, 팔 동작의 반복 속도는 목표 템포의 정수비에 대해 0.6% 이내의 오차로 동조되었으며, Target Timing 기반 목표 반응 보상 적용 시 Beat Timing 부근의 z축 이동 범위가 증가하고 반응 시점의 변동이 감소하였다. 이를 통해 BPM에 따른 반복 주기 조절과 Target Timing에 동기화된 관절 운동 반응을 확인하였다.

Abstract

This study proposes a reinforcement learning-based method for periodic human motion transformation using BPM-based cycle alignment and target-timing response rewards. The method adjusts motion cycles to BPM conditions and induces directional hand-joint responses near target timings in a physics-based simulation. Experiments show that the arm-swing rate was entrained to integer ratios of the target tempo with an error within 0.6%. The target response reward also increased z-axis travel near beat timings and reduced variation in response peak timing. These results demonstrate that the proposed method can control repetition cycles and target-timing-synchronized joint responses in a single periodic motion.

Keyword : Reinforcement Learning, Periodic Human Motion, Motion Transformation, Cycle Alignment Reward, Target-Timing Response Reward

Copyright © 2026 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

걷기, 뛰기, 팔 흔들기와 같은 인간의 기본 동작은 일정한 리듬과 반복성을 갖는 주기적 특성을 가진다. 이러한 주기적인 기본 동작은 가상현실, 증강현실, 게임, 메타버스, 인터랙티브 아바타, 로봇 제어 분야에서 자연스러운 움직임을 구현하기 위한 핵심 요소로 활용된다. 특히 외부 조건에 따라 동작의 주기나 방향성이 변화해야 하는 환경에서는 단순한 모션 데이터의 재생이 아니라 조건에 반응하는 동작 변형 기술이 필요하다.

기존의 모션 생성 및 변형 방법은 주로 사전에 수집된 모션 데이터를 재생하거나 사용자가 직접 키프레임을 수정하는 방식으로 이루어져 왔다. 이러한 방법은 정밀한 동작 편집이 가능하다는 장점이 있지만 외부 조건 변화에 따라 동작을 자동으로 조절하거나 실시간으로 반응하는 데에는 한계가 있다. 또한 데이터 기반 생성 모델은 학습 데이터와 유사한 동작을 생성하는 데 효과적이지만 특정 조건에 맞추어 동작의 주기성과 공간적 방향성을 명시적으로 제어하기 어렵다.

강화학습은 에이전트가 환경과 상호작용을 하면서 보상을 최대화하는 행동 정책을 학습하는 방법으로 물리 기반 동작 제어에 활용할 수 있다. 특히 원하는 동작 특성을 보상 함수로 정의하면 명시적인 정답 동작을 제공하지 않아도 정책이 조건에 적합한 움직임을 학습할 수 있다. 따라서 강화학습 기반 접근은 주기적 동작의 시간적 정렬성과 특정 시점에서의 방향성 있는 관절 운동 반응을 동시에 고려하는 동작 변형 문제에 활용될 수 있다.

본 연구에서는 BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 이용한 강화학습 기반 주기적 동작 변형 방법을 제안한다. BPM 기반 주기 정렬 보상은 실제 동작 주기와 BPM으로부터 계산된 Beat Interval

사이의 오차를 감소시켜 동작의 반복 주기가 외부 리듬 조건에 맞게 조절되도록 한다. Target Timing 기반 목표 반응 보상은 지정된 시점 부근에서 제어 대상 관절이 설정된 목표 방향으로 이동하도록 유도하여 주기 동작에 시점 의존적인 운동 반응을 부여한다.

제안 방법은 두 보상 함수를 AMP 기반 모션 모방 구조와 결합하여 물리 기반 시뮬레이션 환경에서 정책을 학습한다. 이를 통해 원본 주기 동작의 기본 형태와 자연스러움을 유지하면서 입력 BPM에 따라 반복 주기가 변화하고 Target Timing 부근에서 손 관절이 설정된 목표 방향에 대응하는 운동 반응을 나타내도록 동작을 변형한다.

II. 관련 연구

1. 강화학습 기반 동작 모방 방법

강화학습 기반 동작 모방은 모션 캡처 데이터와 같은 참조 동작을 기반으로 물리 시뮬레이션 환경에서 에이전트가 인간의 움직임을 학습하도록 하는 방법이다. 에이전트는 관절 위치와 회전, 관절 속도, 루트 상태, 접촉 정보 등을 관측하고 참조 동작과 생성 동작 사이의 유사도 또는 동작의 자연스러움을 보상으로 활용하여 제어 정책을 학습한다. 대표적인 연구로 Peng et al.의 DeepMimic^[1]은 참조 동작의 관절 자세, 관절 속도, 말단 관절 위치, 루트 상태를 추적하는 모방 보상을 이용하여 걷기, 달리기, 발차기와 같은 다양한 동작을 학습하였다. 모방 보상과 과제 보상을 결합하여 참조 동작의 특성을 유지하면서 목표 방향으로 이동하는 과제를 수행할 수 있다. Chentanez et al.^[2]은 모션 캡처 동작과 캐릭터의 상태 차이를 최소화하는 심층 강화학습 방법을 제안하여 시뮬레이션 과정에서 발생하는 자세 및 속도 오차를 정책이 지속적으로 보정한다.

