Journal of Computing and Communication Vol.1 , No.2 , PP. 69-79 , 2022

Note Frequency Recognition and Finger Motion Capture of a Guitarist: A Survey and Challenges Ahead

John Emad*[a], Ahmed Serag[a], Karim Khaled[a], Ahmed Yehia [a], Karim Mohamed [a], Hager Sobeah[a], and Walaa Hassan[a]

[a] Department of Computer Science, Misr International University, Cairo, Egypt

*Corresponding Author: John Emad [*john1810636@miuegypt.edu.eg*]

ARTICLE DATA

Article history:
Received 03 June 2022
Revised 09 Aug 2022
Accepted 09 Aug 2022
Available online

Keywords:
Frequency Recognition
Finger Motion
Guitarist
Machine Learning

ABSTRACT

One of the main issues that face new guitar aspirants is that there is a lot of further information, which proves a lot to take in for just a beginner as it causes much-disliked confusion. Students also face problems with their left-hand motion and correct pitch frequency.
Researchers have tackled this problem in many ways and have mainly landed on using two approaches. The first was Finger Motion Capture, where previous research specialized in analyzing images and videos of the guitarist. The second was Note Frequency Recognition, where the research's primary purpose was to examine the sound and audio recorded by the guitarist.
This paper surveys all the approaches taken to solve this problem, discussing them in detail and exploring the challenges faced with each approach. Furthermore, this paper proposes a hybrid solution that includes both Note Frequency Recognition and Finger Motion Capture to make a full assessment and give feedback on a guitarist's performance.

1. Introduction

Musical instruments have been around for over 37000 years [1]. Since the dawn of time, man has created musical instruments for amusement. Through time, music has also proved itself a representation of its culture. A nation's music would give you a decent insight into that country's culture [2]. In our present day, the guitar has become a widespread musical instrument. The history of the guitar as we know it extends back to the 15th century [3] and has gone through many transformations until reaching the state that the guitar has today. The guitar has two main variations, Acoustic and Electric. First, the electric guitar is solid-bodied. It produces sound by vibrating the guitar strings and converting them into electrical signals, which run into an amplifier providing a massive music interval [4]. Our paper focuses on the more generic and common hollow-bodied Acoustic guitar. A guitarist playing acoustic guitar would use both hands, the left hand for adjusting and selecting the note and the right hand for producing the note and controlling the pitch of the produced note [5]. Nowadays, many different guitar playing styles exist, two of which are considered the two main styles when it comes to playing guitar, Lead and Rhythm guitarists.

Lead guitar is a style that mainly focuses on playing melodies and solos, which is a lot harder than rhythm guitar, a style that focuses more on playing chords [6]. With all these different styles and variations, a guitar aspirant who is just starting would be stumped by this massive amount of new information. Statistics showed that almost 90 percent of new guitarists quit in the first year [7]. As the CEO of Fender, Andy Mooney has stated that the problem isn't in attracting new aspiring guitarists; it is to keep the new guitarists interested in continuing the learning process [8]. This problem was the subject of interest for many researchers. Most took it upon themselves to develop a solution to facilitate the guitar learning process and ease the confusion beginners experience when first playing. The researchers then discovered that the main issue was that the left-hand motion was never made correctly, and the pressure needed

on the fretboard to produce the correct pitch frequency must be considered. Numerous papers used different approaches, and each faced its challenge. The general idea was one of two.

The first was through Finger Motion Capture [9], where the researchers created a system that would capture the finger motion of the guitarist through videos or images, extract those positions, and assess if they were right or wrong. The second was Note Frequency Recognition [10], where the researchers created a system that extracted note frequencies from recorded audios based on many equations to get the exact frequency and then assess if it is the right note or not. The purpose of this paper is to discuss the many approaches the researchers took in detail and the challenges they faced in creating their systems depending on their applied approach. Moreover, this paper proposes a solution using Finger Motion Capture and Note Frequency Recognition to assess a guitarist's overall performance. The rest of the paper is organized as follows: Section 2 presents the different approaches used for Note Frequency Recognition and Finger Motion Capture and the challenges faced by these approaches. Section 3 presents the proposed methodology for analyzing hand movement. Finally, conclusions are drawn in section 4.

