RESEARCH ARTICLE | NOVEMBER 28 2023

Rhythm game for real instruments

Hoang Tuan Son ; Michael Cohen

AIP Conf. Proc. 2909, 070002 (2023)
https://doi.org/10.1063/5.0182588

Rhythm game for real instruments

Hoang Tuan Son[a)] and Michael Cohen[b)]

Spatial Media Group, University of Aizu
Aizu-Wakamatsu, Fukushima 965-8580
Japan

[a)]Corresponding author: s1262009@u-aizu.ac.jp
[b)]xilehence@gmail.com

Abstract. We describe a project to create a rhythm video game in which players can use real instruments to play with music streamed by the game, capturing sound through the microphone available in mobile devices or computers, and using machine learning to analyze and detect chords played by the real instrument. The game also uses visual, auditory, and tactile elements to guide and support instrumental study.
Keywords: rhythm game, video game, musical instruments, machine learning, multimodality.

INTRODUCTION

A survey [1] of thousands of adolescents considered why they listen to and perform music. Over 50% of respondents played an instrument currently had done so or regularly before giving up, and the sample listened to music an average of 2.45 hours per day. Music is important to adolescents, and that this is because it allows them to (a) identify with a style or genre and also express an 'image' to the outside world, and (b) satisfy emotional needs.

Another study [2] confirmed that machine learning (ML) has seen increased usage in manufacturing over the past 20 years. A branch of artificial intelligence that focuses on self-learning by computers, ML is being actively used in a wide range of fields, including entertainment, business, and industrial contexts, and the video game industry is no exception. Deep reinforcement learning (DRL) is revolutionizing the field of artificial intelligence (AI) and represents a step toward building autonomous systems with a higher-level understanding of the sensed world. Currently, deep learning is enabling reinforcement learning (RL) to scale to problems that were previously intractable, such as learning to play video games directly from examined pixels [3].

From the above points, we can see that music and video games play an integral part in people's lives, especially teenagers'. We can also use music as a way to get rid of the harmful points that video games bring. In a previous project, "Multimodal Metronome" [4], we created an application that uses auditory, visual, and tactile elements to assist and guide musical instrument learners, using features available on the Nintendo Switch device and its paired "Joy-Con" detachable interface. However benefits of that project are limited to supporting musical instrument players, and do not bring an experience like a real video game. With this background, we created this project to combine the benefits of both video games and music, delivering the excitement of video games as well as encouraging the combined skills and physical movement that playing an instrument involves.

METHODOLOGY

System features

The architecture of our system is shown in Fig. 1. Our application features these interface modalities:

1. Control (input)

   Including audio capture through microphone.

2. Display (synesthetic output)

   Including auditory (sound), visual (visual music), and haptic/tactile (vibration) feedback.

In this project we want players to be able to play this video game anywhere. Therefore we have put it on highly portable devices such as smartphones, tablets, laptops. Devices must to have a microphone to capture the musical signal from the instrument.

ETLTC-ICETM2023 International Conference Proceedings
AIP Conf. Proc. 2909, 070002-1–070002-6; https://doi.org/10.1063/5.0182588
Published by AIP Publishing. 978-0-7354-4750-9/$30.00

