Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Fret-Nav: A Computer Vision Tool to Aid Beginner Electric Bass Players with Fretboard Navigation

Zemyna Kulikauskaite

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science

Supervisor: Kenneth Dawson-Howe

April 2025

Declaration

I agree that this thesis was completed in line with the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at `http://www.tcd.ie/calendar`.

I have completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at `https://www.tcd.ie/library/support/plagiarism/story_html5.html`.

I agree that this thesis will not be publicly available but will be available to TCD staff and students in the University's open access institutional repository on the Trinity domain only.

Zemyna Kulikauskaite

April 18, 2025

Fret-Nav: A Computer Vision Tool to Aid Beginner Electric Bass Players with Fretboard Navigation

Zemyna Kulikauskaite, Master of Science in Computer Science

University of Dublin, Trinity College, 2025

Supervisor: Kenneth Dawson-Howe

For beginners learning to play the electric bass guitar, one of the most significant early challenges is navigating the fretboard. Traditional methods, such as printed tablature, require users to mentally translate notation into physical movements on the instrument, often leading to cognitive overload, slow progress, and discouragement. This thesis presents *Fret-Nav*, a computer vision-based learning tool designed to assist novice bassists by overlaying real-time visual guidance directly onto their instrument.

Fret-Nav processes a live video feed of the player's bass guitar, detects the fretboard and strings, and tracks hand and finger positions using the MediaPipe Hands framework. Musical input is parsed from bass tablature and synchronised with real-time finger placement and plucking detection to generate on-instrument visual cues for correct note positioning. The system enforces deliberate practice by advancing only when the correct note is detected.

The system achieved a note detection accuracy of 90.96% and precision of 94.37% on a manually annotated dataset of 157 notes. A user study comparing *Fret-Nav* to traditional tablature showed increased engagement, self-correction ability, and reduced cognitive load among participants using the tool. While the prototype performs reliably under controlled conditions, future work is required to address environmental constraints such as cluttered backgrounds and camera positioning.

By combining computer vision with beginner-focused pedagogy, *Fret-Nav* demonstrates the potential for real-time visual systems to transform early-stage music learning into a more intuitive and confidence-building experience.

Acknowledgments

Thank you to my Mother and my Godmother for their unconditional, endless encouragement throughout the last five years. This is as much yours as it is mine.

Zemyna Kulikauskaite

University of Dublin, Trinity College
April 2025

Contents

Abstract ii

Acknowledgments iii

Chapter 1 Introduction 1

1.1 Objective . . . 1
1.2 Motivation . . . 1
1.3 Executive Summary . . . 3

Chapter 2 Background 6

2.1 The Electric Bass Guitar . . . 7
2.1.1 Parts of the Bass Guitar . . . 7
2.1.2 How Notes Are Played . . . 8
2.1.3 Beginner Plucking Technique . . . 9
2.1.4 Tablature . . . 10
2.2 Related Work . . . 10
2.2.1 Cognitive Load in Learning Musical Instruments and the Challenges of Fretboard Navigation . . . 11
2.2.2 Music Information Retrieval (MIR) and Automatic Music Transcription (AMT) . . . 12
2.2.3 Vision-Based Approaches in Automatic Music Transcription (AMT) 13
2.2.4 Methods Used in Note Detection Algorithms . . . 15
2.2.5 Real-Time Feedback Systems in Music Learning . . . 16
2.2.6 The Research Gap . . . 17
2.3 Relevant Computer Vision Concepts . . . 17
2.3.1 Image Preprocessing . . . 17
2.3.2 Edge Detection . . . 18
2.3.3 Template Matching . . . 19
2.3.4 Line Detection . . . 20

2.3.5 Hand Tracking . . . 20
2.3.6 Motion Analysis . . . 21
2.4 Summary . . . 22

Chapter 3 System Design **24**
3.1 System Overview . . . 24
3.2 Core Components . . . 25
3.2.1 Visual Input Processing . . . 25
3.2.2 Musical Information Processing . . . 26
3.2.3 Guidance and Feedback . . . 26
3.3 Data Flow . . . 26
3.4 User Interaction . . . 28
3.5 Summary . . . 28

Chapter 4 Implementation **30**
4.1 Visual Input Processing . . . 30
4.2 Musical Information Processing . . . 38
4.3 Guidance and Feedback . . . 40
4.4 Summary . . . 42

Chapter 5 Evaluation **43**
5.1 Accuracy Evaluation . . . 43
5.1.1 Experimental Setup . . . 43
5.1.2 Results . . . 44
5.2 User Study . . . 45
5.2.1 Study Design . . . 46
5.2.2 Results . . . 46
5.3 Summary . . . 48

Chapter 6 Conclusion **49**
6.1 Critique and Review . . . 49
6.2 Core Contributions . . . 50
6.3 Future Work . . . 51
6.4 Final Remarks . . . 52

Bibliography **54**

Appendices **57**

List of Tables

5.1 Confusion Matrix for System Evaluation 44
5.2 User Study Likert Question Responses Summary 46

List of Figures

1.1 The Anatomy of a Bass Guitar . . . 2
1.2 An Example of the User Interface of Fret-Nav . . . 3

2.1 The Anatomy of a Bass Guitar . . . 7
2.2 The Bass Guitar Frets and Fretboard . . . 9
2.3 Position of Basic Plucking Technique . . . 9
2.4 An Example of Tablature Representation of a Single Note . . . 10
2.5 A Bass Player Plucking a High Fret . . . 11
2.6 The User Interface of Soloist . . . 13
2.7 The Guitar Tablature Notation Results from TaptoTab . . . 16
2.8 Image Rotation Applied to an Image . . . 18
2.9 Canny Edge Detection Algorithm Applied with Different Thresholds . . . 18
2.10 An Application of Template Matching to Find an Object . . . 19
2.11 An Application of Hough Line Transform to Detect Straight Lines in an Image . . . 20
2.12 MediaPipe Hands Framework Applied to Detect Hand Landmarks of a Bassist 21
2.13 Optical Flow in a Street Scene with Moving Cycles and Pedestrians . . . 22

3.1 The Fret-Nav System User Interface . . . 24
3.2 High-Level Overview of the Fret-Nav System . . . 25
3.3 Modular Architecture of *Fret-Nav* . . . 27
3.4 User Journey Flow Diagram of *Fret-Nav* . . . 29

4.1 An Image of the Original Frame from Camera Input . . . 31
4.2 An Image of the Rotated Frame . . . 31
4.3 Canny Edge Detection Applied to a Frame . . . 32
4.4 Template Edge Image of the Bass Fretboard . . . 32
4.5 Template Matching Image Results . . . 32
4.6 Hough Line Transform Applied to Frame . . . 33
4.7 Hough Line Transform Applied to Frame with Post-Processing . . . 33

4.8 Hough Line Transform Applied to Frame with Extrapolation 34
4.9 Fret Distance Calculation Results 34
4.10 Hough Line Transform Applied to Fretboard 35
4.11 Combined Fret Detection Observations 35
4.12 An Image Hand Position at a Pluck 38
4.13 An Image of a Detected Pluck 38
4.14 Sample Tablature Input for the System 39
4.15 The *Fret-Nav* User Interface with Note Tracking 39

5.1 A Frame from the Note Detection Testing Video 44

Chapter 1

Introduction

This introductory chapter will provide an overview of the domain, the specific problem addressed by this dissertation project, the motivation behind this work, and an executive summary of the document.

1.1 Objective

The primary objective of this project is to explore how computer vision can be leveraged to assist beginner musicians in learning fretted string instruments by providing real-time, visual guidance. Specifically, the project aims to develop and evaluate a system that captures live video of a player's instrument and overlays intuitive visual cues directly onto the instrument, indicating the correct finger placements for specific musical notes.

The system is intended to support learners in developing instrument familiarity, spatial awareness, and muscle memory. The ultimate goal is to create a tool that enhances the early stages of instrumental learning, making the experience more engaging and efficient for novices.

1.2 Motivation

The precision bass guitar is a four-stringed electric instrument commonly used in a wide range of musical styles, from rock and jazz to funk and pop. It plays a foundational role in most ensembles by providing low-frequency rhythm and harmony [33]. Physically, the instrument consists of a solid wooden body, a long neck, and a flat surface along the neck called the fretboard [23].

The fretboard is divided by thin metal strips known as frets, which segment each string into discrete pitches. Pressing a string down behind a fret shortens its vibrating

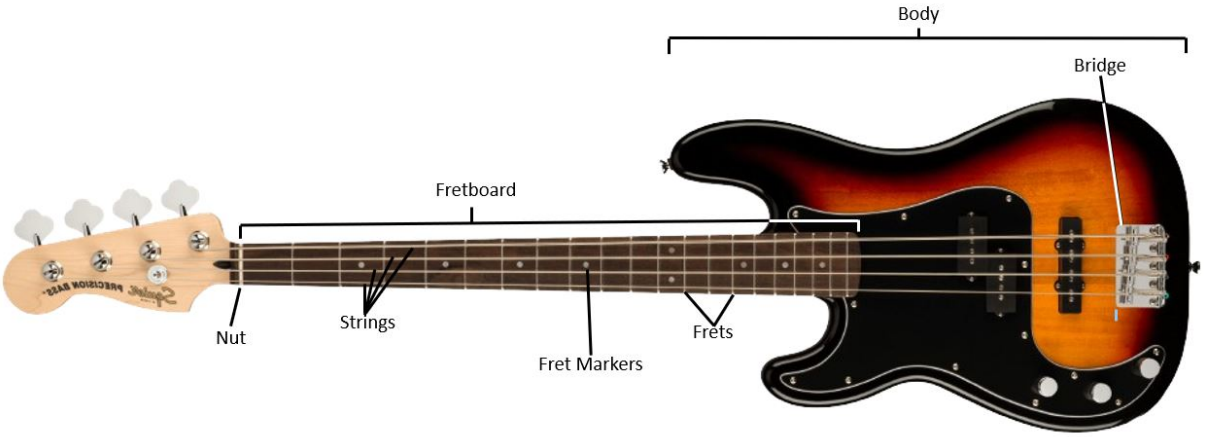

Figure 1.1: The labelled anatomy of a standard four-string precision model bass guitar. Image Credit: xmusic.

length, thereby raising the pitch of the note produced when the string is plucked. The four strings are typically tuned to **E**, **A**, **D**, and **G**, from lowest to highest, and each fret represents a semitone (half-step) increment [32].

For those just starting out, understanding the fretboard and navigating it confidently is one of the first major learning challenges to overcome. Early learners frequently rely on counting frets from the nut (the starting point of the fretboard) to find the correct position, a method that is both slow and cognitively demanding. This not only interrupts the flow of playing but can lead to frustration, discouragement, and ultimately reduced practice motivation. *"Beginners often struggle with learning the layout and note location of the bass fretboard. This fundamental and essential skill is not that easy to acquire when you are starting out."* [38].

While fret markers — small inlaid dots at certain positions — are designed to assist with navigation, they are often not intuitive for novices and require experience to use effectively [8]. The importance of practicing exercises that focus on visualisation to develop a strong mental map of the fretboard are highlighted by J. LaBrooy [29]: *"Without a good understanding of the fretboard, you'll struggle to play notes, chords, and scales, making it difficult to progress as a bassist."* Without a strong mental map of the fretboard, many beginners find themselves overwhelmed, unable to connect what they see in tablature or lessons with the physical space on their instrument.

This project is motivated by the need to reduce that cognitive barrier. By offering real-time, on-instrument visual cues for note placement, the system aims to make the learning process more tangible and less mentally taxing. Rather than guessing or counting frets, learners would be able to see where to place their fingers directly on the instrument — an

approach intended to accelerate familiarity, build confidence, and support the development of essential muscle memory.

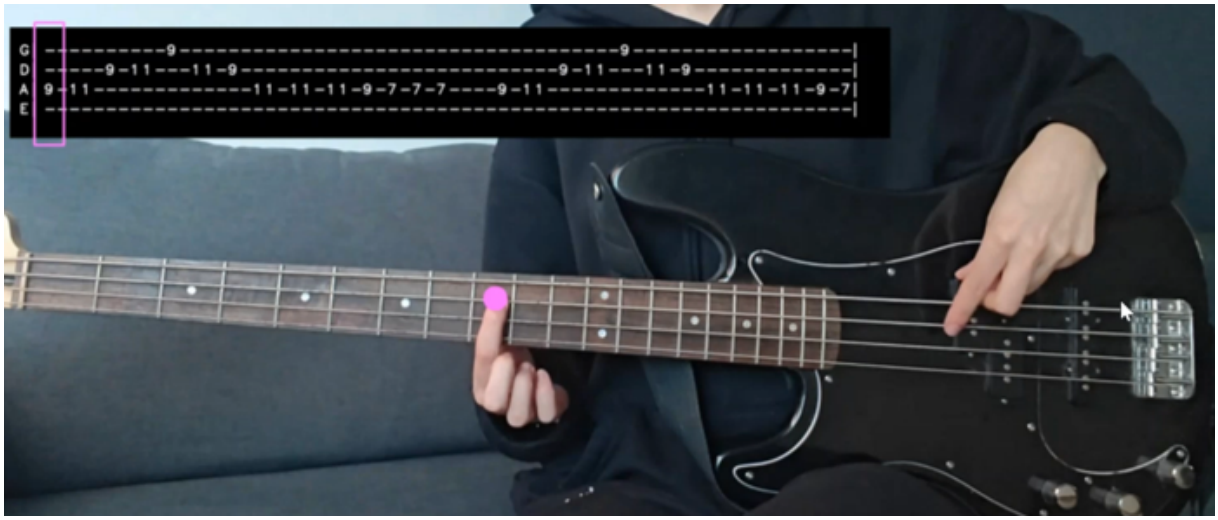

Figure 1.2: The User Interface of *Fret-Nav*, a system displaying music notation and visual markers on both the music sequence and the instrument to guide the user through a practice session.