이러한 연구들은 참조 궤적을 직접 추적하여 특정 동작을 정밀하게 재현하였다. Peng et al.^[3]의 AMP는 참조 동작의 각 프레임을 직접 추적하는 대신 모션 데이터와 생성 동작을 구분하는 판별기를 이용하여 동작 분포를 학습한다. 판별기의 출력은 스타일 보상으로 사용되며 생성 동작이 참조 데이터의 자연스러운 움직임 특성을 유지하도록 유도

a) 숭실대학교 전자정보공학부(Department of Electronic Engineering, Soongsil University)

* Corresponding Author : 박성실(Soungsill Park)

E-mail: pssil1321@gmail.com

Tel: +82-2-828-7194

ORCID: <https://orcid.org/0000-0002-6113-6919>

· Manuscript received July 19, 2026; Revised September 1, 2026;

Accepted September 1, 2026.

한다. 이후 ASE^[4]는 대규모 비정형 모션 데이터로부터 재 사용할 수 있는 동작의 잠재 표현을 학습하여 하나의 정책이 다양한 동작을 수행하고 새로운 과제에 활용될 수 있도록 확장하였다.

그러나 기존의 강화학습 기반 동작 모방 연구는 주로 참조 동작의 정확한 재현과 동역학적 불일치의 보정 또는 다양한 동작 스타일과 기술의 학습에 초점을 둔다. 따라서 BPM과 같은 시간적 외부 조건에 따라 동작의 반복 주기를 조절하거나 주기성을 유지하면서 특정 관절을 주어진 공간 목표 방향으로 유도하는 조건 반응형 동작 변형에는 추가적인 제어 입력과 보상 설계가 필요하다. 본 연구에서는 BPM, Target Position 및 Target Timing을 정책의 조건으로 제공하고, BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 적용하여 주기적 인간 동작의 시간적 반복성과 특정 시점의 방향성 있는 관절 운동 반응을 외부 조건에 따라 조절하는 강화학습 기반 제어 방법을 제안한다.

2. 조건 기반 동작 변형 연구

조건 기반 동작 변형은 입력 조건에 따라 기존 동작의 스타일, 신체 부위별 움직임, 방향성, 표현 강도를 조절하는 방법이다. Jang et al.^[5]은 Motion Puzzle을 통해 신체를 여러 부위로 나누고 각 부위에 서로 다른 스타일을 적용한 뒤 결합하는 모션 스타일 전이 방법을 제안하였다. 이를 통해 하체의 보행은 유지하면서 상체에 다른 스타일을 적용하는 것과 같은 세밀한 편집이 가능하다. Aberman et al.^[6]은 동작의 내용과 스타일을 분리하여 대응되는 모션 쌍 없

이도 참조 영상이나 동작의 스타일을 다른 동작에 적용하였다. Huang et al.^[7]은 CoMo에서 동작을 신체 부위별 의미를 갖는 Pose code로 표현하고 자연어 명령을 이용해 특정 관절이나 시간 구간을 편집하였다. Athanasiou et al.^[8]의 MotionFix는 기존 동작과 텍스트 편집 명령을 조건으로 사용하여 신체 부위 속도, 반복 횟수, 스타일을 변경하는 동작 편집 방법을 제안하였다.

그러나 기존 연구는 주로 스타일 전이, 신체 부위별 자체 편집, 자연어 기반 의미 변형에 초점을 둔다. 반면 본 연구는 BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 통해 하나의 기본 주기 동작의 반복 주기와 특정 시점에서의 관절 운동 반응을 외부 조건에 따라 조절한다는 점에서 기존 연구와 차이가 있다.

III. BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 이용한 강화학습 기반 주기적 동작 변형 방법

본 연구에서는 BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 이용하여 주기적 인간 동작을 변형하는 강화학습 기반 제어 방법을 제안한다. 제안 방법은 물리 기반 시뮬레이션 환경에서 에이전트가 주어진 조건에 따라 동작을 수행하도록 정책을 학습한다.

제안 방법의 전체 과정은 그림 1과 같다. 먼저 원본 모션 데이터에서 추출한 기본 주기 동작과 외부 조건인 BPM, Target Position, Target Timing이 입력된다. 기본 주기 동작은 에이전트가 원본 동작의 자연스러움을 유지하기 위한

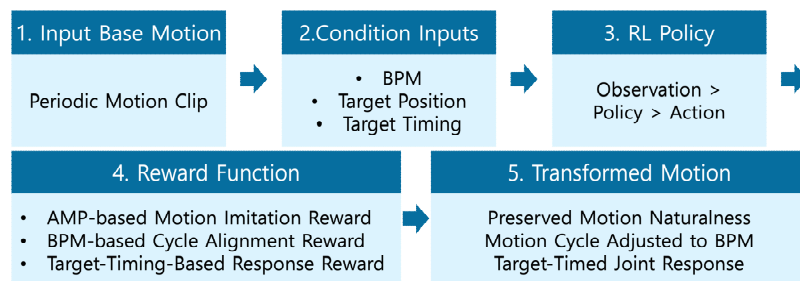


그림 1. BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 이용한 강화학습 기반 주기적 동작 변형 과정

Fig. 1. Reinforcement learning-based periodic human motion transformation using BPM-based cycle alignment and target-timing response rewards

참조 모션으로 사용된다. BPM은 동작의 목표 반복 주기를 결정하는 시간 조건이며 Target Position은 관절의 이동 방향을 위한 공간적 목표이고, Target Timing은 반응 보상이 활성화되는 시점이다.