2. Approaches Applied and Challenges

In this section, we discuss various approaches for Finger Motion Capture and Note Frequency Recognition in detail, along with a discussion of each applied approach's challenges. The section also explores the Different methodologies were used, such as external hardware and the different Deep Learning techniques used.

2.1. External Hardware

Many studies have settled on using external devices such as cameras, sensors, and wrist-worn gloves. These Studies whose main interest was finger recognition and capturing finger motion proved the effectiveness of using such devices.

2.1.1. Cameras

One of the earliest approaches to tackle this problem was the camera alone, allowing researchers to capture finger motion from images and videos through video processing and image segmentation. Currently, systems use cameras and other additional techniques to improve motion capture, as discussed below.

Burns and Anne Marie [11] suggested a method using image segmentation and Hough transformation. The main limitation of the proposed solution was that the camera used only had the vision of the first five frets. Furthermore, in some rare cases, a problem exists when playing chords when two fingers are on the same alignment of the fretboard. The proposed method was correctly able to identify both chords and note sequences.

Chutisant and Saito [12] classified the chord playing using only a stereo camera. Their proposed method involved image segmentation, and they then applied a Bayesian classifier to detect the chords. This method returned good accuracy but required the usage of colored finger markers because the skin tone of the fingers overlapped with the color of the fretboard, which decreased the accuracy.

Alfonso Perez [9] proposed this paper to analyze the finger-string interaction and to help facilitate the difficulty of capturing the guitar player's right-hand motion through high-speed cameras.

This approach posed a challenge for the researchers to detect finger motion with good accuracy since the cameras needed suitable lighting to capture a decent shot. Some papers have even placed the cameras almost on top of the fretboard, which allowed a clearer image of the higher strings but decreased clarity in the lower strings. Thus, lower accuracy resulted [11].

2.1.2 Kinect

Among the many devices used was the Kinect. The Kinect was unique in motion capture because it was mainly used when the researchers needed a solution that would work in real-time and to give real-time instructions and performance evaluations.

Chen et al. [13] proposed a method to enable the user to retrieve the learning exemplar of the target action movement interactively and to immediately acquire motion instructions while learning in front of the Kinect. It uses the Kinect depth camera and time warping algorithm. The system proved to be effective and efficient during the learning exercise movements.

Yang et al. [14] proposed a way to evaluate the user's performance in real-time. This study used Kinect v2 to acquire a professional user's motion as a reference motion and a trainee's motion as a query motion; by utilizing the time warping algorithm, the system then compares the pre-recorded reference motion to the query motion in real-time.

When using this approach, researchers faced challenges such as the only motions recognized by the Kinect where the whole body is visible to the Kinect's sensor, which is considered a limitation to their system [14].

2.1.3 Wrist-Worn Gloves

Other papers focused on correct hand motion and strumming motion capture resorted to using wrist-worn gloves equipped with inertial motion sensors to facilitate capturing the right-hand finger motion.

In an approach proposed by Bruce and Susie Howard [15], lightweight gloves were used, essentially, a device is worn underneath the wrist that casts a light beam onto the finger detecting the finger movement. The lightweight offered more flexibility for the user to move their hands. However, the lightweight gloves included two or more wearable devices: the base system and the host system. The infrared communication between the base system and the glove limited the users' finger range of motion.

Matsushita et al. [16] Soichiro Matsushita investigated a wrist-worn device to recognize the performance of the heavy metal guitar player's right hand. The player would wear this wrist glove equipped with an inertial motion sensor device to capture the right-hand motion during a warm-up procedure.

Mitobe [17] proposed a solution to track hand movement using a Cyber glove (Immersion). The tests ran on pianists playing classical songs for Mozart and Beethoven. After digitizing the hand movements into positions, the data format was specifically BioVision Action data file (BVA), an old format used to describe hand movement in animations. The BVA format was used as an input to be tested in motion capture software. The results showed that the animation was similar to the pianist playing while being recorded with a video camera. Guitarists' left-hand movements are a bit more complex than a pianist's as there are more finger variations, and the left hand doesn't move only horizontally but also vertically.

Shioji et al. [18] proposed a way to establish hand motion recognition and personal authentication at the same time. The study suggested using wrist Electromyography (EMG) and CNN to simultaneously make personal authentication and hand motion recognition.