1.3 Executive Summary

To provide a clear road map for the reader, this dissertation is structured into the following chapters, each building upon the previous one to culminate in the development and evaluation of the *Fret-Nav* system.

Chapter 2 provides a detailed review of the electric bass guitar, with an emphasis on the physical characteristics relevant to fretboard navigation. This includes a breakdown of its components, how notes are played, common beginner techniques, and the use of tablature notation. The chapter then transitions into an overview of related work, including a critical analysis of existing music learning tools and Automatic Music Transcription (AMT) systems. The shortcomings of audio-only systems for bass guitar, particularly the ambiguity caused by multiple fretboard locations producing the same pitch, are highlighted. Vision-based approaches are explored as a more promising solution, with references to recent systems using hand tracking, fretboard segmentation, and optical flow. The research gap is identified as a lack of real-time, beginner-focused tools for bass guitar, particularly those offering direct visual fretboard guidance.

In Chapter 3, the design of the proposed system — referred to as *Fret-Nav* — is introduced at a conceptual level. The chapter outlines the system's three primary components: visual input processing, musical information processing, and real-time guidance

and feedback. A modular architecture is proposed, allowing the system to receive live video of the bass guitar, track hand positions, process musical notation, and overlay visual cues indicating where to place fingers. The interaction flow is described from a user's perspective.

Chapter 4 presents the technical implementation of the system in detail. It begins by describing the visual pipeline, including camera setup, fretboard localisation via template matching and line detection, and fret position estimation using a formula based on the *12th-root-of-2*, a commonly used concept in music theory. The chapter then explains how MediaPipe Hands is integrated to track finger positions and distinguish between fretting and plucking hands. A motion-based pluck detection algorithm using fingertip trajectories and optical flow is introduced, including a special case for the **E** string, which lacks a higher adjacent string. Musical input is handled through a custom tablature parser that tracks the current note in a sequence, while visual cues are overlaid using calculated fret and string coordinates. The guidance system only advances once a correct note is played, enforcing deliberate and accurate practice.

Chapter 5 evaluates the system using both quantitative and qualitative methods. The objective performance was assessed with a dataset of 157 manually annotated bass notes, resulting in an overall note detection accuracy of 90.96% and a precision of 94.37%. Most false positives were traced to ambiguous finger movements or edge cases in pluck detection, especially on the **E** string. A user study involving six complete beginners compared the *Fret-Nav* system against traditional printed tablature. Participants using *Fret-Nav* reported greater ease of learning, improved self-correction, and higher engagement. Feedback also suggested that the direct visual guidance significantly reduced cognitive load compared to manually interpreting tablature. Suggestions for interface improvements, including enlarging the on-screen tablature and adding audio playback of notes, were also noted.

Chapter 6 provides a critical reflection on the work, identifying both strengths and limitations of the system. While the system performed reliably under controlled conditions, its current reliance on a controlled environment limits its deployment in casual practice settings. Core contributions include the creation of a real-time visual learning tool tailored to the needs of novice bassists, the implementation of a custom visual detection and note validation pipeline, and the validation of its educational impact through a small-scale user study. The final section outlines directions for future work, including audio-visual integration, improved pluck detection, UI enhancements, support for additional instruments, and generalisation to less structured environments.

This structure aims to guide the reader through a logical progression, from the initial motivation and theoretical underpinnings to the practical implementation and evaluation

of the *Fret-Nav* system.

Chapter 2

Background

This chapter will present the foundation upon which the *Fret-Nav* system is built. Its primary purpose is to equip the reader with the essential knowledge and context required to fully understand the design and development of the computer vision-based tool for beginner electric bass guitar players.

Following the principle of laying a clear path for the reader, this chapter will first introduce the fundamental concepts of the electric bass guitar and delineate basic playing technique relevant to a beginner. This initial section will ensure a common understanding of the instrument that *Fret-Nav* aims to assist with.

Building upon the understanding of the instrument itself, the chapter will then delve into the relevant academic landscape. It will present the known challenges of learning a new instrument and explain the effects it may have on beginners musicians. It will then explore the interdisciplinary field of Music Information Retrieval (MIR), highlighting its role in bridging the gap between music and technology through computational analysis.

Subsequently, the chapter will examine Automatic Music Transcription (AMT) Systems. While *Fret-Nav* takes a complementary approach by providing visual guidance based on musical notation such as tablature, understanding AMT methodologies is important as it represents a significant area of research in extracting musical information. Furthermore, many AMT systems use computer vision, providing relevant insights into the challenges and techniques involved in visually analysing guitar playing.

Finally, this chapter will present a comprehensive overview of the computer vision techniques that form the technological foundation of the *Fret-Nav* system. This will include a discussion of methods used in object detection for identifying the fretboard, hand tracking and pose estimation for monitoring finger positions, and optical flow analysis for detecting plucking actions. The understanding of the principles and capabilities of these computer vision techniques will further help to illustrate how *Fret-Nav* is able to provide

real-time visual feedback to beginner bass players.

By presenting these interconnected areas — the instrument itself, the field of MIR in music education, related work in AMT, and the underlying computer vision techniques — this chapter aims to provide a thorough and coherent background. This will not only contextualise the *Fret-Nav* project within existing research and technological capabilities but also prepare the reader for the detailed explanation of the system's methodology and implementation in the subsequent chapters. The information presented here will clearly demonstrate why the development of *Fret-Nav* is both relevant and technically feasible within the current landscape of music technology and computer vision.

2.1 The Electric Bass Guitar

This project focuses on assisting beginner musicians learning the electric bass guitar, with an emphasis on the standard precision bass model. While there are many different types of bass guitars — such as the Jazz Bass, short-scale basses, five- or six-string extended range models, and fretless basses [5] — this project narrows its scope to the widely-used four-string Precision Bass [23]. This choice allows for a consistent reference point in terms of fretboard layout, scale length, and string tuning, which is essential for developing targeted instructional tools.

2.1.1 Parts of the Bass Guitar

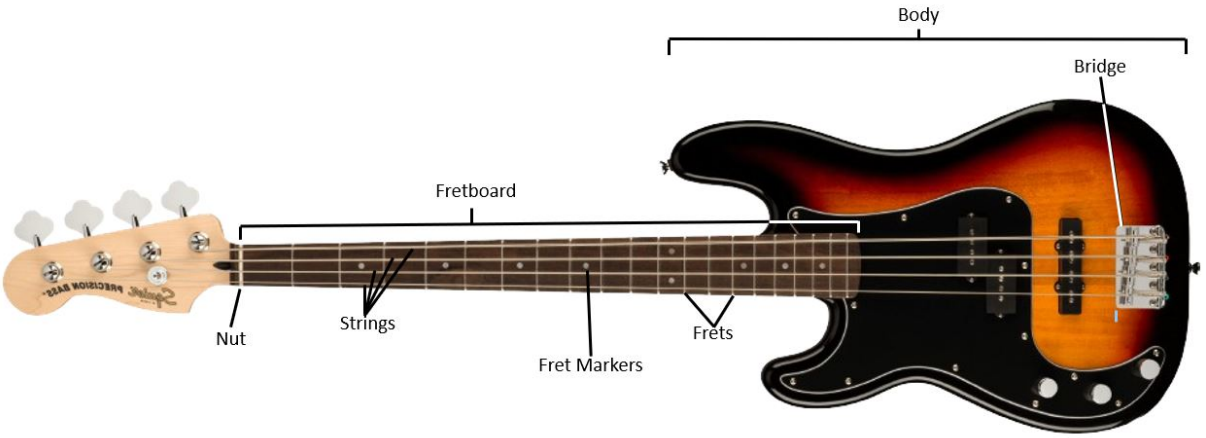

Figure 2.1: The labelled anatomy of a standard four-string precision model bass guitar. Image Credit: xmusic.

The electric bass guitar, like other guitars, consists of several key parts working to-

gether to produce sound [32]. For the purposes of this project, we will focus on the following components — visualised in figure 2.1 — relevant to fretboard navigation and note production.

- **Body:** This is the main wooden part of the bass guitar that provides a foundation for all other components.
- **Fretboard:** A long, typically wooden surface extending from the body. It is where the musician presses down on the strings to produce different notes.
- **Strings:** A standard precision bass has four strings. These are made of metal and vibrate when plucked or strummed to produce sound. The standard tuning for these four strings, from thickest (lowest pitch) to thinnest (highest pitch), is **E, A, D,** and **G**. They can also be referred to as the 4^{th}, 3^{rd}, 2^{nd}, and 1^{st} strings, respectively.
- **Nut:** This is a thick, hard piece located at the top of the fretboard. It has groove that keep the strings spaced correctly and at the proper height above the fretboard. The nut is considered the starting point of the playable length of the strings.
- **Frets:** These are thin metal strips that run perpendicular to the strings of the instrument, dividing the fretboard into semitone intervals.
- **Fret Markers:** Visual aids (often represented as dots) embedded in the fretboard surface to help players locate different fret positions. These are found on the 3^{rd}, 5^{th}, 7^{th}, 9^{th}, 12^{th}, 15^{th}, 17^{th}, and 19^{th} frets.
- **Bridge:** Located on the body of the bass guitar, the bridge anchors the strings at the opposite end from the nut. It plays a role in adjusting the string length and height. The distance between the nut and the bridge determines the **scale length** of the instrument, which is used to calculate fret positions.

2.1.2 How Notes Are Played

Notes on the bass guitar are produced by pressing a string down against the fretboard behind a specific fret with the fingers of the fretting hand — typically the left hand for right-handed players — and then plucking or striking the string with the plucking hand — typically the right hand for right-handed players [22]. The frets act as a stopping point, shortening the vibrating length of the string, which in turn raises its pitch. Each fret represents a semitone, or half step, increase in pitch [41]. For example, playing on the second fret of the **D** string involves pressing the 2^{nd} string down in the space between the first and second frets.

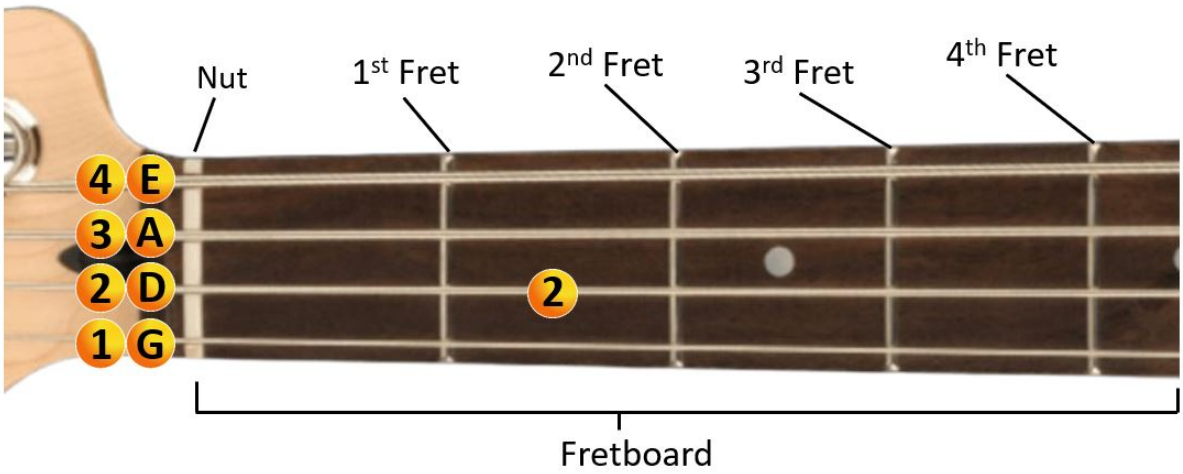

Figure 2.2: A labelled view of the top of the bass guitar fretboard. Image Credit: xmusic

2.1.3 Beginner Plucking Technique

Figure 2.3: The hand position of a bassist performing simple two-finger bass plucking technique, with the thumb anchored on the instrument and the index and middle fingers of the plucking hand resting on the instrument strings.

A common and fundamental plucking technique for beginner bass players involves using the index and middle fingers of the plucking hand while anchoring the thumb on one of the lower strings or the guitar itself. This technique allows for an alternating plucking pattern, where the index finger plucks a string on one beat, and the middle finger plucks on the next, creating a consistent rhythmic foundation [7]. Anchoring the thumb provides a point of stability for the plucking hand, helping to control the movement and ensure consistent tone production. While other plucking techniques exist, this two-finger alternating approach is the most common starting point for beginners [41]. This

project narrows its scope by focusing on this type of playing alone.

