

손 동작 인식 기반 실시간 가상 악기 연주 및 합주 플랫폼

권주연*¹, 김나연*², 박효빈*³, 장지인*⁴, 김유나*⁵, 강지현**

Hand Gesture Recognition-based Real-Time Virtual Musical Instrument Performance and Ensemble System

Ju-Yeon Kwon*¹, Na-Yun Kim*², Hyo-Bin Park*³, Ji-In Chang*⁴, Yu-Na Kim*⁵, and Jiheon Kang**

본 연구는 2025년도 덕성여자대학교 교내연구비 지원에 의해 이루어졌음

요약

최근 디지털 기술의 발전과 비대면 콘텐츠 수요 증가에 따라, 물리적 악기 없이도 음악을 즐기고 협업할 수 있는 새로운 형태의 연주 방식이 요구되고 있다. 본 시스템은 별도의 장비 없이 모바일 기기만으로 사용자의 손 움직임을 인식하고, 이를 기반으로 피아노와 드럼 등 다양한 악기를 연주할 수 있도록 설계되었다. 또한, 네트워크 기반의 실시간 합주 기능을 통해 서로 다른 장소에 있는 사용자들이 함께 참여할 수 있는 합주 환경을 제공한다. 이를 통해 기존의 음악 연주가 지닌 비용, 공간, 시간 등의 제약을 줄이고, 누구나 쉽게 음악 활동에 참여할 수 있는 접근성을 높이는 데 기여하고자 한다. 본 시스템은 교육적 활용뿐 아니라 음악치료, 디지털 콘텐츠 제작, 원격 공연 등 다양한 분야에서의 확장 가능성을 갖고 있다. 향후에는 사용자 경험 개선과 더불어, 고도화된 기능 개발을 통해 음악과 기술의 융합 영역을 넓혀갈 수 있을 것으로 기대된다.

Abstract

This study proposes a virtual musical instrument performance and real-time ensemble system based on hand recognition. As digital technologies advance and demand for contactless content grows, new forms of music interaction that do not require physical instruments are needed. The system enables users to play instruments like piano and drums through hand movements detected by a mobile device, without extra hardware. It also supports real-time ensemble over a network, allowing people in different locations to perform together. By lowering barriers such as cost, space, and time, the system broadens access to musical activities. With applications in music education, therapy, digital content creation, and remote performances, and with further enhancements, it can expand the integration of music and technology.

Keywords

hand recognition, virtual piano, virtual drum, real-time online ensemble

* 덕성여자대학교 소프트웨어전공 학사과정
- ORCID¹: <https://orcid.org/0009-0001-3210-8986>
- ORCID²: <https://orcid.org/0009-0005-3148-2756>
- ORCID³: <https://orcid.org/0009-0001-9463-384X>
- ORCID⁴: <https://orcid.org/0009-0008-9055-8847>
- ORCID⁵: <https://orcid.org/0009-0008-6990-9622>

** 덕성여자대학교 소프트웨어전공 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-5423-3895>

· Received: Sep. 22, 2025, Revised: Oct. 16, 2025, Accepted: Oct. 19, 2025
· Corresponding Author: Jiheon Kang
Dept. of Software, Duksung Women's University, South Korea
Tel.: +82-2-901-8452, Email: jhkang@duksung.ac.kr

I. 서 론

최근 디지털 기술이 빠르게 발전하고 비대면 콘텐츠와 모바일 플랫폼의 사용이 일상화되면서 음악 연주 방식 또한 새로운 변화가 요구되고 있다[1][2]. 음악 연주는 단순한 예술 활동을 넘어, 표현력 향상과 정서적 안정, 창의성 증진 등 다양한 교육적·심리적 효과를 지닌 중요한 활동으로 널리 알려져 있다[3]. 그러나 실제 악기를 배우고 연주하는 데에는 현실적인 장벽이 존재한다. 악기를 구매하는 데 드는 고가의 비용, 악기 보관 및 운반의 번거로움, 그리고 소음 문제로 인해 연습 공간을 확보하기 어려운 점 등은 많은 이들이 음악을 가까이하기 어렵게 만드는 요인이다.

특히 음악 교육이나 합주 활동은 여전히 대면 중심으로 이루어지는 경우가 많아, 시간이나 장소의 제약으로 인해 참여 자체가 제한되는 사례가 빈번하다. 이러한 문제를 해결하기 위해서는 언제 어디서나 접근 가능하며, 실제 악기와 유사한 몰입감 있는 연주 경험을 제공할 수 있는 물리적 악기에 대한 의존도를 낮춘 가상환경과 같은 새로운 대안적 시스템 개발의 필요성이 대두되고 있다[3].