Yoshida et al. [19] Kai Yoshida and Soichiro Matsushita designed a motion visualization system for guitar players to find their improvement points on the chord strumming technique; the angular velocity signals of the wrist motion proved helpful in recognizing the features of the strumming action. The main challenge faced by researchers when working with this approach was that although the wrist-worn glove was lightweight in most cases, it still denied the guitarists from reaching their full range of pointing motion while wearing the wrist-worn glove [15].

2.2. Neural Network and Deep Learning

In this day and age, deep learning has offered many motion capture technologies, the most commonly used of which is Neural Network, with Convolutional Neural Network (CNN) being the most employed type of neural network in motion recognition, especially in working with images.

Abesser et al. [20] proposed an approach to help automatically estimate the player's technique, instrument level, and score level. This study uses the Support Vector Machine (SVM) algorithm to evaluate the player's technique, achieving high accuracy.

Anusorn et al. [21] proposed a system for the researchers to track hand movements to identify Thai sign language. This paper differs from other sign language classifiers as it used a framework called MediaPipe. Their approach provided an accuracy greater than 90 percent, producing results close to the traditional classifiers for sign language. This paper is highly relevant as we will use a similar methodology to detect the fingers across the guitar fretboard.

Das et al. [22] proposed using the spectrograms of audio clips of different genres and feeding them into the network to determine the genre from which that audio clip was taken out. The study shows the usage of both CNN and VGG16; however, the study also showed that VGG16 was more accurate.

Enkhbat et al. [23] created an efficient framework that recognizes hand typing motions and gestures for making a virtual keyboard using a single RGB (Red, Green, Blue) camera. The paper uses CNN to recognize the typing motion and scored highly accurate results.

Yoshitaka and Ochi. [24] proposed a study to focus on minimizing left-hand movement so that new guitarists could avoid unnecessary moves when practicing scales or chords. The researchers tried using the Kinect's skeleton function to recognize the position of the left hand on the guitar. This attempt failed as the Kinect confused the human skeleton with the guitar neck when a guitarist was standing. As the Kinect's function failed, they trained their own CNN to recognize each finger position on the fretboard. Table 1 explores the accuracy of this approach for each of the four fingers on the left hand. The recognition worked only for the 7th,8th, and 9th frets.

Table1: Accuracy of Finger to String Detection

Fingering Fingers	Accuracy
Index Finger	95
Middle Finger	94
Ring Finger	96
Pinky Finger	95

This method could be used in guitar training after being fully utilized for both single-note picking and chords. However, this method only extracts the finger positions relative to the strings and frets. It does not provide a trained model to recognize whether or not the guitarist minimized the left-hand movement. That required a larger dataset to recognize all of the finger positions.

Liu et al. [25] proposed a pitch detecting model for the transcription of a polyphonic piano to perform in the frame and note-based metrics. The study showed the usage of both CNN and note onset/offset detection in pitch/frequency detecting to achieve the highest accuracy possible.

Lloyd et al. [26] Erik Lloyd and Ning Jiang proposed a way to investigate how the CNN will perform in classifying sEMG signals that have not been trained with Novel gestures. This study used various datasets for each of the four fingers and obtained high accuracy results. Still, the algorithm is not yet suitable for prosthetic control and requires more investigation.

Mohamad et al. [27] designed a technique to estimate the plucking point and magnetic pickup location along an electric guitar's strings from an isolated guitar tone recording. The system could also estimate the plucking points with varying genres and techniques of playing.

Menusha Munasinghe [28] proposed a design for a system that can recognize hand gestures and movements in front of a webcam in real-time. The study used a feedforward neural network to achieve this system design. The results, however, differed depending on the lighting in the room where the webcam is situated.

Using a slightly different methodology for chord classification Takumi, Tran [29] proposed a method using neural networks and machine learning to recognize chords played despite background noise. The attempt is to build a system that can realize the finger patterns of guitar players in video and can automatically generate the corresponding chord played. It did establish a musical score generation system from images of a guitarist playing videos. A Three-chord classifier can achieve a recognition rate of approximately 90 percent, and a five-chord classifier can reach approximately 70 percent. The five-chord classifier was relatively low according to the recognition rate.

Shibata et al. [30] described a statistical method for estimating musical scores for lead, bass, and rhythm guitars from polyphonic audio signals of typical band-style music, using CNN; the system was tested on different datasets and came up with high accuracy results.