2.1.4 Tablature

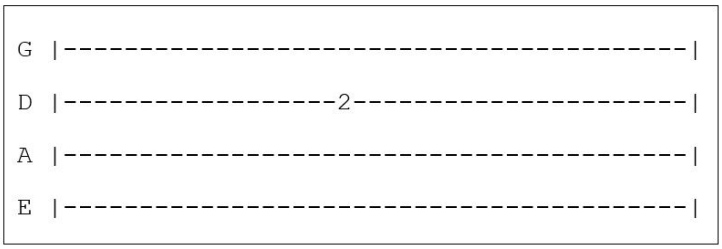

Figure 2.4: Tablature notation representing a single note on the **D** string and 2nd fret.

Tablature (or "tab") is a common and direct form of musical notation specifically designed for fretted string instruments like the bass guitar. Unlike standard musical notation which focuses on pitch and duration, tablature visually represents the instrument's fretboard [37]. It uses horizontal lines to represent the strings. Conventionally, in bass tablature, these lines are often shown with the highest pitched string (**G**) at the top and the lowest pitches string (**E**) at the bottom, which is a flipped representation compared to how the strings might appear visually on the instrument itself, where lower pitched strings are physically higher [6].

Numbers written on these lines indicate which fret to press on that specific string. A **0** indicates an **open string** — playing the string without pressing down any frets. For example, in figure 2.4 , the simple bass tablature indicates that the player should pluck the **D** string while pressing down on the 2nd fret. The horizontal dashes indicate that no note is played on those strings at that particular time. Tablature provide a more direct way for guitarists and bassists to understand finger placement compared to standard notation, as it explicitly hows the string and fret.

For beginner bass players, tablature is often the first written form of music they encounter due to its intuitive representation of the fretboard [12]. This project aims to leverage tablature as a basis for providing visual guidance on the actual instrument.

2.2 Related Work

This section delves into the existing body of knowledge relevant to the *Fret-Nav* project. It is structured to provide a comprehensive examination of the current state of research in Music Information Retrieval (MIR) and Automatic Music Transcription (AMT), partic-

ularly concerning their application to string instruments such as the guitar and the bass guitar.

2.2.1 Cognitive Load in Learning Musical Instruments and the Challenges of Fretboard Navigation

Learning a musical instrument, particularly for beginners, often presents a significant cognitive challenge due to the multitude of tasks that must be managed simultaneously [11]. For guitar or bass playing, novice musicians must coordinate finger placements on the fretboard, maintain rhythmic timing, and memorise the spatial relationships between notes and fret positions [6]. These concurrent demands can lead to cognitive overload, a state where the learner's working memory capacity is exceeded, hindering the learning process and potentially leading to frustration and discouragement [26].

One of the key aspects contributing to this high cognitive load for beginner string instrument players is the challenge of fretboard navigation and memorisation. Unlike some other instruments where notes are visually and spatially discrete, e.g., piano keys, the guitar and bass fretboard presents a continuous surface divided by frets, with multiple locations for the same note [19]. As beginners lack the ingrained muscle memory for fretboard positions, they often resort to the tedious process of fret counting to locate desired notes, which slows down their playing and diverts cognitive resources away from other crucial aspects like creativity and musicality [39]. For instance, if a tablature indicates an '11' on any given string, a beginner may need to count each fret from the nut — 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 — before playing the note. This constant counting makes it difficult to achieve any sense of fluidity in their practice.

Figure 2.5: A bassist plucking a note on a high fret.

Visual real-time learning tools for string instruments have the potential to significantly reduce cognitive load by directly supporting spatial memory through visual cues [34]. This direct visual guidance can help establish a stronger association between musical notation and their corresponding physical locations on the instrument. To understand how visual tools might assist beginners, it's helpful to first examine the current technological landscape in Music Information Retrieval (MIR) and Automatic Music Transcription (AMT).

2.2.2 Music Information Retrieval (MIR) and Automatic Music Transcription (AMT)

Music Information Retrieval (MIR) is an interdisciplinary field dedicated to the extraction, analysis, and retrieval of meaningful information from music using computational techniques [10]. It bridges the gap between music and technology by combining principles from computer science, musicology, and signal processing. The goal of MIR is to enable efficient and effective ways to interact with and manage musical data. This includes tasks such as music classification, recommendation, similarity analysis, and, crucially for this project, music transcription and analysis for learning purposes.

Automatic Music Transcription (AMT) is a significant task within MIR, concerned with automatically transcribing musical performances into a symbolic representation, such as sheet music [27]. While traditional AMT primarily focuses on audio analysis, there is a growing body of research exploring the use of computer vision techniques for AMT, particularly for string instruments [9]. These vision-based approaches analyse video recordings of musicians playing to extract information about their hand movements, finger positions, and even vibrations of the strings to infer the musical notes being played.

Music Information Retrieval (MIR) and Automatic Music Transcription (AMT) technologies have significantly impacted music education and practice by providing innovative tools that enhance learning experiences and facilitate skill development [18], [2]. AMT technologies specifically contribute to music pedagogy by automating the transcription process, making music more accessible to learners who may not be proficient in reading traditional notation. Tools that generate guitar tablature from audio or video inputs provide practical learning aids for guitarists and bassists. These systems can transcribe performances into tablature notation, which is particularly helpful for beginners unfamiliar with standard musical scores [28]. By bridging the gap between performance and notation, AMT tools facilitate self-directed learning and enable musicians to learn new pieces more efficiently.

For instance, Wang, et al. developed *Soloist* [40], a system that transforms exist-

Figure 2.6: The user interface of Soloist. Image Credit: Wang, et al.

ing guitar instructional videos into interactive tutorials using audio processing. *Soloist* addresses the lack of personalised feedback and inefficient navigation commonly associated with traditional instructional videos. By analysing recordings of a user's practice and comparing them with reference performances, Soloist provides real-time feedback through melody visualisation, highlighting any calculated deviations in pitch and timing. This immediate feedback helps learners identify and correct mistakes quickly, leading to improved practice sessions.

Similarly, Avila, et al. proposed an AI-based design framework aimed at supporting musicians' practices by integrating digital tools into their preparation process [4]. Based on ethnographic studies and interviews with working musicians, the authors identified challenges such as balancing instruments while operating computers and annotating scores — a physically demanding and inefficient task. Musicians often need to manage controls of online videos with one hand while playing their instrument with the other, leading to interruptions in the flow of practice. By automating tasks like real-time transcription, harmonic analysis, and automatic chord chart generation, MIR technologies can streamline practice routines, reduce physical strain, and allow musicians to focus more on their performance.

However, existing MIR and AMT tools often assume players already possess fundamental knowledge of their instrument, neglecting the crucial issue of fretboard memorisation that beginner bass guitarists commonly struggle with. Consequently, these systems do not fully address the practical needs of novice musicians.

2.2.3 Vision-Based Approaches in Automatic Music Transcription (AMT)

Audio-based methods have been the traditional approach in Automatic Music Transcription (AMT), relying solely on the analysis of sound signals to extract musical information

such as notes, rhythm, and structure. However, these methods face significant limitations, particularly when applied to string instruments like guitars and bass guitars. One of the primary challenges is the difficulty in determining the exact string and fret positions from audio data alone [25], [1].

The same note can often be played on multiple strings and at various fret positions in string instruments. This redundancy creates ambiguity in note identification when using audio-only systems. Goldstein [25] highlighted that audio-based methods struggle to identify the specific string and fret being played because different strings can produce the same pitch. The author stated, *"Audio-based methods often face difficulty when the same pitch is produced by multiple strings. In contrast, visual data makes it easier to determine which string is vibrating."* [25].

Abeßer [1] also discussed the limitations of audio-only approaches in their work on automatic string detection for bass guitar and electric guitar. They pointed out that *"estimating this position is important for generating tablature notation automatically,"* but audio recordings do not provide sufficient information to determine the exact fretboard position. This limitation is particularly problematic for generating tablature notation, which is essential for guitarists and bassists who rely on precise finger placement information for learning and performance. Additionally, bass guitars tend to produce lower-frequency sounds that are significantly quieter than other instruments, making them more difficult to capture accurately with regular microphones [30]. As a result, many notes may be missed entirely in transcription, further limiting the effectiveness of audio-only AMT.

These limitations highlight the need for visual-based solutions to overcome challenges with fretboard navigation, particularly for beginners who require clear visual feedback to internalise fret positions effectively. Thus, for beginner players, visual feedback systems directly address the critical need for precise fretboard guidance and memory reinforcement.

One of the primary advantages of vision-based methods is their ability to precisely detect which strings are being pressed and picked, thereby identifying the exact fret positions used during a performance. Asmar [3] presented a novel approach for the automatic transcription of guitar music using computer vision and depth data from RGBD cameras. Their system focuses on generating tablature by solely relying on visual input, effectively eliminating the complexities associated with audio signal processing. The author stated: *"We propose a method of transcription based only on video, as an approach using audio processing seems complex, especially when dealing with chords, frets, and strings."* [3]. By using RGB frames for fretboard segmentation and hand pose estimation, along with depth maps for enhancing finger position detection, the system can accurately determine which fingers are pressing which frets. The depth data is particularly useful in resolving

ambiguities when fingers hover over the strings without pressing them, a scenario where audio data would provide little to no information. This precise detection is instrumental for generating accurate guitar tablature and offering detailed feedback to musicians.

Despite these advancements, purely vision-based methods also face challenges, such as lighting conditions, visibility constraints, and occlusions of the fretboard or hands. However, they offer a direct and intuitive way to capture the physical actions of a musician on a string instrument, providing information crucial for accurate tablature generation that is often difficult to obtain from audio signals alone.

2.2.4 Methods Used in Note Detection Algorithms

Current techniques for segmenting the fretboard and detecting hand positions involve a combination of computer vision methods, depth sensing, and machine learning algorithms. In the work by Asmar [3], the process for detection begins with fretboard segmentation, where a deep learning model is trained to produce a binary mask of the fretboard area. This segmentation isolates the fretboard from the background, improving the accuracy of subsequent steps such as string and fret detection. Additionally, the system uses optical flow analysis to detect string plucking or picking motions, which is essential because pressing a string without plucking does not produce sound. This ensures that the system transcribes notes only when a string is actively played.

Variable lighting conditions and visibility constraints pose substantial challenges in the accuracy of visual detection systems. The effectiveness of computer vision algorithms, particularly those involving background subtraction and chromatic segmentation, is highly dependent on consistent and adequate lighting.

Schramm & Jung [35] acknowledge that their system's hand segmentation technique works well in controlled environments but might require manual adjustments based on lighting conditions or the specific camera used. The tool uses the HSV colour space and adaptive thresholds to segment the hand from the rest of the body. However, inconsistent lighting can lead to inaccurate segmentation, causing the system to misinterpret hand movements. The paper states: *"This technique works well in controlled environments but might require manual adjustments based on lighting conditions or the specific camera used."* [35].

Similarly, Emad, et al. [21] highlight that early methods using cameras often had limited accuracy due to lighting conditions and visibility constraints. The survey notes: *"Early methods using cameras often had limited accuracy due to lighting conditions and visibility constraints (e.g., capturing only a portion of the fretboard). Additionally, skin tone similarities with the guitar fretboard decreased detection accuracy."* [21].

These lighting issues can cause the visual system to struggle with accurately detecting and tracking hand and finger movements, leading to reduced performance in tasks such as finger motion capture and note identification. Emad, et al. [21] highlight the importance of considering the mitigation of lighting issues when designing these systems.

In [20], Elashmawi, et al. present a hybrid system that integrates finger motion capture with note frequency recognition to assess and improve guitarists' performance. The system employs the MediaPipe framework for hand and finger tracking, capturing detailed finger movements and placements on the fretboard. For audio analysis, the Harmonic Product Spectrum (HPS) method is used to refine note frequency recognition, achieving a high accuracy of 95%.

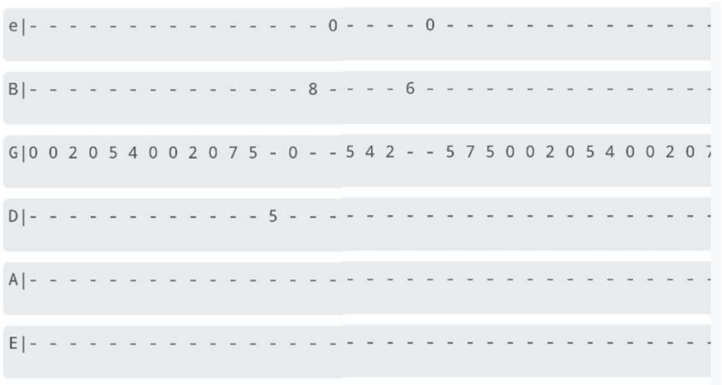

Figure 2.7: The guitar tablature notation results from *TaptoTab* Image Credit: TaptoTab

TapToTab [24] introduces a system that generates guitar tablature by combining video inputs with audio analysis. The visual component uses deep learning models like YOLOv8 to detect the fretboard and hand positions, even in challenging conditions such as occlusions or rapid note transitions. This overcomes limitations of traditional image processing techniques, which often struggle with varying lighting and angles. On the audio side, the system applies the Fourier Transform to convert time-domain signals into frequency-domain data, extracting precise note information. By synchronising the visual detection of hand positions with the derived note frequencies, *TapToTab* can generate accurate tablature transcriptions.