이에 본 연구는 영상분석 기반 손 인식 기술을 활용하여 가상 악기 연주 및 합주 시스템을 제안한다. 본 시스템은 사용자가 별도의 물리적 악기나 복잡한 장비 없이도 모바일 환경에서 피아노와 드럼을 연주할 수 있도록 설계되었다. 또한, 네트워크 기반의 합주 기능을 통해 물리적 거리에 상관없이 여러 사용자가 동시에 실시간으로 함께 연주할 수 있는 환경을 제공한다.

본 시스템은 손 인식 기술을 바탕으로 직관적인 연주 인터페이스를 구현하여, 모바일 환경에서도 실제 악기와 유사한 연주 경험을 제공하는 것을 목표로 한다. 나아가, 실시간 반응성과 낮은 지연 시간을 확보하고, 네트워크를 통해 다수의 사용자가 동시에 협업할 수 있는 환경을 조성하였다.

이러한 시스템을 구현하기 위해 본 연구는 Google의 MediaPipe Hand Landmark Detection 기술을 활용하여 손의 위치와 움직임을 실시간으로 추적하고, 이를 MIDI 프로토콜과 FluidSynth 음원 엔

진을 통해 실제 소리로 변환하는 구조를 채택하였다. 더불어, 다수의 사용자 간 실시간 데이터 송수신이 가능하도록 Photon Engine 기반의 네트워크 구조를 구성하여 원격 환경에서도 협연이 가능하도록 하였다.

II. 관련 연구

가상 악기 연주 시스템은 공간적 제약이나 물리적 악기 소유의 한계를 극복하기 위해 컴퓨터 비전, 인간-컴퓨터 상호작용(HCI), 디지털 음향 처리, 실시간 네트워크 통신 등 다양한 기술 융합을 통해 발전해 왔다[4]. 본 절에서는 손 인식 기반 가상 악기 시스템, 영상 분류 기반 연주 기술, MIDI 기반 음향 처리 시스템, 그리고 실시간 네트워크 합주 시스템에 대한 주요 선행연구를 검토하고, 본 연구와 관련된 연구들과의 차별성을 기술한다.

2.1 손동작 인식 기법

최근 손동작 인식 기술은 AI와 컴퓨터 비전의 발전에 힘입어 빠르게 고도화되고 있다. 대표적으로 센서 기반 방식과 영상 기반 방식이 공존하지만, 최근에는 별도의 하드웨어 없이 카메라만으로 손의 움직임을 인식하는 영상 기반 접근이 주류로 자리 잡고 있다[5].

영상 기반 손동작 인식은 이미지나 영상에서 손의 주요 지점(키포인트)을 추출하고, 이들의 공간적·시간적 변화를 분석해 다양한 제스처를 할 수 있다. 예를 들어, YOLO11이나 MediaPipe와 같은 최신 모델은 손목, 손가락 관절, 손끝 등 20개 이상의 키포인트를 실시간으로 추정하며, 이를 바탕으로 핀치, 스와이프, 원형 동작 등 복잡한 손동작을 높은 정확도로 구분할 수 있다.

이러한 기술은 스마트폰, 태블릿, 웹캠 등 범용 카메라만으로도 구현 가능해 접근성이 뛰어나며, AR/VR, 스마트홈, 수화 번역, 게임, 원격 제어 등 다양한 분야에서 활용되고 있다[6]. 특히 MediaPipe는 경량화된 구조와 높은 실시간성, 다양한 플랫폼 지원으로 연구와 산업 현장에서 널리 사용되고 있다.

핸드 트래킹과 제스처 인식의 정확도는 데이터셋의 다양성, 조명 변화 대응력, 사용자별 적응성 등에 따라 좌우되며, 최근에는 딥러닝 기반의 키포인트 추정과 시계열 분석을 결합해 더욱 정교한 인식이 가능해지고 있다. 이처럼 영상 기반 손동작 인식 기술은 하드웨어 부담을 줄이면서도 실시간성과 정확성을 동시에 확보하며, 인간-컴퓨터 상호작용의 새로운 패러다임을 제시하고 있다[7].

2.2 가상 악기 연주 시스템

기존의 손동작 인식 기반 가상 악기 시스템은 다양한 하드웨어 센서를 활용하여 손가락 관절을 움직임을 패턴을 추적하고, 그 정보를 기반으로 소리를 출력하는 방식으로 구현되었다. 대표적으로, 적외선 카메라를 활용한 시스템에서는 사용자의 손에 부착된 반사 테이프의 위치 정보를 두 대의 카메라로 실시간 추적하고, 이를 기반으로 3차원 좌표의 이동을 분석하여 타격 위치와 강도를 판단해 음을 출력하였다[8]. 이 방식은 높은 정확도를 제공하지만, 적외선 카메라, 반사 테이프와 같은 추가적인 하드웨어 장비에 대한 의존도가 높아 일반 사용자의 접근성이 제한된다.