에이전트는 현재 신체 상태와 조건 정보를 관측값으로 입력받고, 강화학습 정책을 통해 각 시간 단계의 관절 제어 행동을 출력한다. 생성된 동작은 AMP 기반 모션 모방 보상, BPM 기반 주기 정렬 보상, Target Timing 기반 목표 반응 보상으로 평가되며, 각 보상은 각각 원본 동작의 자연스러움 유지, 반복 주기 조절, 목표 방향 운동 반응을 유도한다. 정책은 이들 보상의 결합값을 최대화하도록 학습되어 BPM과 Target Timing 조건이 반영된 변형 동작을 생성한다.

1. 전체 동작 변형 구조

제안하는 동작 변형 구조는 입력 모션 데이터, 조건 입력, 강화학습 정책, 물리 기반 시뮬레이션 환경, 보상 함수, 변형 동작 출력으로 구성된다. 입력 모션 데이터는 걷거나 팔 흔들기와 같이 반복적 형태를 갖는 주기 동작이며, 에이전트가 기본 동작의 특성을 학습하기 위한 참조 모션으로 사용된다. 에이전트의 관측 정보는 root 높이와 회전, root 선형 및 각속도, 관절 위치와 속도, 주요 신체 부위의 위치와 같은 신체 상태와 BPM, Target Position, Target Timing의 외부 조건 정보로 구성된다.

AMP 관측값은 root 상태, 관절 자유도 정보 및 주요 신체 부위 위치를 기반으로 구성되며, 생성 동작과 참조 모션의 동작 분포를 비교하는 데 사용된다. 각 시간 단계에서 정책은 관측값을 입력받아 관절 제어 행동을 출력하고, 물리 기반 시뮬레이션 환경은 이를 적용하여 다음 상태와 보상값을 계산한다. 정책은 누적 보상을 최대화하도록 갱신되며, 각 시뮬레이션 단계에서 AMP 관측값도 함께 갱신된다.

본 연구에서는 AMP 기반 모션 모방 구조에 BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 결합하였다. 이를 통해 기본 동작의 특성을 유지하면서 입력 BPM에 따라 반복 주기를 조절하고, 지정된 Target Timing 부근에서 제어 대상 관절이 목표 방향으로 운동 반응을 나타내도록 유도한다.

2. BPM 기반 주기 정렬 보상

주기적 인간 동작은 일정한 시간 간격으로 반복되는 특성을 가진다. 본 연구에서는 이러한 반복성을 외부 BPM 조건과 정렬하기 위해 BPM 기반 주기 정렬 보상을 설계하였다.

$$T_{interval} = \frac{60}{BPM} \quad (1)$$

$$E_{cycle} = (T_{cycle} - T_{interval})^2 \quad (2)$$

$$R_{cycle} = \exp\left(-\frac{E_{cycle}}{2\sigma_c^2}\right) \quad (3)$$

BPM은 1분 동안 발생하는 Beat 수를 의미하며 이를 이용하여 목표 Beat Interval $T_{interval}$ 을 계산한다. $T_{interval}$ 은 수식 (1)과 같으며 하나의 Beat가 발생하는 시간 간격을 의미한다. 예를 들어 100BPM이 입력된 경우 $T_{interval}$ 은 0.6초가 된다. 동작의 실제 반복 주기는 특정 관절의 위치 변화에서 검출된 연속 피크 사이의 시간 간격으로 정의한다. 본 연구에서는 손 관절의 z축 위치 변화를 사용하여 동작의 상승 및 하강 패턴을 분석하고 연속된 피크 사이의 시간 차이를 실제 동작 주기 T_{cycle} 로 정의하였다. 실제 동작 주기와 목표 Beat Interval 사이의 오차 E_{cycle} 은 수식 (2)와 같이 계산한다. 수식 (2)에서 T_{cycle} 은 실제 동작에서 측정된 반복 주기이며 $T_{interval}$ 은 BPM에서 계산된 목표 주기이다. E_{cycle} 값이 작을수록 실제 동작의 반복 주기가 입력 BPM 조건과 가깝게 정렬된 것을 의미한다. 최종적인 BPM 기반 주기 정렬 보상은 주기 오차가 적을수록 높은 값을 갖도록 수식 (3)과 같은 가우시안 형태로 정의된다. 수식 (3)에서 R_{cycle} 은 BPM 기반 주기 정렬 보상이며 σ_c 는 주기 오차에 대한 허용 범위를 조절하는 파라미터이다. σ_c 가 작을수록 목표 Beat Interval과 실제 동작 주기의 정확한 일치를 강하게 요구하며 σ_c 가 클수록 오차에 대해 완화된 보상을 제공한다.