2.2.5 Real-Time Feedback Systems in Music Learning

Soloist [40] addresses the limitations of traditional instructional videos, which often lack personalised feedback and efficient navigation. The system provides real-time feedback by analysing the user's performance and comparing it with reference performances from instructional videos. The authors state: *"Soloist provides real-time feedback on the user's*

performance by analysing recordings of their practice and comparing it with the reference performance in the instructional video." [40]. By offering instant feedback through melody visualisation and note correctness scores, Soloist enables musicians to correct mistakes as they occur, enhancing the effectiveness of their practice sessions.

2.2.6 The Research Gap

While significant progress has been made in the development of audio-based and vision-based Automatic Music Transcription (AMT) systems, much of this research has been heavily focused on the standard six-string guitar, with comparatively little attention given to the bass guitar. The limited representation of bass guitar in current AMT and MIR research leaves a substantial gap, especially for novice players who lack fretboard familiarity and require guided, real-time feedback. The *Fret-Nav* project aims to address this underexplored area by developing a vision-based system specifically designed to support bass guitarists in learning fretboard navigation through intuitive visual cues and real-time performance feedback. Addressing this gap not only contributes to AMT research but also opens the door to more inclusive, accessible music education tools tailored for underrepresented instruments like the bass guitar.

2.3 Relevant Computer Vision Concepts

Following the literature review, this section outlines the key computer vision concepts employed in the development of *Fret-Nav*.

2.3.1 Image Preprocessing

Image preprocessing involves operations applied to images before performing further analysis. These techniques help enhance image quality or transform the data into a more usable format for computer vision tasks.

Image Rotation

Image rotation is a geometric transformation that repositions an image around a central point, typically its centre [13]. This technique is commonly used for data augmentation, aligning images, or correcting orientation. The rotation is achieved using a transformation matrix:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Figure 2.8: A gometric rotation transformation applied to an image. Image Credit: Kenneth Dawson-Howe

where θ is the angle of rotation. Libraries like OpenCV provide built-in functions such as `cv2.getRotationMatrix2D` to apply this transformation efficiently.

2.3.2 Edge Detection

Edge detection identifies points in an image where the brightness changes sharply. These points often represent object boundaries and are vital for shape analysis and segmentation.

Canny Edge Detection

Figure 2.9: The Canny Edge Detection alorithm applied at different thresholds. Image Credit: Kenneth Dawson-Howe

Canny edge detection is a multi-stage algorithm designed to detect a wide range of edges in images [14]. It involves the following steps:

1. **Noise Reduction:** The input image is first smoothed using a Gaussian filter to reduce the impact of noise on the edge detection process.

2. **Gradient Calculation:** This step produces two gradient images that are used to compute the gradient magnitude and direction at each pixel, which together highlight areas of rapid intensity change — the likely locations of edges.

3. **Non-Maximum Suppression:** To thin out the detected edges and retain only the most likely edge pixels, the algorithm applies non-maximum suppression. This process compares the gradient magnitude at each pixel with its neighbors in the gradient direction. If the pixel's magnitude is not the maximum, it is suppressed (set to zero). This results in a cleaner edge representation.

4. **Double Thresholding:** To differentiate between strong, weak, and non-edge pixels, double thresholding is applied using two user-defined thresholds: a high threshold and a low threshold. Pixels with gradient magnitudes above the high threshold are considered strong edges, while those between the low and high thresholds are classified as weak edges. Pixels below the low threshold are discarded.

5. **Edge Tracking by Hysteresis:** Weak edges are examined to determine whether they are connected to any strong edges. If a weak edge is connected to a strong one, it is retained; otherwise, it is suppressed. This process ensures that relevant edges are preserved while isolated, noise-induced weak edges are removed.

This technique is widely used due to its accuracy in detecting meaningful edges.

2.3.3 Template Matching

Template matching is a technique for finding portions of an image that match a template image [15]. It's often used for object detection when the object's shape and orientation are known in advance.

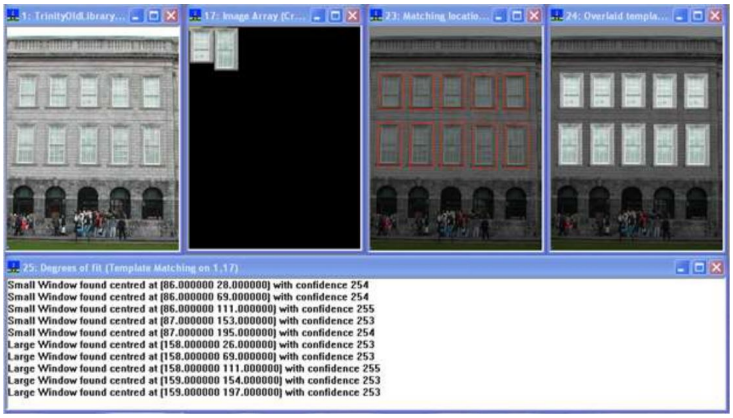

Figure 2.10: An application of template matching to find a window object within an image. Image Credit: Kenneth Dawson-Howe

The algorithm slides the template over the input image and at each location, a similarity metric is calculated between the template and the overlapping region of the image.

The location with the best score — either the highest correlation or the lowest difference — is considered the point where the template best matches the image. This position is returned as the detected location of the object.

OpenCV implements this with the `cv2.matchTemplate` function.

2.3.4 Line Detection

Line detection is used to identify straight lines in an image, often as part of scene understanding or object detection [16].

Hough Line Transform

Figure 2.11: The original image, an edge representation, and an application of Hough Line Transform to detect straight lines in an image. Image Credit: Kenneth Dawson-Howe

The Hough Line Transform detects lines by transforming points in Cartesian coordinates (x, y) into the Hough space (ρ, θ), where:

$$\rho = x \cos \theta + y \sin \theta$$

Each point in the image votes for all possible lines it could belong to in the accumulator space. Peaks in this space represent the most likely lines.

OpenCV's `cv2.HoughLines` function implements this technique, commonly used in applications such as road lane detection.

2.3.5 Hand Tracking

Hand tracking detects and follows the position and movement of hands in real time, useful in gesture recognition, AR, and human-computer interaction.

Figure 2.12: The MediaPipe framework applied to the hands of a bassist to detect finger landmarks.

MediaPipe Hands

MediaPipe Hands is a high-fidelity hand and finger tracking solution developed by Google [42]. It performs:

1. **Palm Detection:** Instead of detecting the entire hand directly, MediaPipe first detects a palm bounding box using a lightweight convolutional neural network (CNN). This approach simplifies the problem by reducing variation in hand poses and orientations.

2. **Hand Landmark Estimation:** Once the palm is detected, the system crops and normalises the region to feed into a regression model that predicts 21 3D hand landmarks. These landmarks represent key anatomical points on the hand, including fingertips, joints, and wrist. The output includes (x, y, z) coordinates relative to the image, providing a detailed pose estimation of the entire hand.

3. **Tracking:** To ensure stable and smooth hand motion over time, MediaPipe uses temporal smoothing and keypoint tracking between frames. This step reduces jitter and makes the landmark positions more reliable in real-time applications, especially during fast movement or brief occlusions.

It processes RGB input in real time and returns landmarks as normalized coordinates. MediaPipe is known for its efficiency, enabling use on mobile and embedded devices.

2.3.6 Motion Analysis

Motion analysis involves understanding movement within a sequence of frames. It's essential for activity recognition, video compression, and tracking.

Optical Flow

Optical flow estimates the motion of objects or camera between consecutive frames by analyzing pixel intensity changes [17]. It assumes:

- **Brightness Constancy:** Pixel intensity remains constant over time.
- **Small Motion:** Movement between frames is minimal.
- **Spatial Coherence:** Neighbouring pixels have similar motion.

Figure 2.13: Optical flow in a street scene with moving cycles and pedestrians. Image Credit: Arijit Kar.

OpenCV provides two main optical flow methods:

- `cv2.calcOpticalFlowPyrLK`: Computes sparse optical flow, typically tracking selected keypoints..
- `cv2.calcOpticalFlowFarneback`:Computes dense optical flow across the entire image using polynomial approximations.

The output is a vector field representing motion direction and magnitude at each point. In this project, the focus is not on precise motion estimation, but simply detecting if any movement occurs. For this purpose, `cv2.calcOpticalFlowFarneback` is preferred due to its dense output, which allows for easy thresholding to detect motion presence.

2.4 Summary

Despite considerable advances in Music Information Retrieval (MIR) and Automatic Music Transcription (AMT) technologies, a significant gap remains in addressing the unique needs of beginner bass guitarists. Most current systems prioritise standard six-string guitar and assume a baseline of musical and instrumental proficiency, offering little targeted

support for novice bass players. These learners frequently struggle with fretboard memorisation and physical coordination, yet existing tools offer minimal assistance for these foundational skills.

Audio-based AMT systems, while effective in many cases, suffer from key limitations when applied to bass guitar. The instrument's lower frequencies are harder to capture accurately, and the redundancy of pitch positions across the fretboard leads to ambiguities in transcription. These problems are compounded for beginners, who depend heavily on clear, visual guidance to understand finger placement and string interactions.

Vision-based approaches offer a promising alternative by leveraging RGB and depth data to pinpoint finger positions and string activity. These systems can generate accurate tablature and track user performance in real time. However, even state-of-the-art solutions — such as *TapToTab* and *Soloist* — focus on guitar or general music instruction, largely overlooking bass-specific challenges. While these tools integrate audio and video processing for richer feedback, they often presume that users already have a working knowledge of the fretboard.

Furthermore, real-time feedback systems are typically designed with advanced users in mind, providing corrective insights that are more valuable to intermediate or professional musicians. Beginner-focused interaction designs, such as intuitive fretboard visualisation, guided finger placement, or progressive learning scaffolds, are conspicuously lacking.

The *Fret-Nav* project is proposed as a direct response to this oversight, aiming to fill the educational and technological gap by developing a vision-based system specifically tailored to beginner bassists. By focusing on real-time fretboard feedback, finger tracking, and accessible visual cues, *Fret-Nav* addresses a critically underrepresented user group in the landscape — those just starting their journey with the bass guitar.

Chapter 3

System Design

This chapter outlines the abstract design of *Fret-Nav*, a computer vision-based tool intended to assist beginner electric bass players with fretboard navigation and playing practice. The chapter is structured into several sections, each contributing to a comprehensive understanding of the proposed solution. The System Overview section will present a high-level abstract description of the proposed approach, independent of specific implementation details. Following this, the Core Components section will present the essential functional modules of *Fret-Nav* and their abstract responsibilities. The Data Flow section will describe the abstract movement and processing of information within the system. The User Interaction section will detail how a beginner bass player would interact with *Fret-Nav*.

3.1 System Overview

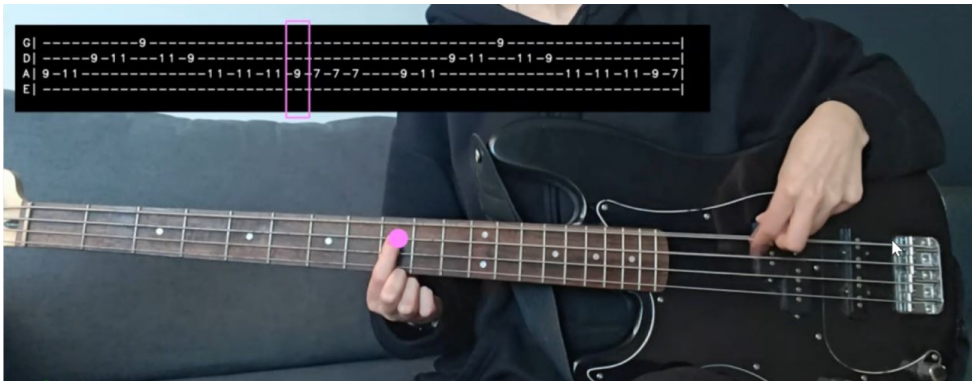

Figure 3.1: The User Interface of Fret-Nav, a system displaying music notation and visual markers on both the music sequence and the instrument to guide the user through a practice session.

Fret-Nav is conceived as a system that provides real-time visual guidance to beginner electric bass players directly overlaid on their instrument during practice. The core problem it addresses is the difficulty beginners face in memorising fret positions and locating notes on the bass guitar fretboard, which can hinder their progress and enjoyment. Existing learning methods often rely on static representations like virtual fretboards on a screen, which do not foster direct engagement with the physical instrument.

The *Fret-Nav* system proposes an alternative approach by utilising computer vision techniques to perceive the bass guitar and the player's interaction with it. It will process a live video stream of the instrument to identify the fretboard and track the player's hands. Simultaneously, *Fret-Nav* will process musical information — such as tablature — representing a piece of music to be learned. By integrating the visual information with the musical information, the system will generate real-time visual cues indicating the correct finger placement for each note in the musical piece. Furthermore, it will provide feedback on whether the played note matches the expected note, facilitating self-correction and reinforcing correct playing.