한편, 센서 없이 영상 정보만으로 손 제스처를 분석하는 비디오 분류 기반의 가상 피아노 시스템도 존재한다. 녹화된 영상 내 손 위치와 건반 영역을 인식하고, 건반 위 그림자의 밝기 및 면적을 분석하여 타건 여부와 강약을 추정하는 방식이다[9]. 이러한 방식은 장비 의존도가 낮고 구현이 간편하지만, 실시간 처리의 한계가 있다.

2.3 실시간 합주 시스템

실시간 합주 시스템 분야는 네트워크 기반 멀티 디바이스 환경에서 서로 다른 위치의 사용자들이 동기화된 연주를 수행할 수 있도록, MIDI 이벤트를 낮은 지연으로 전송·재생하는 기술 개발에 집중해 왔다.

웹 기반 MMMD(Multi Media Multi Device) 환경에서 MIDI 파일을 활용한 AET(Analyzer - Extractor

- Transmit) 프로세스는 Analyzer 단계에서 디바이스 환경과 컨텍스트를 분석하고, Extractor 단계에서 악기별 MIDI 트랙을 분리·추출한 뒤 Transmit 단계에서 이를 각 재생 장치로 전송한다. 특히 MIDI 파일 내 쉼표(Rest point) 구간의 시간 정보(Acrp)와 사전 정의된 기준(Rest point-time, Asrp)을 비교하여 발생한 지연(Gap)을 보정함으로써 기기 간 재생 시점 차이를 최소화하는 동기화 알고리즘을 구현하였다[10].

UDP 전송 기반 접근의 RemoteMIDI VST 연구에서는 DAW(Digital Audio Workstation) 환경에서 VSTi 플러그인 형태로 동작하는 RMV(Remote MIDI VSTi) 모듈을 설계·구현하였다. 이 모듈은 MIDI 메시지를 분석한 뒤 Ethernet 기반의 프로토콜을 통해 전송·연결함으로써, 기존 DAW 시스템의 확장성과 응용성을 보완하는 것을 목표로 한다[11].

III. 손 인식 기반 가상 악기 구현

3.1 개발환경 및 시스템 구성

그림 1은 본 연구에서 제안하는 가상 악기 연주 및 합주 시스템의 구조를 나타낸다. 시스템은 사용자 손의 움직임을 인식하고 가상 피아노 및 드럼을 연주할 수 있도록 설계되었다. 또한, Photon 서버를 기반으로 실시간 연주 데이터를 사용자 간에 전송함으로써 원격으로 합주하는 기능을 지원한다. 사용자는 별도의 하드웨어 없이 모바일 환경에서 연주 및 합주 기능을 사용할 수 있다.

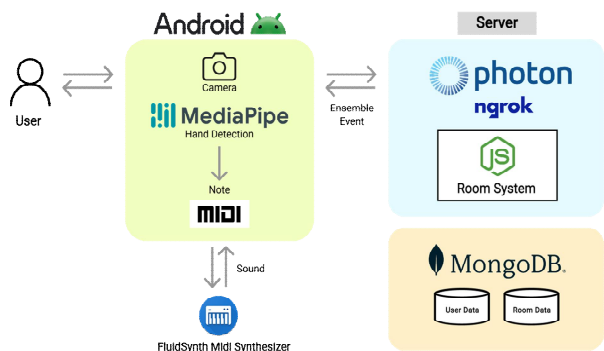


그림 1. 시스템 구조도
Fig. 1. System architecture

3.2 Mediapipe 기반 손 관절 인식

그림 2와 같이 손의 움직임을 인식하기 위해 Google에서 제공하는 MediaPipe 프레임워크의 Hand Landmarker 모듈을 활용한다. 해당 모듈은 영상 내의 양손 각각에 대해 총 21개의 관절 위치를 3차원 좌표(x, y, z)로 반환하여 프레임 단위로 손가락의 움직임을 실시간으로 추적한다. 본 시스템에서는 손 끝 관절들인 [4, 8, 12, 16, 20]번 landmark를 추적 대상으로 설정하여 연주 이벤트를 판단한다.

손 추적 정확도를 높이기 위해 min_hand_presence_confidence 및 min_tracking_confidence 값을 0.5 이하로 설정하고, GPU 가속을 활성화하여 평균 30fps 이상의 실시간 프레임 속도를 확보하였다. 이를 통해 손이 인식된 시점부터 MIDI 음 재생 시간까지 평균 50ms 내외로 측정되었다.