Jungpil et al. [31] proposed recognizing the position of the human hand and detecting the joints of the hand's finger in a 2D image. The study utilized CNN to analyze the 2D image and determine the position of the joints and the depth information.

Sumarno et al. [32] investigated the influence of sampling frequency that does not follow the Shannon sampling theorem in a chord recognition system with a transform domain approach that uses a DST and segment averaging.

Tono et al. [33] proposed to deal with the automatic segmentation of the guitars appearing in RGB images to understand multimedia concerts' contents better. This paper uses CNN between two classes, guitar, and non-guitar, to detect the image pixels which belong to a guitar region. The results indicated that HRNet-based Networks tend to perform better in guitar-based segmentations.

Zhao Wang and Jun Ohya [34] proposed a way to interpret how well a guitarist played and give it an evaluation. This study uses CNN, SVR, and Discrete Cosine Transform (DCT). The system is trained to segment finger motion on guitar. The system would evaluate the player based on videos of pro guitarists.

Xie et al. [35] presented a simple, highly modularized network architecture for image classification. That network is constructed by repeating a building block aggregating a set of transformations with the same topology. This study uses neural networks to perform the mentioned task.

Zhang et al. proposed [36] a method to capture a hand gesture by capturing frames and analyzing an RGB image, and an optical flow snapshot would be fed into a CNN to learn long-term features. The output would be fed again into Long Short-Term Memory (LSTM) network, where a final classification result is predicted.

The main challenge faced by researchers when using this approach was that despite how advanced this approach is compared to others, working with multiple varying data may cause a significant decrease in the recognition accuracy [29].

2.3. Other

Aoki et al. [37] suggested photo reflectors dependent on casting light rays onto a guitar's fretboard. The method worked but failed to capture movements when different light sources were around.

Cournat et al. [38] propose encoding for tablatures in which the hand positions on the fretboard are encoded to ease computational analysis and generation of guitar music. The study shows unusual guitarist gestures around the G/B string pair due to its interval of a major third compared to the perfect fourth between other adjacent pairs of strings.

Gan et al. [39] proposed a methodology to help with visual sound separation tasks, using a Video-Analysis network and Audio-Visual separation network. This study explicitly allowed the modeling of musicians' body and finger movements as they perform.

Qiao et al. [40] used leap motion technology and tracked the movements of the hands to create a real-time virtual functioning piano. The system came up with a virtual piano that functions normally. Still, the real-time feature caused problems mainly because the player could not locate their virtual hands because their real hands got in the way.

Raboanary et al. [41] motivation were that bass guitar scores contain only tablature and have no fingerings. This study used two dynamic programming methods; one gives an optimal fingering, which is easier physically, and the other depends on the guitarist's preference.

Wortman et al. [42] Kevin Wortman and Nicholas Smith proposed an application to consider a dynamic approach to the problem of generating playable guitar chord fingerings. This app should be able to work on android mobile phones. The paper is concerned with generating chords given run-time parameters regarding the guitar configuration and the player's hand.

Marky et al. [43] created a system that can act as an assisting tool for guitarists. The system uses capacitive sensing to capture the finger motion on a 3D-printed guitar fretboard. The results showed that the visual indicators used to take little time to adjust to the finger position of the player with high accuracies.

Erdem et al. [44] explored and expanded on the ongoing experiment of what is known as 'air instruments', which means they are digital instruments that do not require physical interaction between player and instrument, the dataset used included many bioelectric muscle signals, along with motion and audio capture and trained it with LSTM. The system resulted in it being able to predict some audio energy features.

Geelen et al. [45] developed their own multi-camera motion capture system to capture the human body's movements in 3D. The system is composed of multiple microcomputers coupled with cameras and one extra microcomputer which acts as a controller. Using tools such as DeepLabCut and AniPose allowed video frames to convert and provide biochemical analysis. The system was only used in small movements but can be extended to larger ones.

Table 2 summarizes the most important papers which used the previously mentioned approaches to solve their problems and came up with the best results.