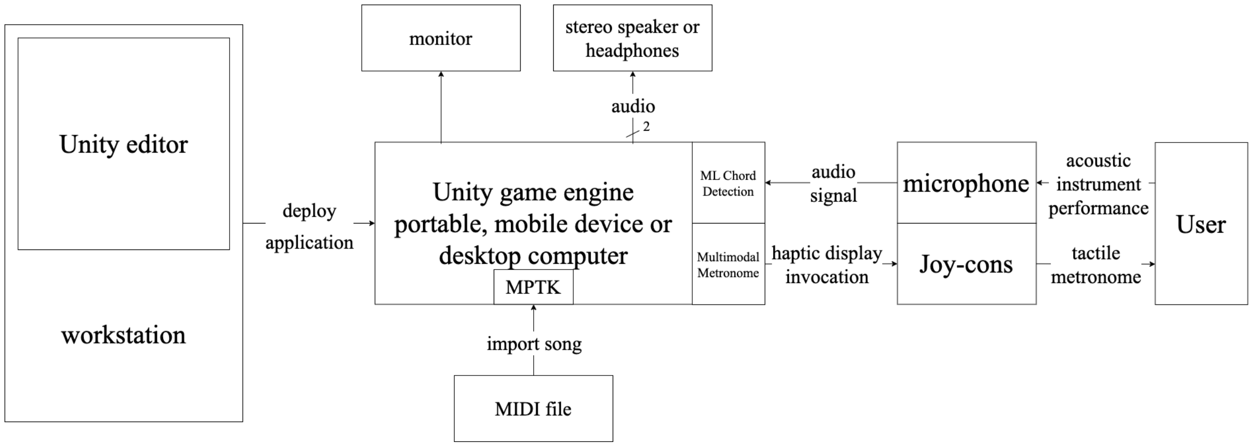

FIGURE 1. Software Functional Diagram

This project inherits all the features developed in the Multimodal Metronome project. Players use songs they like through the MIDI file import feature, using the visual and auditory metronome system. In addition to some handheld game consoles that support separate handles such as the Nintendo Switch, we provide tactile metronome features that are also inherited from the Multimodal Metronome project as well as a motion tracking system to confirm that the right rhythm is being played.

The primary goal of the project is to give users a truly seamless experience. When a player imports a MIDI file of a song they like, the system must render the MIDI file as an audio source and provide the metronome the tempo information. From there the game analyzes the sound source to give the correct chords for that song, the chords are displayed on the screen in real-time so that the player can accompany the song. During performance, the device's microphone captures sound from the instrument, analyzing its current notes to confirm if the chord matches the harmony of the song. For each correct chord, the system adds points to the score. Players can change the tempo of the music to suit their level and play style, and the metronome system automatically changes with that tempo.

Implementation

Unity, developed by Unity Technologies, is a comprehensive game and simulation development tool that allows developers to create interactive content such as video games, architectural visualization, simulation, and real-time animation. It features a fully integrated professional game engine for modeling physical interaction and rendering graphics and sound. Unity supports only C# language programming scripting, so we use C# as the programming language.

MIDI Player Tool Kit (MPTK) [5] is an asset for Unity that supports reading, playing, and rendering MIDI files, using MIDI messages and streams to trigger and modulate synchronous events. We use it to parse song data and render such events as synthesized music and event triggers. Since it's really hard to get most of the music on the market, we came up with an idea for players to actively import their favorite tracks onto the software to play with them. Doing the above brings two main advantages to the software, the first is that players have full control to choose music without being constrained by a given playlist of the game. The second benefit is that the software is lighter, reducing download and installation times, especially for low-capacity or profile devices.

To realize the above idea, we used the MPTK package, which is both a music player from MIDI files and a tool to read and return data from that MIDI file. After the player imports a MIDI file into the game, the system adds the file to the playlist and MPTK analyzes the MIDI file and sends data about the chords and tempo to the system. When the player plays that song, MPTK is responsible for reading the MIDI file and rendering it into an audio signal to output to the speakers. The system uses the chord data to make comparisons with the captured performance, using the tempo data to parameterize the metronome to the player.

The Unity Machine Learning Agents Toolkit (ML-Agents) [6] is an open-source project that enables games and simulations to serve as environments for training intelligent agents using deep reinforcement learning and imitation learning. We use it to analyze and recognize chords. Another project [7] created using ML-Agents to detect chords in continuous audio.