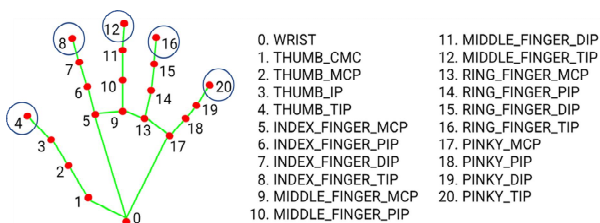
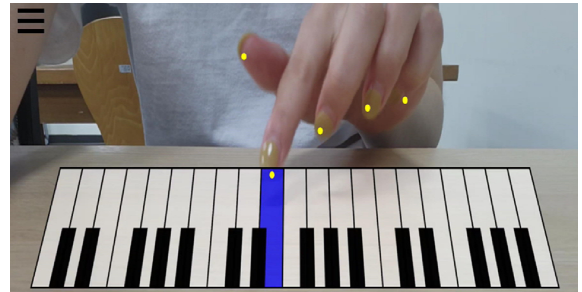


그림 2. Mediapipe 손 랜드마크
Fig. 2. Mediapipe hand landmarker

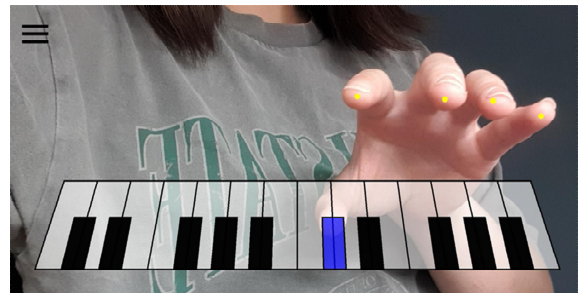
3.3 가상 악기 인터페이스 설계

본 연구는 모바일 환경에 최적화된 가상 피아노 및 드럼 인터페이스를 구현하였다. 전체 UI의 배치 및 구성은 그림 3과 같이 모바일 장치 화면에 AR과 같이 오버레이 되어 표시된다. 그림 3(a)와 3(b)는 화면 하단에 배치된 가상 피아노 인터페이스로, 흰 건반과 검은 건반이 실제 피아노의 음계 패턴을 반영하여 유사한 비율로 배치되며 1~3 옥타브까지 다양하게 표시될 수 있도록 하였다. 그림 3(c)는 사용자가 선택한 드럼 구성에 따라 다양한 드럼 및 심벌즈 이미지가 배치된 예시를 보여준다.

화면에 표시된 노란색 점이 손가락 끝의 좌표를 나타내며, 피아노 건반 및 드럼의 2차원 화면 좌표와의 비교를 통해 특정 영역에 속해있는지, 손가락 끝의 방향성과 속도 변화를 기준으로 해당하는 음이 재생된다.



(a) 가상 피아노 흰 건반 연주
(a) Press a white key on the virtual piano



(b) 가상 피아노 검은 건반 연주
(b) Press a black key on the virtual piano



(c) 가상 스내어 드럼 연주
(c) Hit the snare on the virtual drum

그림 3. 가상 악기 인터페이스
Fig. 3. Virtual instrument interface

본 연구에서는 손 동작 및 위치 인식을 위해 Google에서 개발한 MediaPipe 프레임워크를 활용하였다. MediaPipe는 21개의 손 관절 좌표를 실시간으로 추출할 수 있는 머신러닝 기반 손 추적 기술로, 별도의 센서 없이 스마트폰의 카메라만으로도 손 관절과 손가락 끝의 위치와 이동 속도를 추적할 수 있다. 본 시스템은 이러한 기술을 활용하여, 피아노와 드럼의 연주 영역을 구성하고 손끝의 위치 및 속도를 기반으로 타격 여부를 판별한다. 이를 통해 하드웨어 부담 없이 모바일 환경에서도 실시간 연주가 가능한 인터페이스를 구현하였다.

또한, 음원 출력 방식에서도 mp3 재생 방식이 아닌, MIDI기반의 디지털 메시지를 통해 음을 생성

및 재생하였다. MIDI는 건반 번호, 세기, 길이 등의 정보를 제어할 수 있기 때문에 실제 악기에 가까운 연주 구현이 가능하다. 본 시스템에서는 오픈소스 신디사이저인 FluidSynth를 활용하여, MIDI 메시지를 실시간으로 해석하고 다양한 악기 음색에 대한 출력이 가능하도록 하였다.

사용자는 그랜드 피아노, 일렉트릭 피아노 등 다양한 음색을 선택할 수 있으며, 피아노 건반 수(예: 7, 14, 21개)와 드럼 구성(패드 수, 심벌 포함 여부)도 자유롭게 커스터마이징할 수 있다. 이러한 설정은 내부 ViewModel 구조를 통해 유지되어 매 연주 세션마다 자동으로 적용되어 개인화된 인터페이스 환경을 지속적으로 제공한다.

전체 인터페이스는 Android의 View 기반으로 구현되었으며, 화면 크기와 비율에 따라 유동적으로 조절되는 벡터 그래픽(Path, RectF, Region 등)을 활용하였다. 이를 통해 다양한 해상도와 기기에서도 가상 악기 형태의 왜곡 없이 일관된 시각적 경험을 제공할 수 있다.