This abstract design focuses on the functional aspects of the system, allowing for flexibility in choosing specific computer vision algorithms (like segmentation, hand pose estimation, optical flow, Hough Line Transform, Canny edge detection, template matching) and methods for processing musical information. The key is the integration of visual perception of the instrument with the flow of musical instruction to create an interactive and intuitive learning experience for beginners.

3.2 Core Components

The *Fret-Nav* system can be divided into the following core components:

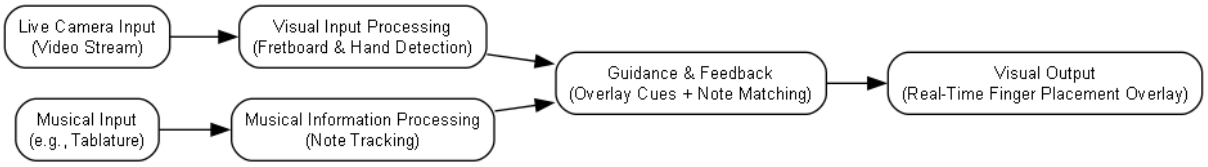

Figure 3.2: A block diagram showing the three main areas of *Fret-Nav:* Visual Input, Musical Input, and Guidance & Feedback.

3.2.1 Visual Input Processing

Responsible for capturing and interpreting visual information from the bass guitar. It includes:

- **Fretboard Localiation:** Identifying the precise location and orientation of the fretboard within the video frame.
- **Hand and Finger Tracking:** Detecting and tracking the presence and position of the player's hands and fingers relative to the fretboard.
- **String Plucking Detection:** Identifying when and where a string is being plucked or played.

3.2.2 Musical Information Processing

Handles the musical content that the user is learning, including:

- **Musical Score Input:** Interpreting notation such as tablature representing the notes to be played.
- **Current Note Tracking:** Keeping track of the next note(s) to be played in the sequence.

3.2.3 Guidance and Feedback

Provides real-time guidance and feedback to the player:

- **Visual Cue Generation:** Creating visual overlays to indicate the correct finger position for the current note.
- **Note Accuracy Assessment:** Comparing the user's played note (from visual inference) to the expected note.
- **Feedback Delivery:** Providing feedback indicating whether the note played was correct or incorrect, and when to advance.

3.3 Data Flow

The abstract data flow within the system proceeds as follows:

1. A live video feed of the bass guitar is captured and sent to the Visual Input Processing module.
2. This module localises the fretboard, tracks hands/fingers, and detects string plucks.

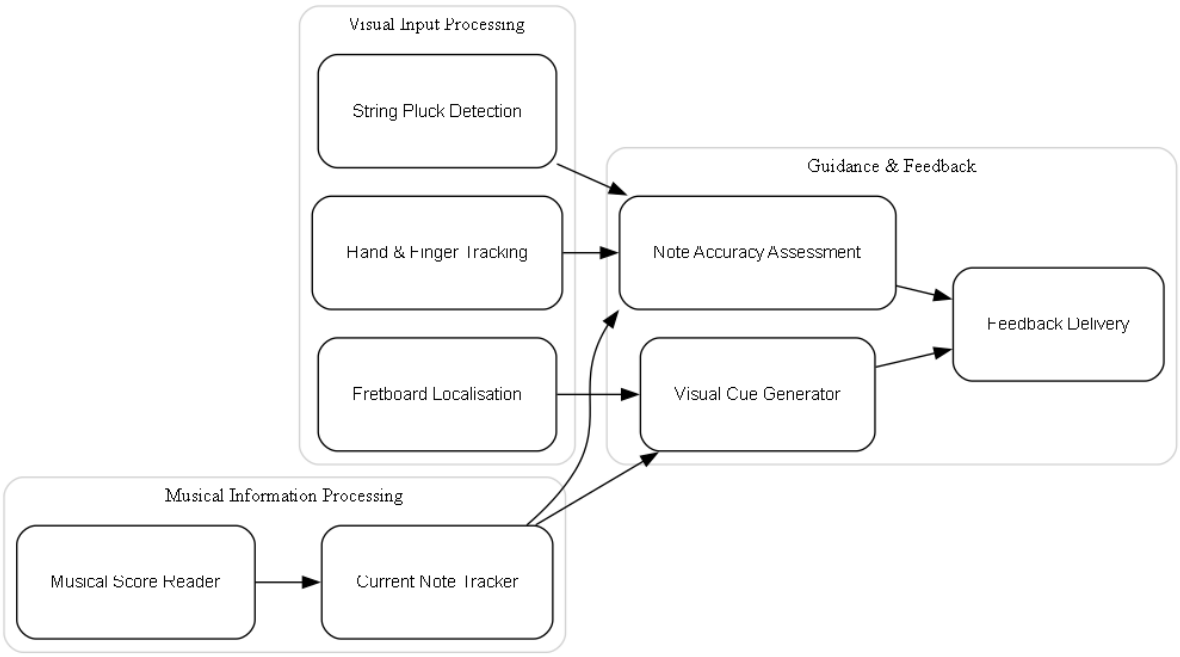

Figure 3.3: Each core component of *Fret-Nav* as a modular architecture. Each cluster has clear sub-modules, breaking down its responsibilities. Arrows represent the data dependencies and flow between components.

3. Simultaneously, musical score data (e.g., tablature) is input into the Musical Information Processing module.

4. This module determines the next expected note.

5. Guidance and Feedback receives information about both the fretboard and the expected note.

6. A visual cue is generated and overlaid onto the video feed to show the correct finger placement.

7. The user adjusts their hand and plays the note. The visual system re-evaluates hand position and pluck location.

8. Played note information is sent back to the Guidance and Feedback module.

9. The system assesses note accuracy and gives feedback. If correct, it signals progression to the next note.

10. The cycle repeats, allowing continuous real-time interaction.

3.4 User Interaction

The expected interaction flow for a user is:

1. **Setup:** Position a camera to capture the fretboard and start the *Fret-Nav* system.
2. **Musical Selection:** Choose a song or exercise, loaded into the system in a supported format (e.g., tablature).
3. **Practice:** Watch the live feed with overlaid visual markers indicating finger placement.
4. **Playing and Feedback:** Play according to the cue; the system confirms correct notes or indicates errors.
5. **Progression:** If correct, the next note cue is displayed automatically.
6. **Iteration:** Continue this guided practice to develop muscle memory and improve accuracy.

3.5 Summary

This chapter has presented an abstract system design for *Fret-Nav*, a real-time visual guidance tool for beginner bass guitar players. By combining computer vision techniques with musical knowledge processing, the system creates an interactive and engaging environment tailored for early-stage learners. The design outlines a flexible and modular architecture that can be implemented with a variety of methods, offering a promising alternative to traditional static or audio-only learning tools.

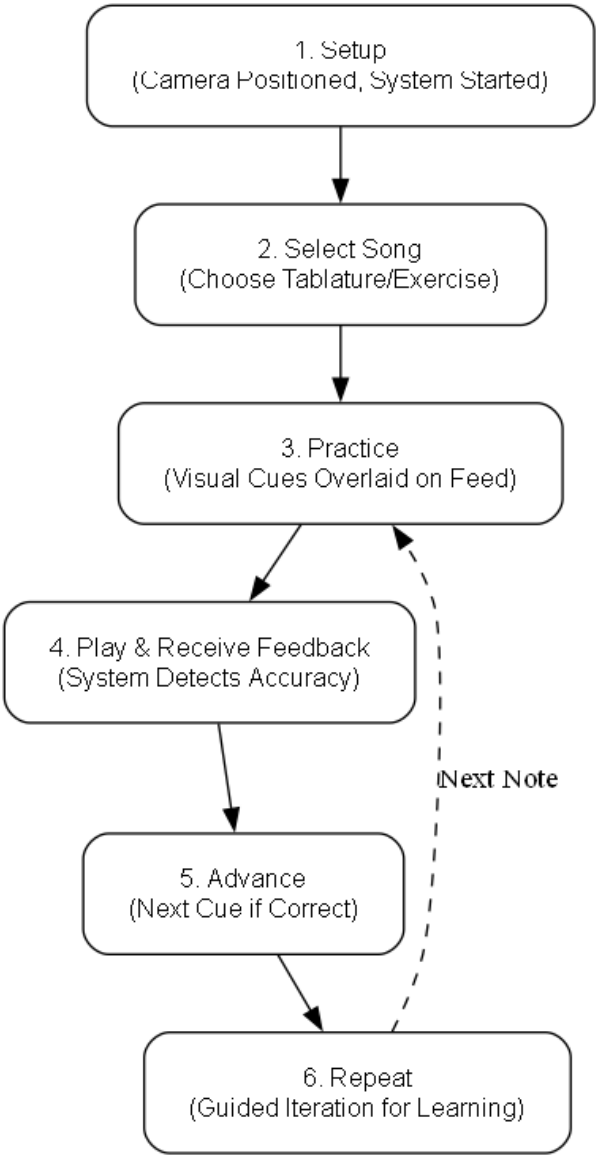

Figure 3.4: A step-by-step flow diagram showcasing the linear flow from setup to iteration of *Fret-Nav* as it would be experienced by a user. The dashed loop represent continuous guided play.

Chapter 4

Implementation

This chapter provides a detailed account of the implementation of *Fret-Nav*, a computer vision tool designed to assist beginner electric bass players with fretboard navigation and practice. Building upon the abstract design outlined in the previous chapter, this section describes the specific methods, algorithms, and technologies employed to realise the *Fret-Nav* system. For this project, OpenCV and Python were chosen to develop the system. The chapter is structured to cover the key stages of development.

4.1 Visual Input Processing

The visual input processing component of *Fret-Nav* was implemented to capture and interpret the live video stream of the bass guitar. This involved several key steps:

Camera Setup

A standard RGB camera was used to capture the video feed of the bass guitar and the player's fretting and plucking hands. The camera was positioned to provide a clear, unobstructed view of the entire scale length of the bass guitar. While depth information could be integrated in future work, the initial implementation focused solely on RGB frames.

Fretboard Detection and Localisation

Accurate fretboard detection and localisation form the foundation of the *Fret-Nav* system, as all subsequent processes - such as string detection, fret calculation, hand tracking, and note verification - depend heavily on precisely identifying the fretboard region within the video frame.

Region of Interest (ROI) Identification

The first and most critical step in visual input processing is the identification of the fretboard's Region of Interest (ROI) from the live camera feed. This is accomplished through a combination of image preprocessing and template matching techniques.

Figure 4.1: An image of the original frame from camera input, before any procesing is applied.

To reduce inconsistencies arising from camera angle variations, the system first applies an orientation correction.

Figure 4.2: An image of the frame after the rotation transformation is applied.

The live video frame is rotated to align the detected lines (strings) horizontally, ensuring a consistent basis for further processing. This rotation angle is estimated by detecting all dominant near-horizontal lines in the frame using the Hough Line Transform and computing their average slope. In a controlled environment, it is aumed the most dominant horizontal slopes detected will be the ones pertaining to the strings of the bass guitar. This creates a normalised frame where strings run near-horizontally, making downstream tasks (like edge detection and matching) significantly more reliable. Once aligned, the image is passed through a Canny edge detector to enhance contrast along structural lines and edges. The fretboard is defined by a regular pattern of straight lines — frets and strings — that often have strong contrast against the body of the guitar. However, raw image data also contains a lot of visual noise, shadows, and texture details unrelated to the structure. The edge detection focuses attention on structural edges, such as fret wires

and string boundaries, which are essential for locating the fretboard. The processed frame now presents a cleaner, high-contrast version of the image, where meaningful structural elements are easier to match against the reference template.

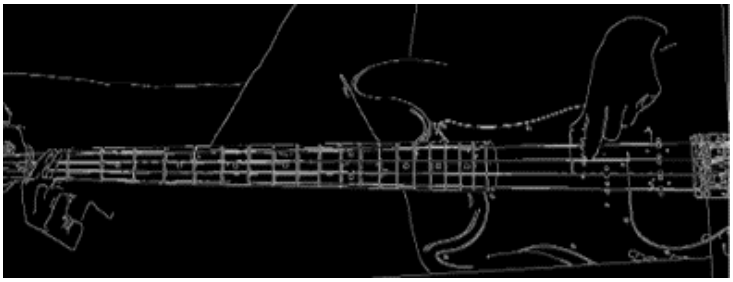

Figure 4.3: The frame with the Canny Edge algorithm applied.

Figure 4.4: The template edge image used for fretboard detection.

Even after edge detection, the fretboard must be explicitly located within the frame. In cases where the background is cluttered or the hands are partially occluding the instrument, relying on generic edge patterns is insufficient. Therefore, the system needs to find a spatially consistent, known pattern — this is done using template matching. Template matching is employed using OpenCV's `cv2.matchTemplate()` function. A predefined edge image template of a standard electric precision bass guitar fretboard sourced from the bass model manufacturer's website [23] (captured in ideal lighting and perspective conditions) is slid across the edge-detected frame. At each location, the correlation between the template and the input frame is computed, and the location with the highest correlation is assumed to contain the fretboard. This identified rectangular region becomes the ROI for all downstream detection tasks.