3. Target Timing 기반 목표 반응 보상

Target Timing 기반 목표 반응 보상은 BPM 조건에 따라 생성되는 Target Timing에서 손 관절이 목표 위치 방향으로 움직이도록 유도하기 위해 사용된다. 특정 시점 t 에서 제어 대상 관절의 현재 위치를 $P_{joint}(t)$, 목표 위치를 $P_{target}(t)$ 라고 할 때 위치 오차 $E_{target}(t)$ 는 수식 (4)와 같이 정의한다.

$$E_{target}(t) = \|P_{joint}(t) - P_{target}(t)\|^2 \quad (4)$$

$$R_{target}(t) = \exp\left(-\frac{E_{target}(t)}{\sigma_t^2}\right) \quad (5)$$

수식 (4)에서 $P_{joint}(t)$ 는 시간 t 에서 특정 관절의 3차원 위치를 의미하며 $P_{target}(t)$ 는 해당 관절에 설정된 Target Position을 의미한다. $E_{target}(t)$ 값이 작을수록 현재 관절 위치와 설정된 Target Position 사이의 거리가 작음을 의미한다. Target Timing 기반 목표 반응 보상은 위치 오차가 작을수록 높은 값을 갖도록 수식 (5)와 같이 정의한다. R_{target} 은 시간 t 에서 Target Timing 기반 목표 반응 보상이며 σ_t 는 목표 위치 오차에 대한 허용 범위를 조절하는 파라미터이다. 제어 대상 관절과 Target Position 사이의 거리가 가까울수록 높은 보상이 부여되고, 거리가 증가할수록 보상은 감소한다. 이를 통해 에이전트는 주기적 움직임을 유지하면서 Target Timing 부근에서 제어 대상 관절이 설정된 방향으로 움직이도록 학습한다.

본 연구에서 Target Position은 Beat 구간 진입 시점의 관절 위치를 기준으로 상대적으로 설정되며, 절대 공간 좌표상의 고정된 목표 위치와는 차이가 있다. 따라서 본 보상은 절대적인 목표 위치 도달보다는 Target Timing 부근에서 특정 방향의 관절 운동 반응을 유도하는 데 사용된다.

4. 최종 보상 함수 및 정책 학습

$$R_{total} = w_{amp}R_{amp} + w_cR_{cycle} + w_tR_{target} \quad (6)$$

본 연구의 최종 보상 함수는 원본 모션의 자연스러움을

유지하기 위한 AMP 기반 모션 모방 보상, BPM 조건에 따른 주기 정렬 보상, Target Timing 조건에 따른 목표 반응 보상을 결합하여 구성한다. 수식 (6)과 같이 R_{total} 은 최종 보상으로, R_{amp} 는 AMP 기반 모션 모방 보상, R_{cycle} 은 BPM 기반 주기 정렬 보상, R_{target} 은 Target Timing 기반 목표 반응 보상을 의미한다. w_{amp} , w_c , w_t 은 각 보상 항목에 대한 중요도를 조절하는 파라미터이다. AMP 기반 모션 모방 보상 R_{amp} 는 생성된 동작이 원본 모션 데이터의 자연스러운 움직임 분포를 유지하도록 유도한다. BPM 기반 주기 정렬 보상 R_{cycle} 은 생성된 동작의 반복 주기가 입력 BPM 조건과 정렬되도록 하며 Target Timing 기반 목표 반응 보상 R_{target} 은 지정된 시점 부근에서 특정 관절이 설정된 Target Position 방향으로 움직이도록 유도한다. 정책은 최종 보상 R_{total} 을 최대화하는 방향으로 학습되며 이를 통해 원본 동작의 자연스러움을 유지하면서도 BPM 조건과 Target Timing 조건이 반영된 변형 동작을 생성한다. 각 보상 항의 가중치는 보상 값의 범위와 학습 안정성을 고려하여 경험적으로 설정하였으며, 각 항에 대해 별도의 정규화는 적용하지 않았다. 각 보상 항이 최종 동작에 미치는 역할을 분석하기 위해 실험에서는 R_{amp} , R_{cycle} , R_{target} 의 구성에 따른 결과를 비교하였다. 다만 일부 조건에서는 R_{target} 의 추가와 함께 R_{cycle} 의 가중치가 조정되었으므로, 해당 조건에서 나타난 팔 동작 반복 속도의 차이를 R_{target} 단독의 효과로 해석하지 않는다.

IV. 실험 및 결과

1. 실험 구성

본 실험은 Linux 환경에서 NVIDIA Isaac Sim을 이용하여 수행하였다. Isaac Sim은 물리 기반 시뮬레이션 환경을 제공하며 본 연구에서는 휴머노이드 에이전트가 하나의 기본 주기 동작을 기준으로 BPM, Target Timing 조건에 따라 시간적 반복 특성과 관절 운동 반응을 변화시키는지를 분석하였다. 에이전트는 관절 위치, 관절 속도, 신

체 부위 위치 등의 상태 정보를 관측값으로 입력받고 학습된 정책은 각 시간 단계에서 관절 제어 행동을 출력한다.

2. BPM 조건에 따른 주기 동작 변형 결과

BPM 조건에 따른 실험은 하나의 기본 주기 동작에 서로 다른 BPM 값을 적용하여 동작의 반복 주기가 입력 조건에 따라 변화하는지를 분석하였다. BPM 값이 증가할수록 Beat Interval은 짧아지고 BPM 값이 감소할수록 Beat Interval은 길어진다. 이에 따라 각 BPM 조건에서 생성된 오른손 관절의 반복 움직임을 분석하여 동작 주기의 변화 양상을 확인하였다. 그림 2는 실험에 사용된 기본 주기 동작을 나타내며, 그림 3은 AMP 기반 모션 모방을 통해 학습된 기본 동작의 결과를 나타낸다.