3.4 타격 인식 알고리즘 및 소리 출력

그림 4는 본 시스템에서 제안하는 가상 악기의 특정 영역(건반 또는 드럼)에 손이 진입하고, 타격 조건을 만족할 경우 이를 연주 이벤트로 인식하는 절차를 나타낸다. 모바일 단말기에 표시되는 2차원 공간에서 피아노 건반 또는 드럼이 위치한 곳에 단순히 손이 놓여져 있는 것인지, 실제 누름이나 타격인지 판단하기 위하여, 피아노와 드럼 모두 손가락이 뒤에서 아래 방향으로 움직임인지 판단하기 위하여 손가락 끝 좌표 중 y축 위치 변화와 프레임 간 속도(미분값)를 분석하여 손의 하강 속도 및 가속도 변화를 적용하였다. 이를 통해 단순 위치 기반 접근이 아닌, 시간 축 상의 동적 움직임에 따른 정밀한 타격 판별 알고리즘을 구현하였다. 또한, 이전 프레임의 위치 및 속도를 저장하여 손의 떨림이나 무의식적인 미세 이동에 따른 오작동을 방지하였다. 이 방식은 실제 악기 연주에 가까운 직관적이고 자연스러운 사용자 반응을 유도한다.

특히, 피아노의 검은 건반 인식은 입력 충돌을 피하기 위해 별도의 타격 우선순위 기반 알고리즘

을 활용한다. 검은 건반의 타격은 손의 y축 속도가 일정 임계값 이상으로 빠르게 감소한 이후, 다시 정지하거나 속도가 급격히 줄어드는 지점에서 발생한다. 이는 손가락이 흰 건반 영역을 빠르게 통과한 후, 검은 건반 위에서 멈춰야만 타격으로 인식되도록 설계된 것이다. 따라서 흰 건반 위에서는 속도 감소가 충분하지 않아 입력으로 간주되지 않고, 손의 속도가 검은 건반 위에서 급감하거나 정지할 때만 연주 이벤트가 발생한다. 이와 같은 설계 방식을 통해 검은 건반의 좁은 면적과 위치적 특성으로 인해 발생하는 가상 악기 구현의 어려움을 효과적으로 해소할 수 있도록 하였다.

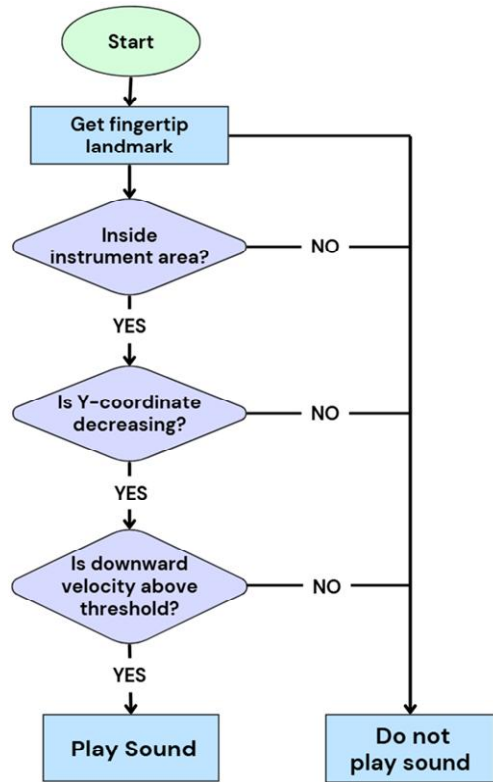


그림 4. 타격 인식 알고리즘
Fig. 4. Press/Hit recognition algorithm

드럼은 이전 프레임의 y좌표와 비교한 낙하 속도가 사전에 정의된 임계값(ex: 25px/frame) 이상이며, 손가락이 특정 드럼 패드 영역 안에 위치해 있을 경우 타격 이벤트로 인식한다. 이는 손이 드럼 패드를 스치거나 정지한 상태에서 발생하는 불필요한 입력을 제거한다. 특히, 드럼 연주 중 손이 UI 영역을 지나쳤다가 다시 올라오는 경우에 중복 타격이 되는 것을 방지한다.

연주 이벤트가 감지되면, 해당 정보는 MIDI (Musical Instrument Digital Interface) 메시지로 변환된다. MIDI는 악기의 음 높이, 강도, 시간 정보를 디지털로 표현하는 국제 표준 프로토콜이다. 본 시스템에서는 이를 FluidSynth 엔진에 전달하여 실시간으로 음을 출력하는 역할을 수행한다. FluidSynth는 미리 설정된 SoundFont(.sf2) 파일을 참조하여 MIDI 메시지에 대응하는 악기 음색을 찾아 출력한다. 이러한 방식 단순한 사운드 클립 기반 시스템보다 음의 지속성과 연주 자연성 측면에서 우수하며, 다양한 음색 전환을 보다 정밀하게 구현할 수 있다는 장점을 가진다.