Figure 4.5: The detected ROI for the bass guitar fretboard located in the frame.

String Detection Using Hough Transform

Figure 4.6: The rotated frame with Hough Line Transform applied to the previously detected fretboard ROI, with lines filtered only to near-horizontal observations.
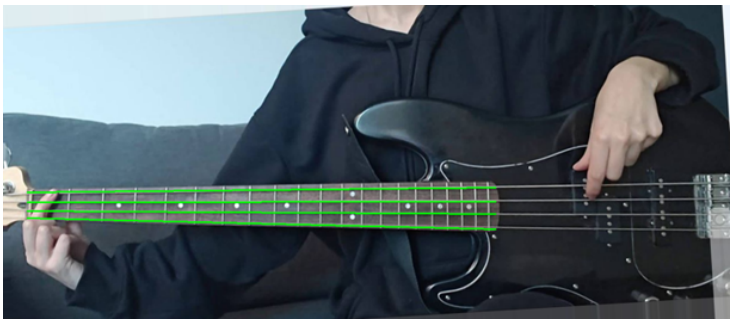
Figure 4.7: The previously seen Hough Line Transform results with post-processing performed to join strongest observations into four concrete strings.

Within the established ROI, the next task is to detect the four bass strings. The string positions serve as anchors for calculating both finger locations and plucking positions.

The system applies the Hough Line Transform within the ROI to detect near-horizontal lines, which correspond to the strings on the fretboard. The output often contains multiple short line segments due to variations in lighting, reflections, and fret markers. To consolidate these detections, a custom post-processing step merges overlapping or closely spaced line segments into single representative lines.

The four most prominent lines at the beginning and end of the fretboard are selected and connected to represent the strings. In cases where insufficient data is available (e.g., weak edge responses or occlusion), fallback geometry is used. If fewer than four viable lines are detected, the expected Y-positions of all strings are interpolated or extrapolated based on the known physical string spacing of a standard 4-string precision bass guitar (sourced from the manufacturer's website [23]). This hybrid approach ensures that string detection remains usable even under suboptimal conditions.
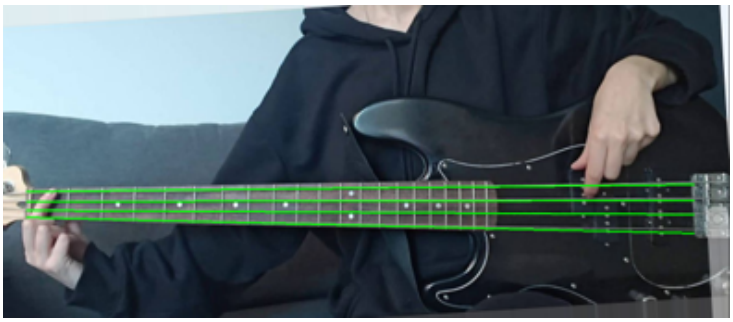
Figure 4.8: The string detection results along with extrapolation of observations to reach the bridge of the bass guitar.

Fret Position Estimation via Mathematical Modelling

Figure 4.9: Fret distance calculation results using distance formula drawn on fretboard.

Unlike strings, frets are not visually detected due to their thinness, poor visibility under common lighting conditions, and the lack of strong edge response. Instead, fret positions are calculated algorithmically using a well-established mathematical model.

The *12th-root-of-2 rule* [31], used in the design of stringed instruments, was implemented to estimate the location of each fret [36]. This formula defines the distance from the nut to the n-th fret as:

$$x_n = L \times \left(1 - 2^{-\frac{n}{12}}\right)$$

Where:

- x_n is the distance from the nut to the n-th fret,
- L is the total scale length (distance between the nut and the bridge),
- n is the fret number.

By identifying the pixel coordinates of the nut and the bridge within the ROI, the system computes the scale length in pixels. The formula is then applied iteratively to

Figure 4.10: Hough Line Transform applied to the fretboard region, with observations filtered to only display near-vertical lines.

Figure 4.11: The combined observations of the Hough Line Transform and the distance formula, displaying the final fret detection results.

calculate the pixel position of each fret from the nut along the longitudinal axis of the fretboard. These calculated positions are then drawn as vertical lines within the ROI, providing an accurate and consistent spatial reference for all frets.

To further refine accuracy, a secondary Hough Line Transform is performed specifically to detect near-vertical lines within the ROI. If any of these detected lines fall within a small margin of error (± 5 pixels) from a calculated fret position, they are used to correct the calculated positions. This blending of visual detection with mathematical prediction results in a highly reliable fret mapping system that can adjust dynamically to different fretboard sizes and camera setups.

Hand and Finger Tracking

Accurate hand and finger tracking is a critical component of the *Fret-Nav* system, as it allows the software to infer which note is being played based on finger positions and interactions with the instrument. To achieve this, the system employs MediaPipe Hands [42], a real-time hand tracking framework developed by Google.

MediaPipe Hands Integration

MediaPipe Hands is a high-performance framework that estimates 21 3D landmarks for each detected hand in real-time from an RGB image. The model pipeline consists of two main components:

- **Palm Detection:** A fast neural network first detects the presence of a palm in the video frame and returns a bounding box.
- **Hand Landmark Estimation:** Within the detected palm region, a second regression model estimates the 3D coordinates (x, y, z) of 21 specific landmarks corresponding to key hand joints, including fingertips, knuckles, and the wrist. The landmarks are indexed as follows:
  - **Index 0:** Wrist
  - **Indices 1-4:** Thumb
  - **Indices 5-8:** Index finger
  - **Indices 9-12:** Middle finger
  - **Indices 13-16:** Ring finger
  - **Indices 17-20:** Pinky

  In the context of the *Fret-Nav* system, MediaPipe Hands is used to detect both the fretting hand and the plucking hand. The video feed from the player's camera is processed in real-time, and the 2D landmark coordinates (x, y) are extracted frame-by-frame. The z-coordinates are not used in this implementation.

Finger Position Extraction

After MediaPipe Hands provides the 21 landmark coordinates, the system identifies which hand corresponds to which function:

- **Fretting hand:** The hand located within the predefined region of interest (ROI) that overlaps with the fretboard.
- **Plucking hand:** The hand positioned outside the fretboard ROI, closer to the strings toward the bridge end of the instrument.

Hand role classification is performed by comparing the average x-position of the wrist (landmark 0) to the x-range of the fretboard bounding box. This distinguishes the two hands in typical playing posture.

Fingertip Detection

For note analysis, only fingertip landmarks are used, specifically:

- Index tip: 8
- Middle tip: 12
- Ring tip: 16
- Pinky tip: 20

Each fingertip's position is recorded as (x, y) in image coordinates.

Plucking Detection

Plucking detection in the *Fret-Nav* system was implemented through a combination of fingertip tracking and motion analysis using the MediaPipe Hands framework. Specifically, the index and middle fingertips of the player's plucking hand were tracked across consecutive frames, and their positions relative to the detected string lines were stored in a rolling history buffer of up to 15 frames. For each frame, the system computed the perpendicular distance between the fingertip and each of the four string lines to determine the closest string. If the fingertip remained near one string (e.g., **A**) for several frames and then transitioned quickly to a position near another (e.g., **E**), with a significant upward movement along the y-axis in the rotated coordinate system, a plucking action was inferred. This motion-based detection was further validated by applying optical flow within a predefined region of interest encompassing the strings from the end of the fretboard to before the beginning point of the plucking hand. Optical flow was used to verify that motion occurred on the plucked string and, where applicable, on the adjacent lower-pitched string. This was because in observing plucking motions, it was identified that plucking one string resulted in the finger coming to rest on the string above it. This in turn created movement on both the string plucked and string above it, although only the plucked string made sound as the resting finger mutes the string it sits on. A special case was implemented for the lowest-pitched string (**E**), where a phantom string **X** was defined above it to maintain consistent detection logic. To avoid repeated detections of the same pluck due to frame-by-frame jitter, a cooldown mechanism was introduced that suppressed further detections for a short duration after a confirmed pluck. Once a pluck was validated, the system proceeded to identify the fretting finger position to determine the played note.

Figure 4.12: An image hand position at a pluck.

Figure 4.13: An image of a pluck being detected by the system.

4.2 Musical Information Processing

The musical information processing component of the *Fret-Nav* system was responsible for interpreting the musical content used during practice sessions and translating it into structured, machine-readable instructions that the visual feedback system could operate on. This module formed the bridge between static musical notation and dynamic real-time interaction, ensuring that the system could guide the player through a performance in a meaningful, note-by-note manner.

Tablature Input

The system was designed to accept musical input in the form of tablature, a format that is widely used and intuitively understood by beginner bass players. This form of notation lends itself well to computational parsing, and more importantly, aligns closely with the physical interaction required on the instrument.

To process tablature, the system parses a pre-formatted array structure where each note is represented as a tuple containing the string label and fret number (e.g., **(A, 5)** for the fifth fret on the **A** string). The parsed notes are stored as a sequential list representing

```
Tab = {
    ('A', 9), ('A', 11), ('D', 9), ...
}
```

Figure 4.14: A visual of sample tablature input for the system.

the performance order. Internally, these notes are also mapped to their expected spatial positions on the detected fretboard, allowing the system to generate visual cues in the appropriate location on the live video feed. The complete piece of music was displayed on each frame, allowing a user to follow a marker that pointed at the current position in the tablature throughout practice.

Current Note Tracking

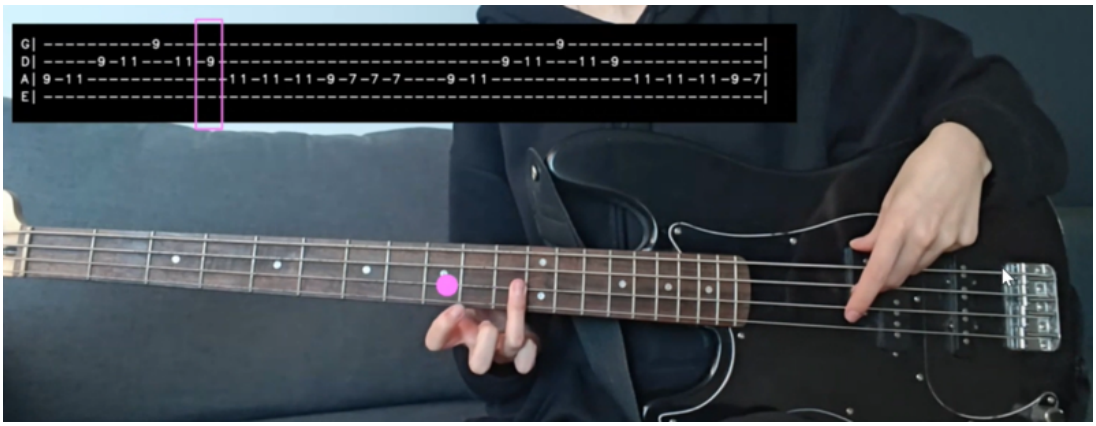

Figure 4.15: The *Fret-Nav* user interface, displaying the current notes being tracked using pink visual markers on both the tablature and the instrument.

A critical functionality of the system was the implementation of a dynamic note tracking mechanism. A pointer was maintained throughout each practice session to indicate the current target note in the tablature sequence. This pointer served as the anchor for the real-time feedback system, ensuring that the visual marker correctly identified the next note to be played by the user. The system waited for a match between the expected note (as defined by the pointer) and the actual note played, as inferred from the fretting and plucking analysis.

Upon successful detection of a correct note, meaning the correct string had been plucked and the finger was positioned on the correct fret, the pointer would automatically

increment to the next note in the sequence. If the note played was incorrect, the pointer remained stationary, and the system continued to provide visual feedback for the same note. This loop ensured that users could not progress through the piece without accurately playing each note, thus encouraging active memorisation of fretboard positions.

4.3 Guidance and Feedback

The guidance and feedback module served as the primary user-facing component of the system, translating the internal analysis and musical information into real-time, intuitive visual assistance for beginner bass players. This module provided immediate, frame-by-frame support to ensure that users could clearly understand what note to play next, receive feedback on their performance, and improve through repeated visual reinforcement.

Visual Cue Generation

Target Fret and String Identification

For each note in the tablature sequence, the system identified the corresponding target fret and string. This mapping was derived from the parsed tablature input, which stored notes as tuples of the form (`string`, `fret`). The system cross-referenced these values with the detected positions of the fretboard and strings — calculated during the initialization phase — to determine the exact pixel coordinates at which visual guidance should be rendered.

Marker Overlay

Once the target fret and string were identified, the system generated a visual marker to overlay onto the live video feed at the appropriate location. The coordinates of the marker were computed by intersecting the relevant horizontal string line and the vertical fret line, as determined through Hough Line Transform and scale length calculations. The marker dynamically followed the current note pointer from the musical information processor and updated in real time as the user progressed through the sequence. The overlay was rendered using OpenCV drawing functions and superimposed onto each processed frame before display.

Note Accuracy Assessment

Played Note Inference

To determine the note that had been played, the system combined data from multiple sources. The string that had been plucked was inferred within the designated string region of interest. Simultaneously, the fretting hand's fingertip positions were analysed to identify which finger was pressing down on the relevant string. This required computing overlaps between fingertip coordinates along the plucked string, and selecting the highest overlapping fingertip. The position of the fingertip on the string was compared to the computed fret regions, resulting in the calculation of the string and fret played.