Table 2: Summarization of previous approaches

Year	Method	Results
2021 [21]	MediaPipe	A model that recognizes the Thai sign language
2020 [19]	Wrist-Worn Gloves	The combination of graphical displays allowed us to diagnose the strumming action effectively.
2020 [39]	Video Analysis Network	Able to perform better on audio-visual source separation of different instruments, with remarkable results on separating sounds of same instruments, which was impossible before.
2020 [31]	CNN	It showed a 3% error of the normalized 3D CNN model
2019 [14]	Kinect v2	Detects the difference between Kinect motions

2019 [26]	CNN	The accuracy of both (LSTM and GRU) and (RNN and FFNN) are similar while in training the (LSTM and GRU) have shorter time.
2018 [34]	CNN and SVR	A guitar fingering assessing model
2018 [24]	CNN	The accuracy of the string fingers system reached higher than 90%
2006 [11]	Image Segmentation	Detection of fingertips by matching its positions on fretboard.

3. Proposed Methodology

Based on the previous discussion of the approaches used and their limitations in section 2, we propose the following system, ignoring the usage of Kinect since it only detected the hand's palm and not the finger's positions. In addition to confusing the guitar neck with the left hand. Instead, Mediapipe [21] was used to locate each finger position and its dipping correctly. Results are shown in Figures 1 and 2.

Figure 1: hand and shoulders tracking when using MediaPipe

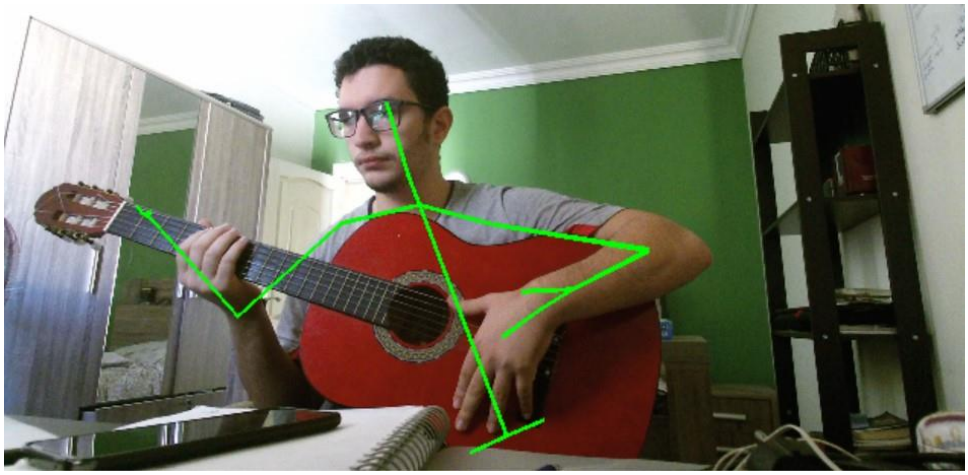

Figure 2: False detection using Kinect

The proposed system comprises six modules; two modules to handle images' data, two modules for note recognition, and two for assessing and displaying the feedback to the guitarist, as described in Figure 3. Each entity of the proposed method is discussed in detail.

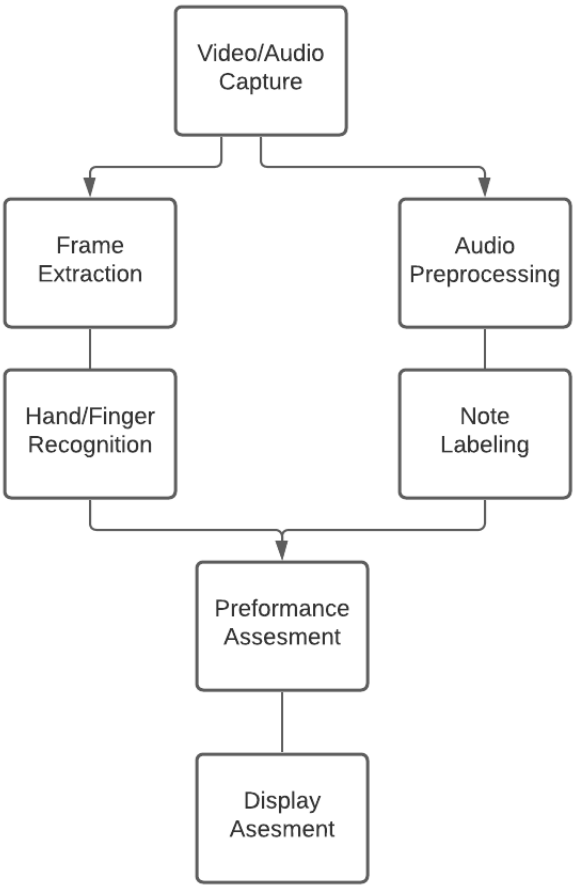

Figure 3: Proposed Method

3.1. Image Modules

3.1.1 Frame Extraction

To Extract the frames of a video, the proposed system treats every frame in the video as a separate image detecting the hands in each frame. There used to resize the video of the guitarist's performance [46], such as cartesian resizing, which involves deleting rows and columns from an image while maintaining the quality of the content.