3.5 실시간 합주

본 시스템은 단일 사용자의 연주에 국한되지 않고, 서로 다른 위치에 있는 사용자들이 동시적으로 하나의 연주에 참여할 수 있는 실시간 합주 기능을 제공한다. 이를 구현하기 위해, 낮은 지연 시간과 높은 안정성을 제공하는 Photon Engine을 기반으로 네트워크 통신 구조를 설계하였다. Photon은 UDP 기반의 경량 프로토콜을 사용하는 실시간 통신 엔진으로, 본 시스템에서는 해당 엔진의 C++ SDK를 기반으로 통신 모듈을 구성하였다.

본 연구에서는 기존 AET 기반 동기화 알고리즘과 UDP 전송 모듈의 한계를 극복하기 위해, Photon Engine을 활용한 경량 실시간 통신 구조를 도입하였다. Photon의 C++ SDK 기반 네트워크 모듈을 Android Kotlin 애플리케이션과 JNI(Java Native Interface)로 연동하여, 사용자가 생성한 MIDI 이벤트를 낮은 지연으로 동기화된 세션에 전송·수신한다. 수신된 이벤트는 UI 처리 루틴과 독립된 전용 스레드에서 우선 처리되어 음 손실 및 타이밍 오류를 방지하며, Photon 룸 시스템과 MongoDB 기반 백엔드(ngrok 터널링 지원)를 통해 원격지 어디서나 안정적인 세션 생성·관리를 지원한다. 이러한 통합적 접근을 통해 본 시스템은 분산 환경에서도 몰입감 있는 실시간 합주를 구현하며, 비대면 음악 교육, 원격 협업 연주, 디지털 창작 도구로서의 활용 가능성을 제공한다.

Android 환경에서 Kotlin으로 개발된 애플리케이션은 JNI(Java Native Interface)를 통해 Photon의 C++ 모듈과 연동된다. 사용자가 연주를 수행하면, Android 앱에서 생성된 MIDI 이벤트가 JNI를 통해 네이티브 C++ 코드로 전달된다. 해당 이벤트는 Photon 네트워크를 통해 동일 세션에 접속한 다른 사용자 단말기로 실시간 전송된다. 수신 측에서는 다시 Kotlin 계층으로 데이터를 전달하여 음원 모듈에서 즉시 재생되도록 하여, 실제로 하나의 공간에서 함께 연주하는 것과 유사한 몰입감을 제공한다.

연주 이벤트는 MIDI note 값, velocity, timestamp 등 필수 정보만을 포함한 경량화된 데이터 구조로 설계되어 전송 지연 및 패킷 손실을 최소화하였다. 또한, 수신된 데이터는 사용자 인터페이스(UI) 처리 루틴과 독립된 전용 스레드에서 우선 처리되도록 하여, 네트워크 지연에 기인한 음 손실 또는 타이밍 오류를 방지하고 원활한 실시간 연주 환경을 보장한다.

본 시스템은 합주 기능의 세션 관리를 위해 방 생성 및 입장 기능을 포함한다. 사용자 간 세션 동기화와 참여자 관리는 Photon의 룸 시스템과 MongoDB 기반 사용자 데이터베이스를 통해 수행된다. 백엔드 서버는 Node.js로 구현되었으며, 이를 통해 실시간 다자간 연주 환경에서 안정적인 세션 유지와 사용자 상태 동기화가 가능하다.

실시간 통신 구조의 안정성 및 체감적인 지연을 검증하기 위해, 네트워크 지연 시간(Latency)을 측정하였다. 연주 이벤트가 발생하면 송신 측에서는 해당 시점의 타임스탬프 T_{send} 를 기록한 뒤, MIDI 이벤트를 Photon 네트워크를 통해 수신 측으로 전송한다. 수신 측에서는 이벤트를 수신한 시점의 타임스탬프 T_{recv} 를 기록하며, 이 두 시점의 차이를 통해 송신/수신측 라운드 트립 지연 시간(Latency)을 계산하였다.

이벤트 식별을 위해 각 MIDI note에 고유한 노트 번호와 발생 시각 정보를 포함시켜 전송하였으며, Android 로그 시스템을 통해 각 단말기의 타임스탬프를 정밀하게 수집하였다. 이 방식을 통해 모바일 환경에서도 실시간 연주 이벤트의 지연 정도를 객관적으로 측정할 수 있었다.

측정 결과, 평균 단방향 지연 시간은 약 127ms로 확인되었으며, 이는 모바일 기반 실시간 합주 애플리케이션에서 사용자가 체감할 수 있는 허용 범위 내에 해당한다고 판단된다. 이러한 성능은 UDP 기반의 경량 통신 구조, JNI 기반 네이티브 모듈 연동, MIDI 이벤트의 경량화된 전송 구조, 그리고 네트워크 처리와 UI 처리를 분리한 비동기 아키텍처 설계가 복합적으로 작용한 결과로 분석된다.

그림 5는 제안된 손 인식 기반 실시간 가상 악기 연주 및 합주 플랫폼을 안드로이드 앱으로 구현하여 원격지에 떨어진 두 사용자가 실시간으로 합주하는 상황을 나타내고 있으며, 실제 사용자 테스트 환경에서도 복수의 사용자가 동시에 연주를 수행하는 과정에서 음 끊김, 타이밍 불일치, 입력 누락 등의 문제가 발생하지 않음을 확인하였다.