Comparison with Expected Note

The inferred played note was then directly compared against the expected note from the tablature. A match occurred only if both the string and fret values aligned precisely. This strict checking ensured the user developed correct muscle memory for specific note locations on the fretboard.

Feedback Delivery

Progression to Next Note

Once the played note was confirmed as correct, the system advanced the current note pointer and updated the visual cue to reflect the next expected note in the tablature. This cycle continued until all notes in the loaded piece had been correctly performed. In the event of an incorrect note, the system did not advance. Instead, it retained the marker on the correct note, effectively signalling to the user that their input did not match expectations. The user was then required to correct their finger placement and plucking action before they could proceed.

Real-time Processing Pipeline

The system was designed for real-time operation. The video frames were continuously captured and processed, however the fretboard localisation was only re-calibrated once a note had been played. This was done to maintain the real-time operation of the system, as the steps involved in fretboard detection proved to be computationally expensive. The outputs of the visual input processing (fretboard location, hand position, plucking detection) and the musical information processing (current target note) were continuously fed into the guidance and feedback component to provide immediate assistance to the user.

4.4 Summary

This chapter detailed the technical implementation of *Fret-Nav*, a computer vision-based tool designed to aid beginner electric bass players with fretboard navigation and practice. The system integrates real-time video processing, machine learning-based hand tracking, and interactive feedback mechanisms to provide a seamless learning experience on the physical instrument.

The implementation began with the Visual Input Processing module, which captures and interprets live RGB video frames of the bass guitar. Using techniques such as template matching and the Hough Line Transform, the system localises the fretboard and detects string lines. Frets are mathematically mapped using the *12th-root-of-2 rule*, and their exact locations are subsequently refined using Hough Line Transform. The MediaPipe Hands framework enables accurate hand and fingertip tracking, while optical flow analysis detects string plucks based on motion vectors and fingertip trajectories.

The Musical Information Processing component manages input in the form of tablature. It maintains a real-time pointer to the current note the player should perform, allowing for sequential note tracking as the user practices.

The Guidance and Feedback system overlays visual cues (coloured markers) on the video feed to indicate the correct fret and string. It infers the note played from hand and plucking data and compares it to the expected note. Accurate performance triggers a progression to the next note, while incorrect notes prompt corrective feedback.

The result was a functional computer vision-based tool that provided on-screen visual cues to a bassist on their instrument, along with music notation to play along to.

Chapter 5

Evaluation

This chapter presents the evaluation of the *Fret-Nav* system through a combination of objective and subjective testing methods. The evaluation aimed to assess not only the accuracy of note recognition using visual input but also the impact of the tool on user experience, learning confidence, and engagement. The chapter is structured into two parts: the first presents quantitative results from an accuracy-based evaluation; the second offers insights from a user study involving beginner bass learners.

5.1 Accuracy Evaluation

The first stage of evaluation focused on the system's objective performance in detecting and recognising notes played on a real electric bass guitar. This evaluation aimed to quantify how reliably *Fret-Nav* identifies fretted notes through its computer vision pipeline, which integrates pluck detection and finger tracking. The results serve to validate the technical performance of the system as a real-time guidance tool.

5.1.1 Experimental Setup

To assess the performance of the *Fret-Nav* system, a test video was recorded in which a sequence of 157 individual notes were plucked across all four strings of a standard four-string precision model electric bass guitar. The goal of the experiment was to measure how effectively the tool could identify the correct note played based on computer vision-based detection of the instrument, finger positions, and plucking motions.

The ground truth was established by manually annotating each note played in the video with the corresponding frame number and note (string and fret) [1].

[1]While considerable care was taken to ensure the accuracy of this manual labelling, it is important to acknowledge the possibility of small inconsistencies due to the manual nature of the annotation process.

The system's predictions were then compared against this annotated ground truth. An observation was counted as a true positive if the predicted note (string and fret) and the ground truth matched, with a small frame tolerance (±5 frames). Conversely, false positives were defined as detections made by the system when no note was played, and false negatives occurred when a note was played but not detected by the system.

Figure 5.1: A frame from the testing video used to evaluate the objective performance of the system.

5.1.2 Results

To summarise the evaluation, a confusion matrix was constructed, categorising the results into true positives, false positives, and false negatives[2] .

	Actually Positive	**Actually Negative**
Predicted Positive	True Positives (TP): 151	False Positives (FP): 9
Predicted Negative	False Negatives (FN): 6	True Negatives (TN): N/A

Table 5.1: A confusion matrix representation of the obtained results of the objective evaluation, comparing system detections to a testing video with labelled ground truth.

From the above confusion matrix displayed in table 5.1, the following accuracy and precision metrics were computed:

Accuracy (note detection):

$$\text{Accuracy} = \frac{TP}{TP + FP + FN} = \frac{151}{151 + 9 + 6} = 0.9096 \approx 90.96\% \tag{5.1}$$

[2]True negatives are not directly measured in this setup, since the evaluation focused on frames where notes were expected.

Precision (how many detected notes were correct):

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{151}{151 + 9} = \frac{151}{160} \approx 94.37\% \tag{5.2}$$

The overall computed accuracy of approximately 91% demonstrates strong overall system performance. The high precision (94.37%) indicates that most detections made by the system were correct. This suggests that the visual detection pipeline, including both pluck recognition and fret detection, is generally robust in a controlled testing environment such as the one provided in the evaluation video.

Most false positives were caused by ambiguous plucking gestures and finger movement in the test video that were misclassified as note events. A significant amount of the false positives ($\approx 44\%$) were concentrated on the **E** string. This can likely be attributed to the string's position at the top of the instrument and the lack of a higher string above it, a factor that limits the system's ability to verify plucking gestures through motion patterns that depend on placing a finger on the string above the one being played, as it does for the **A**, **D**, and **G** strings. The increased sensitivity of the **E** string detection is a side effect of the model's design and although a phantom string is used to account for the differences of the top string, it may require future calibration or separate plucking detection rules to prevent over-triggering.

False negatives were most commonly caused by errors in finger tracking. In several instances, a string pluck was correctly identified, but the MediaPipe-based finger tracking component reported fingertip positions that fell just outside the expected fret region. This could be due to imprecise fingertip landmark estimation or noise in the video data. These results highlight that the reliability of note identification is highly dependent on the accuracy of both components — pluck detection and fret positioning — and even minor inaccuracies in either can result in missed detections.

The performance of *Fret-Nav* in this evaluation confirms that it is a viable assistive tool for beginner bass players, capable of providing accurate, real-time guidance in most common scenarios. However, the analysis also reveals key areas where further refinement could significantly enhance its reliability and pedagogical effectiveness.

5.2 User Study

In addition to the objective performance metrics, a subjective evaluation was conducted to assess the usability, intuitiveness, and learning impact of the *Fret-Nav* system.

5.2.1 Study Design

This study involved six anonymous participants, all of whom had varying levels of prior musical experience but crucially had never attempted to play the electric bass guitar before. They were divided into two groups for comparative analysis.

- **Group 1 - Traditional Method:** Participants learned a simple bass sequence using printed tablature.
- **Group 2 - *Fret-Nav* Users:** Participants used the *Fret-Nav* system to learn the same simple bass sequence.

All participants received a short introduction to two-finger plucking technique and how to read tablature. They were given one hour to learn and practice the same sequence.

5.2.2 Results

Following the session, each participant completed a structured survey combining Likert Scale questions and open-ended responses. The table below, table 5.2, provides an overview of the study results for the Likert Scale questions.

Survey Category	**Group 1 (Traditional)**	**Group 2 (*Fret-Nav*)**
Engagement	Agree to Strongly Agree	Agree to Strongly Agree
Ease of Learning	Neutral to Agree	Agree to Strongly Agree
Learning Method Effectiveness	Mostly Agree	Mostly Strongly Agree
Need for Additional Guidance	Neutral to Agree	Neutral to Disagree
Ability to Self-Correct	Mostly Neutral	Mostly Agree

Table 5.2: Summary of participant responses to Likert-style survey questions comparing traditional and *Fret-Nav* learning methods.

Reported Difficulties

Group 1 participants found multitasking between plucking and reading tablature particularly difficult. One user commented on the challenge of *"coordinating fingers while reading."* Another noted that counting frets manually made learning slower. Most notably, one participant reported: *"It was difficult to figure out whether the strings are going from up to down or down to up. Other things I can think of is the higher frets being trickier to quickly figure out where to place my fingers for a note."* These comments reinforce a key motivation behind the development of *Fret-Nav* — to reduce the burden of mental translation from static notation to physical fretboard positions.

In contrast, Group 2 participants cited early awkwardness adjusting to watching the fretboard rather than their hands: *"Getting used to watching the fretboard instead of looking at my hands was tricky at first."* This suggests that while *Fret-Nav* successfully anchors the user's attention to the physical instrument, it also requires a brief habitual shift for players used to looking at their hands. This points to a learning curve introduced by the system itself, which should be considered when designing further advancements in the system.

Suggested Improvements

Group 1 users suggested demonstration videos or more immediate feedback. Group 2 users requested larger on-screen tablature: *"Music displayed was pretty small, I would've liked for it to be a bit more zoomed in."*, and optional audio playback of correct notes: *"Maybe if you could also play the sound of the correct note, so I could hear it along with seeing it."*. One participant suggested game-like difficulty progression as a future feature. This feedback highlights several ways to improve *Fret-Nav* and aspects to consider when developing it further, such a making the system more intuitive, customisable, and multi-modal.

Engagement

While both groups reported being engaged in the session, the source of their engagement differed. The open-ended responses from Group 1 participants suggested the motivation was often driven by the experience of trying a new instrument rather than by the learning method itself. Group 2 also consistently marked "Agree" to "Strongly Agree" regarding engagement, however their open-ended responses explicitly referenced the tool itself as a positive influence.

Ease of Learning

The real-time feedback was a differentiating factor. Group 1 participant responses were mixed. One participant found tablature coordination to be difficult and recommended providing a *"demonstration video"*. Another participant rated the ease of learning as "Neutral". For Group 2, all students found the tool clear and easy to follow. Participants highlighted visual cues as the critical supports, helping them *"know exactly where to place fingers"*. This group reported benefiting from direct visual reinforcement, especially for absolute beginners. This supports the idea that *Fret-Nav* reduces cognitive load by externalising fretboard navigation.

Guidance

Traditional methods did little to resolve user uncertainty in real time. Most participants in Group 1 reported a need for more guidance. Comments also pointed toward a lack of confidence in knowing whether notes were played correctly. The traditional method placed a greater burden on participants to interpret notation without feedback. By contrast, *Fret-Nav* supported the learning process with visual cues. In Group 2, most users reported less need for additional help.

Self-Correction

Self-efficacy was higher in the *Fret-Nav* group. Two participants from Group 1 expressed neutrality on whether they could identify and fix mistakes. This lack of confidence in self-correction suggests the learning process was largely passive. *Fret-Nav* users uniformly reported stronger self-assessment. They felt they could correct mistakes in real time. This reporting suggests *Fret-Nav* transforms abstract tasks like finger positioning and note validation into concrete and observable actions.

5.3 Summary

The evaluation of the *Fret-Nav* system highlights two key outcomes:

- Objectively, the system achieved over 90% accuracy and 94% precision in note recognition, demonstrating the usability of visual note detection on a real instrument in controlled conditions.
- Subjectively, the system improved learners' confidence, self-correction ability, and engagement compared to traditional learning methods. Participants viewed the tool as intuitive, effective, and enjoyable.

While areas for refinement were identified — such as pluck detection calibration and UI adjustments — the results strongly support the tool's value as a beginner-oriented learning aid. The combination of real-time feedback and intuitive visual guidance clearly reduces the cognitive load associated with early-stage fretboard navigation.

Chapter 6

Conclusion

At the outset of this dissertation, a key gap in current music education technology as highlighted: while numerous digital tools exist to aid in musical performance or transcription, few directly support fretboard navigation for beginner electric bass players, showing an underdeveloped area of musical pedagogy. This project set out to address that challenge by designing and evaluating *Fret-Nav*, a computer vision-based learning tool that overlays real-time note guidance onto a user's physical instrument.

The central idea was simple: rather than asking beginners to interpret music via disconnected screens or abstract representations, *Fret-Nav* would place the instruction directly on the fretboard. This would allow users to develop familiarity and confidence through visual, spatial, and immediate interaction with their instrument.

The development of this system combined methods from Music Information Retrieval (MIR) and Computer Vision, resulting in a working prototype that could detect fretting and plucking actions from a live video stream and compare them against a target note extracted from tablature. Evaluation of this system involved both quantitative accuracy testing and a qualitative user study.

Having now implemented, tested, and evaluated the *Fret-Nav* system, this chapter concludes the dissertation with a detailed review and critique, followed by an outline of the project's core contributions and suggestions for future work.