Chord analysis and detection

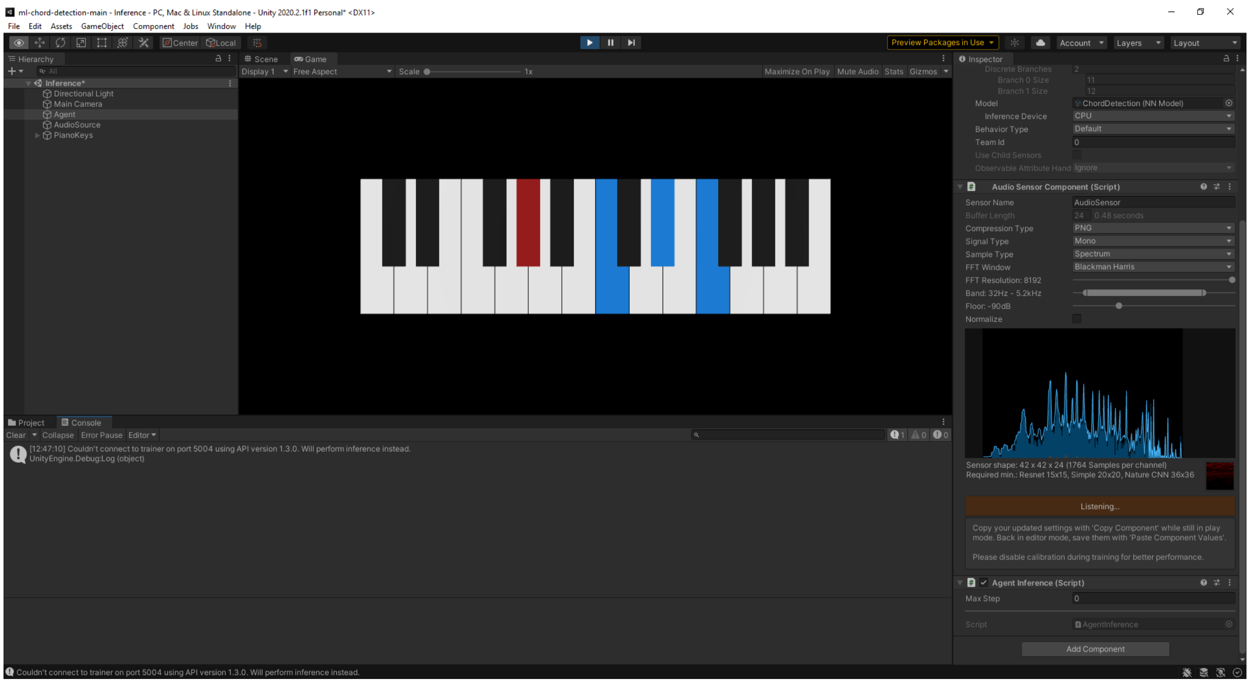

FIGURE 2. Chord detection learning progress

In Chord Detection [7], an ML-Agent is trained to identify 11 different chord types and their constituent notes. Types, keys, inversions, and transpositions are randomized. Training audio is generated on the fly with a couple of sample instruments, which is intended to generalize for various types of sounds. A bass note in the chord's key is added 50% of the time, as are random drum loops. During training, the agent observes 0.48 s slices of audio, which is a compromise between training performance and reaction speed. Apparently, the agent has difficulties learning chord types with shorter observations. Fig. 2 shows the learning progress by chord type.

The visualization interface shows the cumulative agent guesses as key colors, updated at 0.02 s intervals. Generally, detection seems to work best for clean, unambiguous signals. The more complex the music, the harder it becomes for the agent to determine the particular chord (Fig.3).

With experiences from the Multimodal Metronome project on spectrum data analysis, instead of analyzing and detecting an existing audio file, the system can analyze and detect chords from an audio signal captured through the microphone in real-time. The input source can be switched between microphone or audio capture interface, and the system automatically recognizes the source to collect data from in real-time. That means players can use any instrument they want to participate in this game.

Scoring system

Similar to rhythm games on the market, we use a point-per-play system to increase competitiveness in this video game. The software evaluates performance based on the time difference from when each chord appears in the song until the player starts playing that chord on the instrument. The shorter that time period, the higher the score, and if the time expires or the player plays the wrong chord, they are not awarded points [8].

There is a notion in rhythm games called combo or streak, which refers to the number of times a player plays correctly over a sustained period of time. In this game we also use a combo counting system to reward excellent players. A player's score is accumulated by the number of points for each correct play multiplied by the number of consecutive correct plays before, but if a player plays a chord wrong then that combo multiplier is reset.