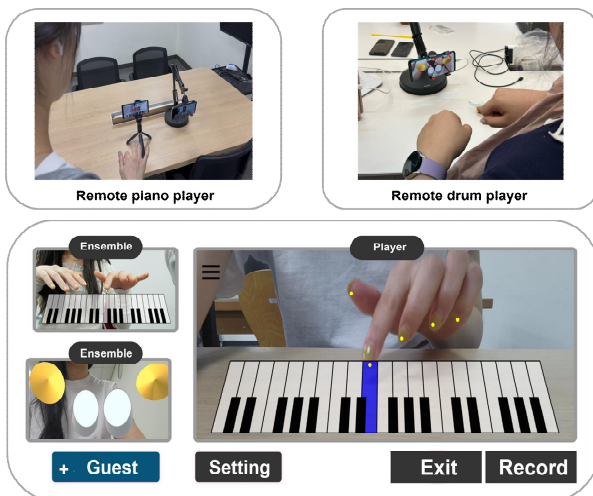


그림 5. 원격지 다중사용자 실시간 합주
Fig. 5. Remote multi-user real-time ensemble

IV. 결론 및 향후 과제

본 논문에서는 손 인식 기술을 기반으로 한 가상 악기 연주 및 실시간 합주 시스템의 구현 내용을 소개하였다. 제안된 시스템은 모바일 기기만으로도 피아노와 드럼 등 다양한 악기를 직관적으로 연주할 수 있으며, 네트워크 기반의 합주 기능을 통해서도 다른 공간의 사용자들이 하나의 연주에 동시에 참여할 수 있도록 설계되었다.

본 시스템은 Google의 MediaPipe Hand Landmarker

를 활용하여 손가락의 세밀한 움직임을 실시간으로 추적하며, 인식된 움직임은 MIDI 프로토콜로 변환되어 FluidSynth 음원 엔진을 통해 실제 악기와 유사한 음향으로 출력된다. 이 과정을 통해 단순한 위치 인식이 아닌 손의 속도, 방향 등의 동적 요소를 반영하여 더욱 자연스럽게 몰입감 있는 연주 경험을 제공한다. 또한 Photon Engine을 기반으로 설계된 네트워크 구조는 낮은 지연 시간과 높은 안정성을 바탕으로 사용자 간 실시간 합주 환경을 구현하였다. 특히 별도의 고가 장비 없이도 모바일 환경에서 원활한 연주가 가능하다는 점은 본 시스템의 높은 접근성과 실용성을 입증한다.

결론적으로, 본 시스템은 인공지능을 활용한 손 인식 기술을 통해 악기 연주의 물리적 한계를 극복하고 누구나 손쉽게 음악을 즐길 수 있는 새로운 방식의 연주 환경을 제시한다. 이는 음악 교육은 물론, 음악치료, 디지털 퍼포먼스, 비대면 공연 등 다양한 분야에서 활용 가능성이 높으며 향후 AR, VR, XR 관련 기술과 결합될 경우 확장성이 뛰어난 플랫폼으로 발전할 수 있을 것으로 기대된다.

본 시스템의 기능 개선을 위해 드럼 구성의 확장, 악기 종류 추가, 피아노의 경우 양손 연주가 가능할 수 있도록 지원할 예정이며, 현재 모바일 단말기에 증강현실처럼 표시되어 화면에 보이는 건반 및 드럼의 위치와 사용자 손의 위치를 연주자가 직접 확인 및 일치시키는 과정에서 발생하는 지연 및 허공을 누르게 되어 실감성이 다소 부족하다는 실험자들의 의견을 반영하여 현실감 있는 인터페이스로 개선할 예정이다.

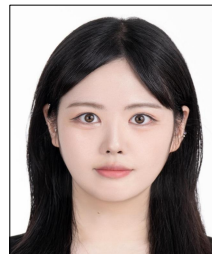
향후 본 시스템은 음악 교육, 디지털 콘텐츠 제작, 원격 공연, 음악 치료 등 다양한 분야에서 실질적인 활용 가능성을 지닌다. 나아가 AR 기술, AI 기반 자동 반주, 음성 인식 기능 등과의 융합을 통해 더욱 몰입되고 직관적인 음악 창작 환경으로 확장될 수 있다. 손 추적 정밀도 향상, 악기 종류 및 연주 방식의 다양화, 사용자 맞춤형 기능 강화 등을 통해 시스템의 완성도를 지속적으로 높이고자 한다. 이러한 기술적 발전을 바탕으로 본 시스템은 물리적 악기의 한계를 넘어서는 새로운 디지털 음악 플랫폼으로 자리매김할 수 있을 것으로 기대된다.