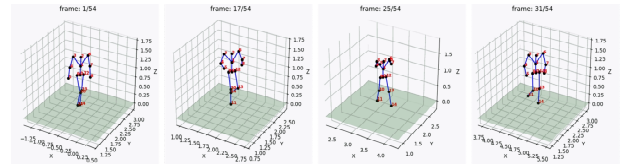


그림 2. 기본 주기 동작 - 양손을 앞뒤로 흔들며 걷기

Fig. 2. Base periodic motion: walking while swinging both hands back and forth

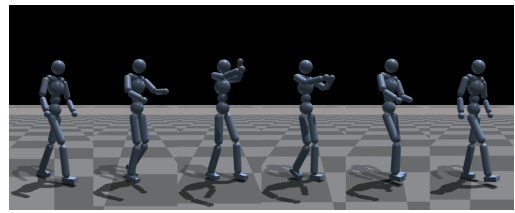


그림 3. 기본 주기 동작 - AMP 기반 동작 모방 결과

Fig. 3. Base motion - AMP-based motion imitation result

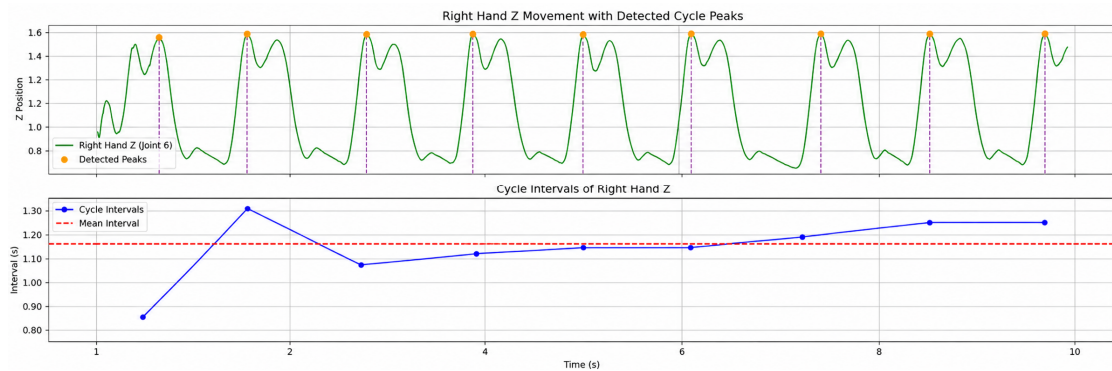


그림 4. 60BPM 조건에서의 주기 변화 결과 - 평균 주기 간격 1.132초

Fig. 4. Cycle variation result under the 60BPM condition - mean interval 1.132s

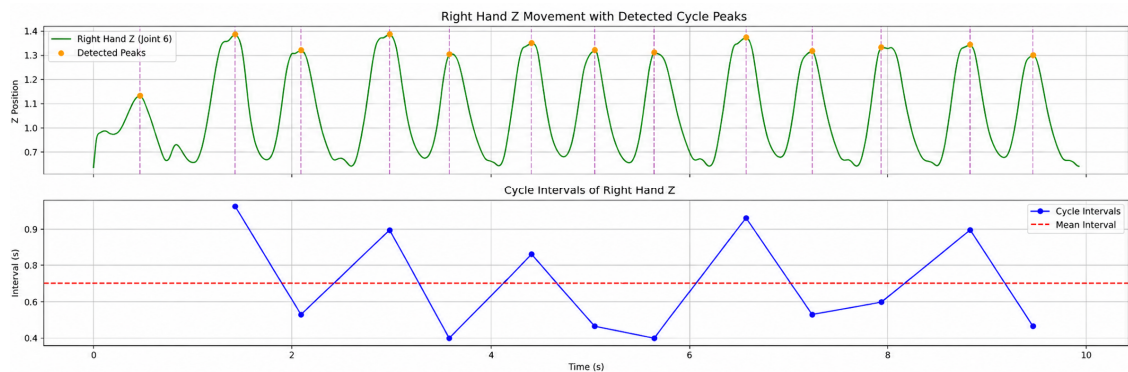


그림 5. 130BPM 조건에서의 주기 변화 결과 - 평균 주기 간격 0.73초

Fig. 5. Cycle variation result under the 130BPM condition - mean interval 0.73s

그림 4와 그림 5는 각각 60BPM과 130BPM 조건에서 생성된 동작의 오른손 관절 z축 위치 변화를 나타낸다. 오른손 관절의 반복 움직임에서 측정된 평균 주기 간격은 60BPM 조건에서 약 1.132초, 130BPM 조건에서 약 0.73초로 나타났다. 이를 통해 입력 BPM이 증가함에 따라 오른손 관절의 반복 움직임에서 주기 간격이 짧아지는 경향을 확인하였다. 다만 입력 BPM과 실제 동작 주기의 동조 여부는 평균 주기 간격만으로 판단하지 않고, 이후 정량 분석에서 팔 동작의 분당 반복 횟수, 목표 템포와의 정수비 관계 및 주기 변동계수를 함께 이용하여 평가하였다.

3. Target Timing 조건에 따른 관절 운동 반응 결과

Target Timing 조건 실험에서는 BPM에 따라 결정되는 Beat Timing 부근에서 오른손 관절이 설정된 목표 방향에 대응하는 운동 반응을 나타내는지를 분석하였다. Target Timing 기반 목표 반응 보상은 지정된 Beat Timing의 ± 0.1 초 구간에서 활성화되며, 해당 구간에 진입한 시점의 오른손 관절 위치를 기준으로 Target Position을 설정하였다.