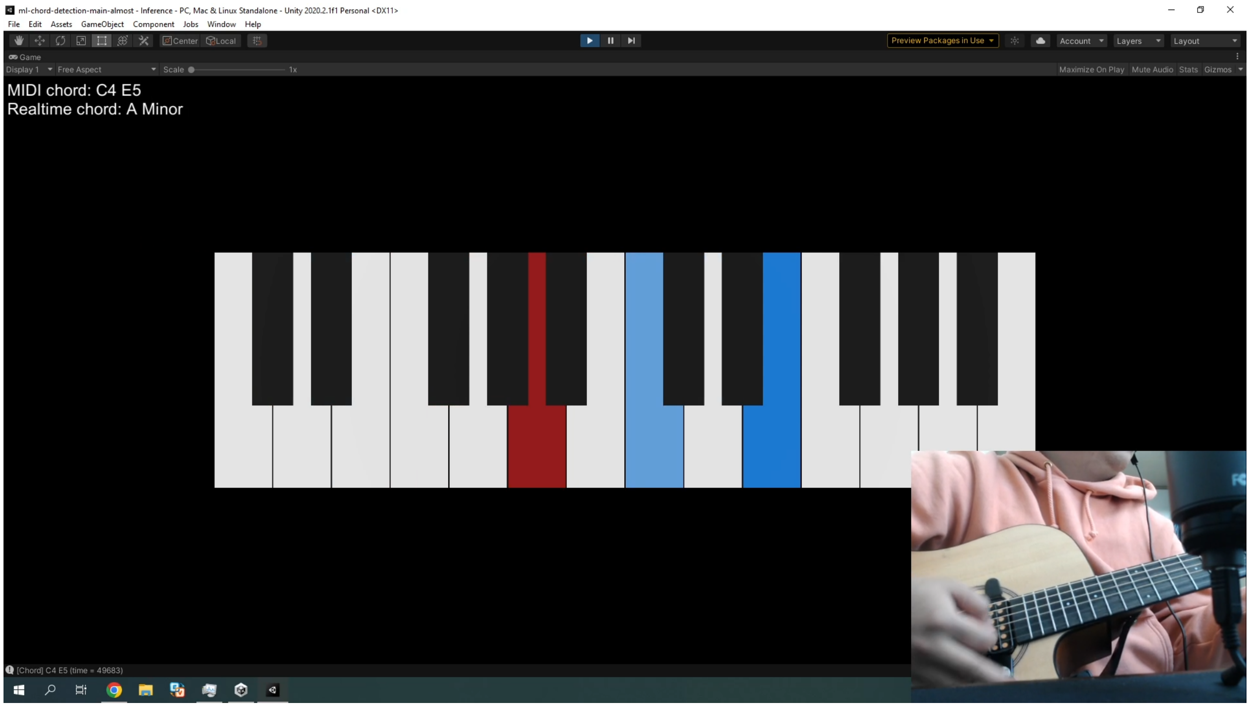

FIGURE 3. Real-time chord detection and analysis system.

Summarizing, player's score is calculated as (Fig. 4):

$$points \ += accuracy \times (1 + \frac{Combo}{25}) \times 100\% \tag{1}$$

RESULT

Having validated obtained, it can be seen that this project brings very positive results. Using a variety of musical instruments such as keyboards, guitars, ukuleles bring about high accuracy but not absolute sensitivity. Some of the results returned when testing in a high-noise environment yield low accuracy, and the system continues to analyze the sound even though the player is not playing, which consumes resources. The lack of optimization of the algorithm leads to a large latency in chord analysis and playback. Finally, the new system only detects and predicts chords, but does not distinguish which step the chord is in.

FUTURE WORK

One of the prerequisites to do in the software improvement process is to improve the data analysis algorithm to improve the accuracy of the results. Finally, we hope bring improvements to the interface and update some features for the software.

Multimodal Metronome

An application that uses auditory, visual, and tactile elements to assist and guide musical instrument learners, using features available on the Nintendo Switch device and its paired Joy-Con detachable interface, to increase the novelty as