3.1.2 Hand Finger Recognition

For finger detection, our system uses media pipe [47], a real-time hand/finger tracking solution that offers a variety of APIS such as hand tracking, pose and object detection. The system's main API uses hand tracking [48], which includes the blaze palm detector and the hand landmark model. The palm detector runs first on the whole image detecting the whole palm and making a bounding box over the hand. The hand landmark model then locates each finger joint of the hand, giving each one a landmark ID identical to Simon's hand key-point detection [49]. After extracting the hand/finger positions, the system outputs it into a CSV file with the corresponding video id and frame number in the video. After running the API on the image, the output is a set of landmarks in the image's X, Y, Z proportions, a flag indicating whether the hand is right or left, and a probability of the hand's presence in the image. To get the essential rows extracted from the repetitive guitarist's performance frames, the proposed method suggests using FDTW [50] fast dynamic time warping, RNN Recurrent neural networks, LSTM long short-term memory, and transformers to detect the repeated patterns in a time series.

3.2. Sound Modeling

3.2.1 Audio Processing

The proposed system first converts video files to Wav format to extract audio. Then, using an audio processing library "librosa" [51], The system extracts the onsets, which are the start times of each note played. The system then splits the wave file into different sub-wav files for each note played in the video. A function is then run that classifies the note played for each wav file using DFT and HPS.

3.2.2 Note Labeling

A note is a labeled frequency, so the system needs to detect the frequency of the sound in space and then assign the corresponding note. Most instrument tuning applications rely on discrete Fourier transformation (DFT), which generates a magnitude spectrum that describes the composition of frequencies that recognize the signal by the machine. We used a refined version called the harmonic product spectrum (HPS) [52], which takes several power spectrums generated by the signal and the product of the frequencies. Using HPS yielded better results as it reduced background noise and provided more accurate note labeling.

3.3. Performance Assessment

Performance assessment uses all the extracted finger positions and classifies whether or not the movement the guitarist made was correct. The main challenges in the assessment phase are that human hands have different sizes and different types of guitars. Thus, to make correct classifications, the dataset required needs to contain various exercises in different settings with different people playing.

4. Conclusion

A guitar instrument is an ancient tool with different genres, styles, and variations. These variations get overwhelming for guitarists, especially new ones who do not know the right way to start playing the guitar. Researchers have been enhancing sound recognition and hand/finger motion detection. This paper surveys the different methods used for finger detection and note labeling with the motivation to help guitarists. Furthermore, this paper proposes a hybrid approach combining audio processing and note labeling with finger motion detection using Mediapipe, which is expected to yield higher accuracy in guitar learning systems. Using Mediapipe in hand/finger motion detection is expected to improve the technique for guitarists as their hand movements will be recorded, not only the notes they played.

References

[1] Frid, E. (2019). Accessible digital musical instruments—a review of musical interfaces in inclusive music practice. Multimodal Technologies and Interaction, 3(3), 57.

[2] Rhiannon Rosas. How music and culture work together: Science behind music. Music House School of Music, 2020.

[3] Guitar history: How the guitar has evolved. College of Contemporary Music, 2018.

[4] MasterClass Staff. Guitar 101: What is an electric guitar? plus tips for perfecting your electric guitar techniques. Master Class, 2021.

[5] Ed Peczeck. Left and right hands. Classical Guitar Academy, 2015.

[6] Is lead guitar harder than rhythm guitar? Guitar Gear Finder, 2020.

[7] Josh Constine. Fender goes digital so you don't have to quit guitar. techcrunch, 2015.

[8] Perez-Carrillo, A. (2019). Finger-string interaction analysis in guitar playing with optical motion capture. Frontiers in Computer Science, 1, 8.