그림 6과 그림 7은 서로 다른 BPM 조건에서 Beat Timing과 오른손 관절의 z축 속도 변화를 나타낸다. 실험 결과, Target Timing 부근에서 오른손 관절의 속도 변화가 반복적으로 관찰되었다. 이러한 반응의 크기와 시점 일관

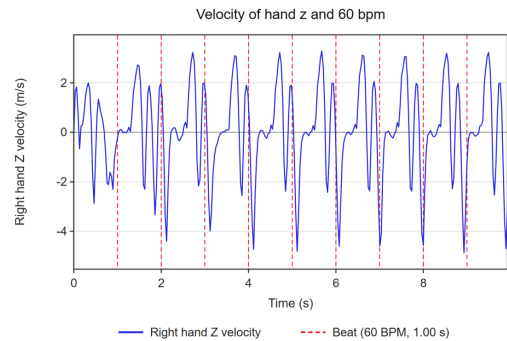


그림 6. 60BPM 조건에서의 Target Timing 기반 관절 운동 반응 - 오른손 관절 z축 속도

Fig. 6. Right-hand joint z-axis velocity response at target timings under the 60BPM condition

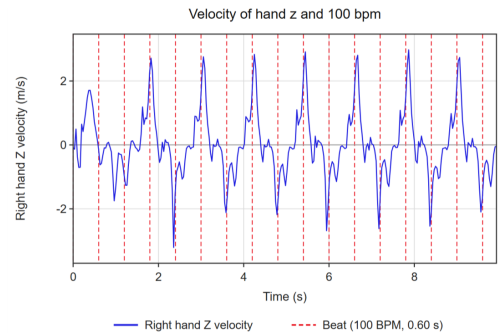


그림 7. 100BPM 조건에서의 Target Timing 기반 관절 운동 반응 - 오른손 관절 z축 속도

Fig. 7. Right-hand joint z-axis velocity response at target timings under the 100BPM condition

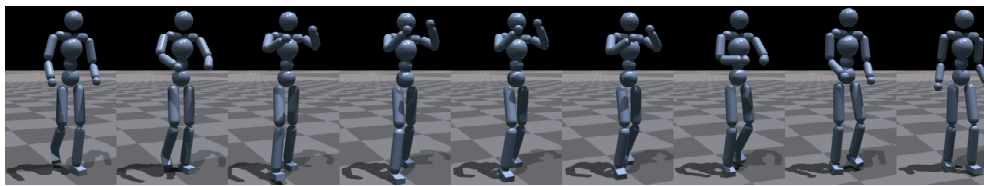


그림 8. 60BPM 조건에서의 동작 변형 결과

Fig. 8. Motion transformation result under the 60BPM condition

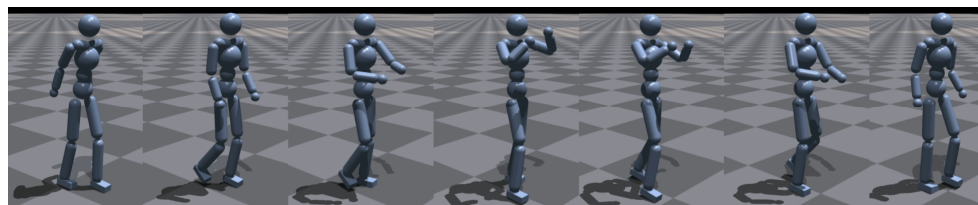


그림 9. 130BPM 조건에서의 동작 변형 결과

Fig. 9. Motion transformation result under the 130BPM condition

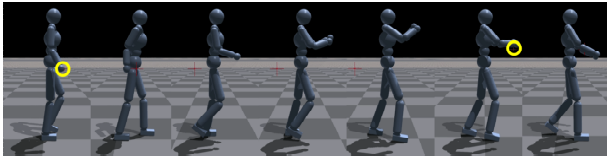


그림 10. 60BPM 조건에서 Beat Timing에 따른 동작 변형 결과
Fig. 10. Motion transformation result according to beat timing under the 60BPM condition

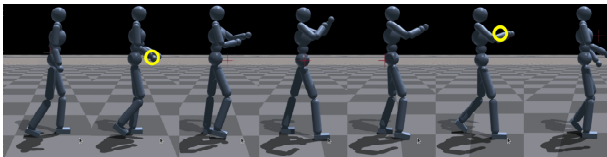


그림 11. 100BPM 조건에서 Beat Timing에 따른 동작 변형 결과
Fig. 11. Motion transformation result according to beat timing under the 100BPM condition

성, 그리고 Target Position과의 거리 오차는 이후 정량 분석에서 추가적으로 평가하였다.

4. BPM 조건에 따른 반복 주기 정량적 분석

BPM 조건에 따른 주기 변화를 분석하기 위해 오른손 관절의 반복 움직임으로부터 평균 반복 주기와 팔 동작의 반복 속도(cycles/min)를 계산하였다. 또한 반복 주기의 규칙성은 주기 변동계수(Coefficient of Variation, CV)로 평가하였다. 주기적 동작은 Beat와 반드시 1:1로 대응하지 않을 수 있기 때문에 본 연구에서는 1:1뿐 아니라 2:3 및 1:2와 같은 정수비 관계도 동조(entrainment)로 고려하였다.