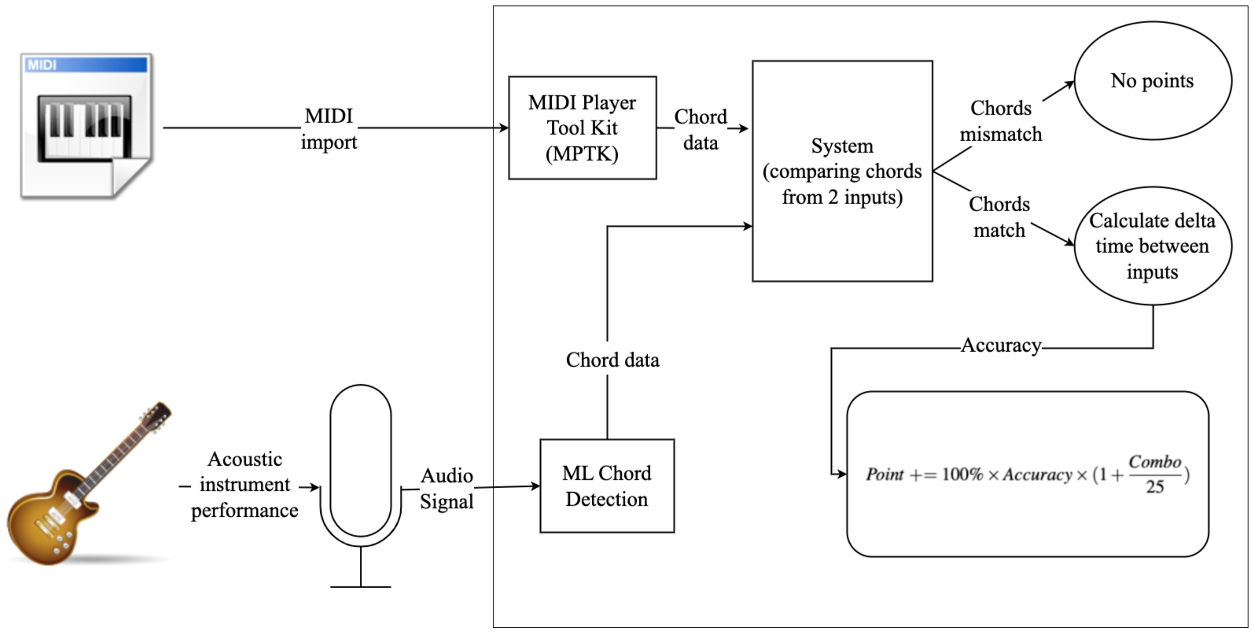

FIGURE 4. Chord detection & scoring method diagram

well as intuitiveness of musical performance while stimulating user experience (Fig. 5). The audio data into spectrum data using MPTK which can be used in audio visualization.

By applying the features present in the multimodal metronome along with some algorithmic improvements to the Rhythm game for real instruments project, players can use the metronome features through sound and images. In addition, players who use the Joy-cons controllers present on Nintendo Switch devices can also use the tactile metronome feature.

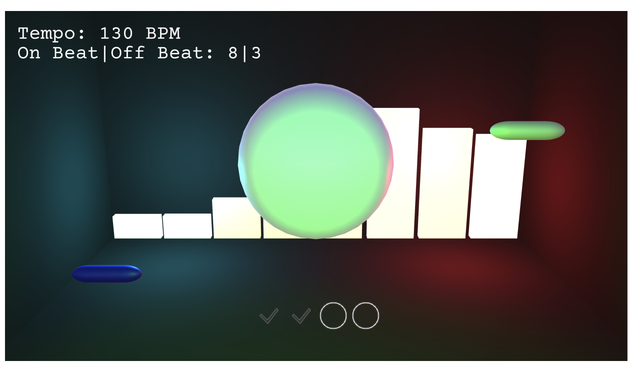

FIGURE 5. Multimodal Metronome

REFERENCES

1. A. C. North, D. J. Hargreaves, and S. A. O'Neill, "The importance of music to adolescents," British J. of Educational Psychology **70**, 255–272 (2000).
2. M. Sharp, R. Ak, and T. Hedberg Jr, "A survey of the advancing use and development of machine learning in smart manufacturing," Manufacturing Systems **48**, 170–179 (2018).
3. K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," IEEE Signal Processing Magazine **34**, 26–38 (2017).

4. T. S. Hoang and M. Cohen, “Multimodal Metronome — Rhythm game for musical instruments,” in *SHS Web of Conf.: ETLTC Int. Conf. on ICT Integration in Technical Education*, Vol. 139, edited by Debopriyo Roy (Aizu-Wakamatsu, Japan, 2022) p. 03024.
5. Bachmann, “Maestro–MIDI Player Tool Kit,” `https://paxstellar.fr/` (2022).
6. Unity, “ML-Agents,” `https://github.com/Unity-Technologies/ml-agents` (2022).
7. mbaske, “Chord detection with unity ml-agents (audio sensor),” `https://github.com/mbaske/ml-chord-detection` (2020).
8. J. Jaime, I. Barbancho, C. Urdiales, L. J. Tardón, and A. M. Barbancho, “A new multiformat rhythm game for music tutoring,” Multimedia Tools and Applications **75**, 4349–4362 (2016).
9. I. Granic, A. Lobel, and R. C. Engels, “The benefits of playing video games,” American Psychologist **69**, 66 (2014).
10. S.-I. Chiu, J.-Z. Lee, and D.-H. Huang, “Video game addiction in children and teenagers in taiwan,” CyberPsychology & Behavior **7**, 571–581 (2004).