[9] Perez-Carrillo, A. (2019). Finger-string interaction analysis in guitar playing with optical motion capture. Frontiers in Computer Science, 1, 8.

[10] Jay K. Patel and E.S.Gopi. Musical notes identification using digital signal processing. National Institute of Technology, India, 2015.

[11] Burns, A. M., & Wanderley, M. M. (2006, June). Visual methods for the retrieval of guitarist fingering. In Proceedings of the 2006 conference on New Interfaces for Musical Expression (pp. 196-199).

[12] Kerdvibulvech, C., & Saito, H. (2007). Real-time guitar chord recognition system using stereo cameras for supporting guitarists. Transactions on Electrical Engineering, Electronics, and Communications (ECTI), 5(2), 147-157.

[13] Hu, M. C., Chen, C. W., Cheng, W. H., Chang, C. H., Lai, J. H., & Wu, J. L. (2014). Real-time human movement retrieval and assessment with kinect sensor. IEEE transactions on cybernetics, 45(4), 742-753.

[14] Yang, C. K., & Tondowidjojo, R. (2019, October). Kinect v2 Based Real-Time Motion Comparison with Re-targeting and Color Code Feedback. In 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE) (pp. 1053-1054). IEEE.
[15] Howard, B., & Howard, S. (2001, October). Lightglove: Wrist-worn virtual typing and pointing. In Proceedings Fifth International Symposium on Wearable Computers (pp. 172-173). IEEE.
[16] Matsushita, S. (2019, October). Wrist-worn Motion Diagnosis Device for Heavy Metal Guitarists. In 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE) (pp. 651-652). IEEE.
[17] Kashiwagi, Y., & Ochi, Y. (2018, March). A Study of Left Fingering Detection Using CNN for Guitar Learning. In 2018 International Conference on Intelligent Autonomous Systems (ICoIAS) (pp. 14-17). IEEE.
[18] Shioji, R., Ito, S. I., Ito, M., & Fukumi, M. (2018, November). Personal authentication and hand motion recognition based on wrist EMG analysis by a convolutional neural network. In 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS) (pp. 184-188). IEEE.
[19] Yoshida, K., & Matsushita, S. (2020, October). Visualizing Strumming Action of Electric Guitar with Wrist-worn Inertial Motion Sensors. In 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE) (pp. 739-742). IEEE.
[20] Abeßer, J., & Schuller, G. (2017). Instrument-centered music transcription of solo bass guitar recordings. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(9), 1741-1750.
[21] Chaikaew, A., Somkuan, K., & Yuyen, T. (2021, March). Thai sign language recognition: an application of deep neural network. In 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering (pp. 128-131). IEEE.
[22] Das, P. P., & Acharjee, A. (2019, December). Double coated vgg16 architecture: An enhanced approach for genre classification of spectrographic representation of musical pieces. In 2019 22nd International Conference on Computer and Information Technology (ICCIT) (pp. 1-5). IEEE.
[23] Enkhbat, A., Shih, T. K., Thaipisutikul, T., Hakim, N. L., & Aditya, W. (2020, October). Handkey: An efficient hand typing recognition using cnn for virtual keyboard. In 2020-5th International Conference on Information Technology (InCIT) (pp. 315-319). IEEE.
[24] Kashiwagi, Y., & Ochi, Y. (2018, March). A Study of Left Fingering Detection Using CNN for Guitar Learning. In 2018 International Conference on Intelligent Autonomous Systems (ICoIAS) (pp. 14-17). IEEE.
[25] Liu, S., Guo, L., & Wiggins, G. A. (2018, April). A parallel fusion approach to piano music transcription based on convolutional neural network. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 391-395). IEEE.
[26] Lloyd, E., & Jiang, N. (2019, October). Convolution Neural Network for EMG-Based Finger Gesture Classification for Novel and Trained Gestures. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 3724-3728). IEEE.
[27] Mohamad, Z., Dixon, S., & Harte, C. (2017, March). Pickup position and plucking point estimation on an electric guitar. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 651-655). IEEE.
[28] Singh, D. K. (2021). 3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling. Procedia Computer Science, 189, 76-83.
[29] Ooaku, T., Linh, T. D., Arai, M., Maekawa, T., & Mizutani, K. (2018, November). Guitar chord recognition based on finger patterns with deep learning. In Proceedings of the 4th International Conference on Communication and Information Processing (pp. 54-57).
[30] Shibata, K., Nishikimi, R., Fukayama, S., Goto, M., Nakamura, E., Itoyama, K., & Yoshii, K. (2019, May). Joint transcription of lead, bass, and rhythm guitars based on a factorial hidden semi-Markov model. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 236-240). IEEE.
[31] Shin, J., Rahim, M. A., Yuichi, O., & Tomioka, Y. (2020, August). Deep Learning-based Hand Pose Estimation from 2D Image. In 2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII) (pp. 108-110). IEEE.
[32] Sumarno, L. (2019, March). The Influence of Sampling Frequency on Guitar Chord Recognition using DST Based Segment Averaging. In 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) (pp. 65-69). IEEE.
[33] Tono, I., Gallego, J., Swiderska-Chadaj, Z., & Slater, M. (2020, September). Guitar Segmentation in RGB Images Using Convolutional Neural Networks. In 2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE) (pp. 1-4). IEEE.
[34] Wang, Z., & Ohya, J. (2018). A 3D guitar fingering assessing system based on CNN-hand pose estimation and SVR-assessment. Electronic Imaging, 2018(9), 204-1.
[35] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).
[36] Zhang, W., Wang, J., & Lan, F. (2020). Dynamic hand gesture recognition based on short-term sampling neural networks. IEEE/CAA Journal of Automatica Sinica, 8(1), 110-120.