References

- [1] I. J. Yang, "Virtual Musical Instruments on Smart Device and Transformation in Musical Experience", *Culture and Convergence*, Vol. 42, No. 7, pp. 71-100, Jul. 2020. <https://doi.org/10.33645/cnc.2020.07.42.7.71>.
- [2] S. Ppali, M. Scorer, E. Ppali, and B. Branch, "Remote Rhythms: Audience-informed insights for designing remote music performances", *Proc. of the 2024 ACM Designing Interactive Systems Conference*, New York, United State, Jul. 2024. <https://doi.org/10.1145/3643834.366071>.
- [3] E. G. Duarte, I. Cossette, and M. M. Wanderley, "Analysis of Accessible Digital Musical Instruments through the lens of disability models: a case study with instruments targeting d/Deaf people", *Frontiers in Computer Science*, Vol. 5, Oct. 2023. <https://doi.org/10.3389/fcomp.2023.1158476>.
- [4] I. J. Yang, "Virtual Musical Instruments on Smart Device and Transformation in Musical Experience. *Culture and Convergence*, Vol. 42, No. 7, pp. 71-100, Jul. 2020. <https://doi.org/10.33645/cnc.2020.07.42.7.71>.
- [5] A. O. Hashi, S. Z. M. Hashim, and A. B. Asamah, "A Systematic Review of Hand Gesture Recognition: An Update From 2018 to 2024", *IEEE Access*, Vol. 12, pp. 143599-143626, Jul. 2024. <https://doi.org/10.1109/ACCESS.2024.3421992>.
- [6] J. Yu, H. Park, and Y. Lee, "An Artificial Intelligence Hand Recognition Interface Using Virtual Distance Detection Sensor based On Mediapipe", *Proc. of 2023 KIIT Summer Conference*, Jeju, South Korea, Jun. 2023.
- [7] S. Park, H. Park, and Y. Lee, "User-centered interfaces applicable to interactive media content", *Proc. of 2024 KIIT Summer Conference*, Jeju, South Korea, Jun. 2024.
- [8] D. Wexler, J. Yip, K. P. Lee, X. Li, and Y. H. Wong, "A Touch on Musical Innovation: Exploring Wearables and Their Impact on New Interfaces for Musical Expression", *Sensors*, Vol. 24, No. 1, Dec. 2023. <https://doi.org/10.3389/fcomp.2023.1158476>.
- [9] K. Tian, "Construction of Virtual Piano Performance System Based on Visual Gesture Recognition, *Mobile Information Systems*, Vol. 2022, Jan. 2022. <https://doi.org/10.1155/2022/9446905>.
- [10] O. K. Kwon, "A Study on the Development of Alternative Instruments Corresponding to MIDI 2.0: Focusing on Augmented Instruments", *Doctoral Dissertation*, Sangmyung University, Korea, 2021.
- [11] M. Lim and N. Kotsani, "An Accessible, Browser-Based Gestural Controller for Web Audio, MIDI, and Open Sound Control", *Computer Music Journal*, Vol. 47, No. 3, pp. 6-8, Sep. 2023. https://doi.org/10.1162/COMJ_a_00693.

저자소개

권 주 연 (Ju-Yeon Kwon)



2021년 3월 ~ 현재 :
 덕성여자대학교 소프트웨어전공
 학사과정
 관심분야 : 빅데이터, 인공지능,
 머신러닝, 데이터 분석

김 나 연 (Na-Yun Kim)



2021년 3월 ~ 현재 :
 덕성여자대학교 소프트웨어전공
 학사과정
 관심분야 : 서버 프로그래밍,
 모바일 프로그래밍, 빅데이터

박 효 빈 (Hyo-Bin Park)



2021년 3월 ~ 현재 :
덕성여자대학교 소프트웨어전공
학사과정
관심분야 : 웹 프로그래밍, 모바일
프로그래밍, 빅데이터, 인공지능

장 지 인 (Ji-In Chang)



2021년 3월 ~ 현재 :
덕성여자대학교 소프트웨어전공
학사과정
관심분야 : 서버 프로그래밍,
모바일 프로그래밍, 빅데이터,
인공지능

김 유 나 (Yu-Na Kim)



2021년 3월 ~ 현재 :
덕성여자대학교 소프트웨어전공
학사과정
관심분야 : 서버 프로그래밍,
모바일 프로그래밍, 빅데이터,
인공지능

강 지 현 (Jiheon Kang)



2007년 2월 : 강원대학교
전자통신공학(공학사)
2013년 8월 : 고려대학교
전기전자공학(공학석사)
2019년 2월 : 고려대학교
전기전자공학(공학박사)

2019년 3월 ~ 2021년 2월 : SK텔레콤 ICT기술센터
매니저

2021년 3월 ~ 현재 : 덕성여자대학교 소프트웨어전공
조교수

관심분야 : 지능형 사물인터넷, 인공지능 모데 경량화 및
최적화