표 1에서 80BPM 조건은 1:1 관계, 100BPM 조건은 보상

표 1. 보상 구성에 따른 팔 동작 반복 속도 및 목표 정수비 오차 비교
Table 1. Comparison of arm-swing rate and target ratio error across reward configurations

Reward configuration	BPM	Arm-Swing Rate (cycles/min)	Error from Target Ratio (%)	Cycle Period CV (%)
R_amp (baseline)	-	58.3	-	3.2
+ R_target	80	80.1	+0.2	3.2
+ R_cycle	100	66.8	+0.2	1.7
+ R_cycle + R_target	100	49.7	-0.6	1.4
+ R_target	100	66.8	+0.1	4.5

구성에 따라 2:3 또는 1:2 관계를 형성하였으며, 목표 정수비와의 오차는 모두 0.6% 이내였다. 본 연구에서는 목표 정수비와의 오차가 2% 이내이고, 기준 조건 대비 반복 속도가 5% 이상 변화하며, 주기 변동계수가 10% 미만인 경우 동조가 나타난 것으로 판단하였다.

5. Target Timing에 따른 관절 운동 반응 정량적 분석

Target Timing에 대한 관절 운동 반응은 각 Beat Timing ± 0.1 초 구간에서 오른손 관절의 z축 이동 범위와 반응 시점을 이용하여 분석하고 동일 길이의 무작위 시간 구간 2,000개와 비교하여 p-value를 계산하였다.

표 2에서 Peak Offset은 Beat Timing과 오른손 관절 z축 속도 피크 사이의 시간 차이를 의미하며 평균 $\pm$ 표준편차로 제시하였다. Target Timing 기반 목표 반응 보상을 적용한 경우, z축 이동 범위가 증가하고 반응 시점의 표준편차가 감소하였다. 특히 100BPM 조건에서 평균 Peak Offset은 -0.001초로 나타났다. 이러한 결과는 목표 반응 보상 적용 시 Beat Timing 부근의 z축 이동 범위가 증가하고 반응 시점의 일관성이 향상되었음을 보여준다.

표 2. Beat Timing에서의 오른손 관절 운동 반응 비교
Table 2. Comparison of right-hand joint responses at beat timings

Condition	BPM	Z-Axis Movement Range (m)	p-value	Peak Offset (s)
R_amp	80	0.177	0.5	+0.281 $\pm$ 0.222
+ R_target	80	0.345	0.0005	+0.236 $\pm$ 0.052
R_amp	100	0.161	0.7	+0.217 $\pm$ 0.177
+ R_cycle	100	0.198	0.064	-0.296 $\pm$ 0.159
+ R_cycle + R_target	100	0.266	0.0005	-0.001 $\pm$ 0.082

6. 보상 항 구성에 따른 비교 분석

최종 보상 함수에 포함된 각 보상 항의 역할을 분석하기 위해 100BPM 조건에서 보상 항 구성에 따른 결과를 비교하였다. 표 3과 같이 AMP 기반 모션 모방 보상만 적용한

표 3. 100BPM 조건에서 보상 항 구성에 따른 동조 비교

Table 3. Comparison of entrainment across reward configurations under the 100BPM condition

Condition	R_{amp}	R_{cycle}	R_{target}	Arm-Swing Rate (cycles/min)	Entrainment
Baseline	✓	-	-	58.3	No
+ Cycle Reward	✓	✓	-	66.8	Yes
+ Target Reward	✓	✓	✓	49.7	Yes

기준 조건에서는 동조가 나타나지 않았으나, BPM 기반 주기 정렬 보상을 추가한 조건에서는 100BPM의 2:3 관계에 해당하는 분당 66.8회의 반복 속도가 나타났다. Target Timing 기반 목표 반응 보상을 적용한 조건에서는 표 2와 같이 Beat Timing 부근에서 관절 운동 반응이 뚜렷하게 나타났다.

이를 통해 AMP 기반 모션 모방 보상은 기본 동작 유지, BPM 기반 주기 정렬 보상은 반복 주기 조절, Target Timing 기반 목표 반응 보상은 특정 시점의 관절 운동 반응 형성에 각각 기여하는 것으로 해석할 수 있다. 다만 목표 반응 보상을 포함한 조건에서는 주기 정렬 보상의 가중치도 함께 조정되었으므로, 반복 속도의 변화를 목표 반응 보상 단독의 효과로 해석하지 않는다.

7. Target Position 거리 오차 분석

Target Position과 오른손 관절 사이의 유클리드 거리 오차도 추가적으로 분석하였다. 목표 반응 보상 조건에서 최소 거리는 80BPM에서 0.153m, 100BPM에서 0.070m였으며, 기준 조건에서는 각각 0.087m와 0.073m로 나타났다. 따라서 목표 반응 보상에 따른 일관된 거리 감소는 확인되지 않았다. 이는 Target Position이 절대 좌표가 아니라 Beat Timing 구간 진입 시 관절 위치를 기준으로 상대적으로 설정되기 때문으로 판단된다. 따라서 본 연구에서는 목표 위치 도달보다는 Beat Timing 부근의 관절 이동 범위와 반응 시점의 일관성을 주요 효과로 해석한다.

추가적으로 난수 시드를 변경하여 반복 학습한 결과, 팔 동작 반복 속도의 표준편차는 30BPM에서 17.6cycles/min, 60BPM에서 12.0cycles/min, 130BPM에서 20.9cycles/min으로 나타났다. 다만 표 1~3에 제시한 대표 결과는 단일 시드 학습 결과를 기반으로 하므로, 향후 다수 시드에 대한

추가적인 안정성 검증이 필요하다.