[37] Aoki, N., Tanahashi, S., Kishimoto, E., Yasuda, S., & Iwakoshi, M. (2004). Capturing guitar fingering by photo-reflector technigue Joint Baltic-Nordic Acoustics Meeting.

[38] Cournut, J., Bigo, L., Giraud, M., Martin, N., & Régnier, D. (2021, July). What are the most used guitar positions?. In 8th International Conference on Digital Libraries for Musicology (pp. 84-92).

[39] Gan, C., Huang, D., Zhao, H., Tenenbaum, J. B., & Torralba, A. (2020). Music gesture for visual sound separation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10478-10487).

[40] Qiao, W., Wei, R., Zhao, S., Huo, D., & Li, F. (2017, August). A real-time virtual piano based on gesture capture data. In 2017 12th International Conference on Computer Science and Education (ICCSE) (pp. 740-743). IEEE.

[41] Toky Hajatiana Raboanary, Fanaja Harianja Randriamahenintsoa, Heriniaina Andry Raboanary, Tantely Mahefatiana Raboanary, and Julien Amédée Raboanary. Finding optimal bass guitar fingerings. In 2017 IEEE AFRICON, pages 65–71. IEEE, 2017

[42] Wortman, K. A., & Smith, N. (2021, January). CombinoChord: A Guitar Chord Generator App. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0785-0789). IEEE.

[43] Marky, K., Weiß, A., Matviienko, A., Brandherm, F., Wolf, S., Schmitz, M., ... & Kosch, T. (2021, May). Let's Frets! Assisting Guitar Students During Practice via Capacitive Sensing. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-12).

[44] Erdem, C., Lan, Q., Fuhrer, J., Martin, C. P., Tørresen, J., & Jensenius, A. R. (2020). Towards Playing in the'Air': Modeling Motion-Sound Energy Relationships in Electric Guitar Performance Using Deep Neural Networks. In Proceedings of the SMC Conferences (pp. 177-184). Axea sas/SMC Network.

[45] Pagnon, D., Domalain, M., & Reveret, L. (2022). Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 2: Accuracy. Sensors, 22(7), 2712.

[46] Pagnon, D., Domalain, M., & Reveret, L. (2021). Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness. Sensors, 21(19), 6530.

[47] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.

[48] Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214.

[49] Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1145-1153).

[50] Choi, W., Cho, J., Lee, S., & Jung, Y. (2020). Fast constrained dynamic time warping for similarity measure of time series data. IEEE Access, 8, 222841-222858.

[51] Raguraman, P., Mohan, R., & Vijayan, M. (2019, March). Librosa based assessment tool for music information retrieval systems. In 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 109-114). IEEE.

[52] Sripriya, N., & Nagarajan, T. (2013, October). Pitch estimation using harmonic product spectrum derived from DCT. In 2013 IEEE International Conference of IEEE Region 10 (TENCON 2013) (pp. 1-4). IEEE.