V. 결 론

본 연구에서는 외부 조건에 반응하여 주기적 인간 동작을 변형하기 위한 강화학습 기반 제어 방법을 제안하였다. 제안 방법은 AMP 기반 모션 모방 구조에 BPM 기반 주기 정렬 보상과 Target Timing 기반 목표 반응 보상을 결합하여 기본 동작의 자연스러움을 유지하면서 반복 주기와 특정 시점의 관절 움직임을 조절하도록 설계하였다.

실험 결과, 일부 BPM 조건에서 팔 동작의 반복 속도가 목표 템포와 정수비 관계를 형성하였으며, 목표 정수비와의 오차는 0.6% 이내로 나타났다. 또한 Target Timing 기반 목표 반응 보상을 적용한 경우 Beat Timing 부근에서 오른손 관절의 z축 이동 범위가 증가하고 반응 시점의 변동이 감소하였다. 보상 항별 비교를 통해 BPM 기반 주기 정렬 보상은 반복 주기 조절에, Target Timing 기반 목표 반응 보상은 특정 Beat Timing에서의 관절 운동 반응 형성에 각각 기여함을 확인하였다.

다만 Target Position과 관절 사이의 거리 오차에서는 일관된 감소가 확인되지 않았다. 또한 본 실험은 주기적 특성을 명확하게 분석하기 위해 팔의 반복 움직임을 일정하게 나타내는 하나의 기본 동작을 대상으로 수행하였으며, 팔의 반복 움직임을 포함하는 다른 주기 동작에 대한 추가 실험은 수행하지 못하였다. 따라서 향후에는 다양한 주기 동작과 반복 학습을 통해 제안 방법의 일반화 성능과 안정성을 검증하고, 절대 좌표 기반의 Target Position을 적용하여 목표 위치 도달 성능을 평가할 예정이다. 또한 음악의 Beat와 Dynamics 조건을 함께 활용하여 동작의 반복 주기, 관절 반응, 진폭 및 강도를 제어하는 방식으로 확장하고자 한다.

참 고 문 헌 (References)

- [1] X. B. Peng P. Abbeel S. Levine and M. van de Panne “DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills” ACM Transactions on Graphics vol. 37 no. 4 Article 143 pp. 1 - 14 2018.
doi: <https://doi.org/10.1145/3197517.3201311>
- [2] N. Chentanez M. Müller M. Macklin V. Makovychuk and S. Jeschke “Physics-Based Motion Capture Imitation with Deep Reinforcement Learning” Proceedings of the 11th ACM SIGGRAPH Conference on Motion Interaction and Games Limassol Cyprus Article 1 pp.1-10 2018.
doi: <https://doi.org/10.1145/3274247.3274506>
- [3] X. B. Peng Z. Ma P. Abbeel S. Levine and A. Kanazawa “AMP: Adversarial Motion Priors for Stylized Physics-Based Character Control” ACM Transactions on Graphics Vol.40 No.4 Article 144 pp.1-20 July 2021.
doi: <https://doi.org/10.1145/3450626.3459670>
- [4] X. B. Peng Y. Guo L. Halper S. Levine and S. Fidler “ASE: Large-Scale Reusable Adversarial Skill Embeddings for Physically Simulated Characters” ACM Transactions on Graphics Vol.41 No.4 Article 94 pp.1-17 July 2022.
doi: <https://doi.org/10.1145/3528223.3530110>
- [5] D.-K. Jang S. Park and S.-H. Lee “Motion Puzzle: Arbitrary Motion Style Transfer by Body Part” ACM Transactions on Graphics Vol.41 No.3 pp.1-16 June 2022.
doi: <https://doi.org/10.1145/3516429>
- [6] K. Aberman Y. Weng D. Lischinski D. Cohen-Or and B. Chen “Unpaired Motion Style Transfer from Video to Animation” ACM Transactions on Graphics Vol.39 No.4 Article 64 2020.
doi: <https://doi.org/10.1145/3386569.3392469>
- [7] Y. Huang W. Wan Y. Yang C. Callison-Burch M. Yatskar and L. Liu “CoMo: Controllable Motion Generation Through Language Guided Pose Code Editing” Proceedings of the 18th European Conference on Computer Vision Milan Italy Part XXIX pp.180-196 2024.
doi: https://doi.org/10.1007/978-3-031-73397-0_11
- [8] N. Athanasiou A. Cseke M. Diomatari M. J. Black and G. Varol “MotionFix: Text-Driven 3D Human Motion Editing” SIGGRAPH Asia 2024 Conference Papers Tokyo Japan Article 44 pp.1-11 2024.
doi: <https://doi.org/10.1145/3680528.3687559>

저 자 소 개

박 성 실



- 2014년 : 덕성여자대학교 디지털미디어학과 (학사)
- 2016년 : 중앙대학교 첨단영상대학원 영상학과 영상공학-CV/VR (석사)
- 2017년 ~ 2019년 : 한국전자기술연구원 연구원
- 2019년 ~ 2021년 : (주)싸인텔레콤 개발부 과장
- 2026년 : 중앙대학교 첨단영상대학원 영상학과 영상공학-CG/VR (박사)
- 2026년 ~ 현재 : 숭실대학교 전자정보공학부 및 G-LAMP 사업단 박사후연구원
- ORCID : <https://orcid.org/0000-0002-6113-6919>
- 주관심분야 : 의료 멀티모달 학습, 인간 동작 생성, 생성형 모델