6.1 Critique and Review

Visual Detection of Fretboard and Playing Actions

The visual component of Fret-Nav achieved its core objective: accurately detecting the fretboard, tracking finger placements, and detecting plucking actions. The use of template matching and the *12th-root* fret spacing formula allowed for an adaptable fretboard

mapping, while MediaPipe Hands provided dependable hand and finger tracking under typical conditions. However, it is important to note the controlled testing environment for this tool. A complex background, occluded views of the instrument, or unexpected patterns in a frame will come with direct impacts on the reliability of both fret and pluck detection.

Notably, string detection on the **E** string proved most error-prone. As the topmost string with no adjacent string for comparative motion detection, it posed a unique challenge for the pluck identification system. Although a workaround using an artificial "phantom" string was implemented, the results suggest this logic needs refinement to avoid false positives.

Overall, despite these limitations, the vision pipeline was capable of achieving over 90% note detection accuracy, indicating that *Fret-Nav* can function reliably in controlled environments and with standard instrument configurations.

Learning Impact on Beginners

The user study confirmed a pedagogical benefit for beginners. Compared to a traditional tablature-based learning method, *Fret-Nav* users demonstrated greater self-correction ability, increased confidence, and a reduced reliance on instructor input. Participants appreciated the clarity of visual cues and the immediate feedback loop, which helped them stay engaged and learn through trial and correction.

The only consistent critique from users was interface readability — specifically, the small size of the on-screen tablature. This was especially notable when viewed at a distance. The absence of auditory feedback (e.g., hearing the expected note) was also raised as a missed opportunity to reinforce correct learning.

Overall, the results indicate that real-time visual cues reduce the cognitive load associated with early-stage fretboard learning — a key outcome of this project.

6.2 Core Contributions

The *Fret-Nav* project made the following significant contributions:

1. A computer vision-based learning tool designed specifically for beginner electric bass players, addressing an underrepresented user group in MIR and AMT research.

2. A visual pipeline capable of fretboard localisation, finger tracking, and pluck detection using RGB video and open-source frameworks such as OpenCV and MediaPipe.

3. A real-time feedback loop that overlays visual markers on the user's instrument, updating based on detected playing input and loaded tablature.

4. An accuracy benchmark of 90.96% and precision of 94.37% in note detection across all strings in controlled test environments.

5. A qualitative evaluation framework, showing increased user engagement and learning retention compared to traditional methods.

6. An extendable system architecture, laying a foundation for further research into real-time vision-based music learning tools for string instruments.

These contributions demonstrate both technical feasibility and pedagogical relevance, highlighting the potential for vision-based tools to fill a gap in beginner-focused music education, especially relating to bass guitars.

6.3 Future Work

While *Fret-Nav* achieved its core objectives, the development and evaluation process uncovered several areas for possible future work.

Audio-Visual Integration

Currently, *Fret-Nav* relies exclusively on visual input to determine finger positions and plucking events. While this is sufficient in many scenarios, the system could benefit significantly from the integration of audio analysis. Incorporating audio would allow for confirmation of note pitch and tonal quality, serving as a secondary validation layer. This would be especially useful in cases where visual data is ambiguous or occluded — for example, when a finger briefly hovers near a fret without pressing it firmly, or when pluck detection is uncertain.

Enhanced Pluck Detection

While the current implementation of pluck detection based on fingertip motion and optical flow performs well, it has limitations — particularly with the topmost string, where there is no adjacent string above to determine and verify a plucking movement. A promising direction for future improvement is to implement string-specific pluck sensitivity thresholds, taking into account the physical asymmetry of the bass guitar.

Interface Improvements

Several users noted that the visual tablature display was small or difficult to interpret from a distance. Improving the interface could involve not only scaling up and optimising the display for various screen sizes, but also implementing features such as zoom, colour-coded feedback, and audio playback of target notes. Providing a visual or auditory indication when a note is played correctly could also reinforce learning and create a more engaging experience. Moreover, support for left-handed players, dark mode, and interface language options could broaden the accessibility of the tool to a more diverse user base.

Generalisation to Real-World Conditions

The system was tested in a relatively controlled environment, with consistent lighting and a fixed camera setup. However, in real-world practice environments, such conditions are rarely guaranteed. Future iterations of *Fret-Nav* should aim to improve resilience to variable lighting, background clutter, and dynamic camera angles. Machine learning models trained on diverse playing environments could help the system generalise across different settings. Improving reliability in less controlled spaces would significantly increase the system's practicality and usability.

Instrument Generalisation

While this version of *Fret-Nav* was designed specifically for the four-string precision model electric bass, the underlying methodology is applicable to a wide range of fretted string instruments, including five- and six-string basses, electric and acoustic guitars, and even ukuleles, banjos, or mandolins. Adapting the tool to other instruments would primarily involve collecting templates for the instrument fretboard, recalibrating fret spacing and note mapping logic, as well as accounting for different string tunings and hand positions. Future work could involve developing a modular instrument profile system, where users can select or configure their instrument type and have the system adjust accordingly. This would enable *Fret-Nav* to become a flexible learning platform for a wide array of beginner learners.

6.4 Final Remarks

This dissertation has explored how computer vision can directly support music learning, not by replacing human instruction, but by enhancing it — making the process more intuitive, interactive, and accessible for beginners.

By transforming abstract note notation into real-time visual guidance placed directly on the instrument, *Fret-Nav* empowers learners to develop muscle memory, confidence, and fretboard fluency. The tool stands as a proof-of-concept for an underexplored application of computer vision in education: enabling musical fluency at the very start of the learning journey.

Bibliography

[1] Jakob Abeßer. Automatic string detection for bass guitar and electric guitar. In *From Sounds to Music and Emotions: 9th International Symposium, CMMR 2012, London, UK, June 19-22, 2012, Revised Selected Papers 9*, pages 333–352. Springer, 2013.

[2] Alberto Acquilino and Gary Scavone. Current state and future directions of technologies for music instrument pedagogy. *Frontiers in Psychology*, 13:835609, 2022. Published: 22 March 2022.

[3] Mark Asmar. A computer vision-based automatic transcription of guitar music from rgbd videos. Master's thesis, Polytechnique Montréal, Montréal, Canada, 2022. Supervised by Lama Séoud and Guillaume-Alexandre Bilodeau. Program: Génie informatique.

[4] Juan Pablo Martinez Avila, Adrian Hazzard, Alan Chamberlain, Chris Greenhalgh, and Steve Benford. An ai-based design framework to support musicians' practices. AM '18, New York, NY, USA, 2018. Association for Computing Machinery.

[5] Tony Bacon and Barry Moorhouse. *The Bass Book: A Complete Illustrated History of Bass Guitars*. Backbeat Books, 2016.

[6] Bass Player Center. Learn bass guitar – how to read bass tab, 2023. URL `https://bassplayercenter.com/learn-bass-guitar-how-to-read-bass-tab/`.

[7] BassBuzz. Basic bass plucking technique (beginner bass basics), 2020. URL `https://www.youtube.com/watch?v=CR8yQCZX2HQ`.

[8] BassBuzz Forum. Learn your bass fretboard notes - easy starter method, 2020. URL `https://forum.bassbuzz.com/t/learn-your-bass-fretboard-notes-easy-starter-method/3131/3`.

[9] Bhuwan Bhattarai and Joonwhoan Lee. A comprehensive review on music transcription. *Applied Sciences*, 13(21), 2023.

[10] J. A. Burgoyne, Ichiro Fujinaga, and J. Stephen Downie. Music information retrieval. In Susan Schreibman, Ray Siemens, and John Unsworth, editors, *A New Companion to Digital Humanities*. Wiley-Blackwell, 2015.

[11] Charles Chaffin. An examination of the cognitive workload associated with conducting in an instrumental music context: A review of literature. *Bulletin of the Council for Research in Music Education*, 189:73–88, 2011.

[12] Stuart Clayton. *The Bass Guitarist's Guide to Reading Music – Beginner Level*. Stuart Clayton Publishing, 2013.

[13] Kenneth Dawson-Howe. *A Practical Introduction to Computer Vision with OpenCV*, chapter 5, page Section 5.3. Wiley Publishing, 1st edition, 2014.

[14] Kenneth Dawson-Howe. *A Practical Introduction to Computer Vision with OpenCV*, chapter 6, page Section 6.1. Wiley Publishing, 1st edition, 2014.

[15] Kenneth Dawson-Howe. *A Practical Introduction to Computer Vision with OpenCV*, chapter 8, page Section 8.1. Wiley Publishing, 1st edition, 2014.

[16] Kenneth Dawson-Howe. *A Practical Introduction to Computer Vision with OpenCV*, chapter 6, page Section 6.3. Wiley Publishing, 1st edition, 2014.

[17] Kenneth Dawson-Howe. *A Practical Introduction to Computer Vision with OpenCV*, chapter 9, page Section 9.2. Wiley Publishing, 1st edition, 2014.

[18] Christian Dittmar, Estefanía Cano, Jakob Abeßer, and Sascha Grollmisch. Music information retrieval meets music education. In Meinard Müller, Masataka Goto, and Markus Schedl, editors, *Multimodal Music Processing*, volume 3 of *Dagstuhl Follow-Ups*, pages 95–120. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012.

[19] Andrew Duncan. Combinatorial music theory. *Journal of the Audio Engineering Society*, 39(6):427–448, June 1991. © 1991, Audio Engineering Society and Andrew Duncan. All rights reserved.

[20] Walaa H. Elashmawi, John Emad, Ahmed Serag, Karim Khaled, Ahmed Yehia, Karim Mohamed, Hager Sobeah, and Ahmed Ali. A novel approach for improving guitarists' performance using motion capture and note frequency recognition. *Applied Sciences*, 13(10), 2023.

[21] John Emad, Hager Sobeah, Walaa Hassan, Ahmed Serag, Ahmed Yehia, Karim Khaled, and Karim Mohamed. Note frequency recognition and finger motion capture of a guitarist: A survey and challenges ahead. *Journal of Computing and Communication*, 1(2):69–79, August 2022. Original Article.

[22] Enthu. Bass guitar notes: A complete guide for beginners, 2023. URL `https://enthu.com/blog/guitar/bass-guitar-notes/`.

[23] Fender Musical Instruments Corporation. Precision bass collection, 2025. URL `https://ie.fender.com/collections/electric-basses-precision-bass`.

[24] Ali Ghaleb, Eslam Elsadawy, Ihab Essam, Seif-Eldin Zaky, Mohamed Abdelhakim, Natalie Fahim, Razan Bayoumi, and Hanan Hindy. Taptotab: Video-based guitar tabs generation using ai and audio analysis. In *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 148–155, 2024.

[25] Shir Goldstein. Vision-based musical notes recognition of string instruments. Master's thesis, Reichman University (Israel), 2017.

[26] Eli Harrison. Challenges facing guitar education. *Music Educators Journal*, 97(1):50–55, September 2010.

[27] Fatemeh Jamshidi, Gary Pike, Amit Das, and Richard Chapman. Machine learning techniques in automatic music transcription: A systematic survey. *arXiv preprint arXiv:2406.15249*, 2024.

[28] Christian Kehling, Jakob Abeßer, Christian Dittmar, and Gerald Schuller. Automatic tablature transcription of electric guitar recordings by estimation of score- and instrument-related parameters. In *International Conference on Digital Audio Effects*, 2014.

[29] Jeremy LaBrooy. Bass fretboard mastery — essential skills for beginners, 2024. URL `https://bassictraining.com/bass-fretboard-mastery-for-beginners/`.

[30] Dave Lockwood. Mixing di bass guitar. *Sound On Sound*, August 2022. URL `https://www.soundonsound.com/techniques/mixing-di-bass-guitar`.

[31] R.M. Mottola. Fret placement. URL `https://www.liutaiomottola.com/formulae/fret.htm`.

[32] LF Neal. Parts of the bass guitar (anatomy explained with diagrams), 2023. URL `https://leftyfretz.com/parts-of-the-bass-guitar/`.

[33] Jeff Owens. Legendary lows: The precision bass story, 2022. URL https://www.fender.comarticles/behind-the-scenes/legendary-lows-the-precision-bass-story.

[34] Laurel S. Pardue and Andrew McPherson. Real-time aural and visual feedback for improving violin intonation. *Frontiers in Psychology*, 10:627, 2019.

[35] Rodrigo Schramm and Cláudio R. Jung. A tool for teaching musical metrics based on computer vision. *The Visual Computer*, 2022. Manuscript accepted, publication details to be inserted by the editor.

[36] Stewart-MacDonald. Fret position calculator. URL https://www.stewmac.com/fret-calculator/.

[37] StudyBass. How to read bass tab, 2025. URL https://www.studybass.com/lessons/reading-music/how-to-read-bass-tab/.

[38] TalkBass Community. Learning the notes on the bass fretboard for beginners: Another way of looking at it, 2010. URL https://www.talkbass.com/threads/learning-the-notes-on-the-bass-fretboard-for-beginners-another-way-of-looking-715093/.

[39] Bryan Wade. How to instantly find any fret on the guitar without counting, 2024. URL https://www.bryanwade.com/how-to-instantly-find-any-fret-on-the-guitar-without-counting/.

[40] Bryan Wang, Meng Yu Yang, and Tovi Grossman. Soloist: Generating mixed-initiative tutorials from existing guitar instructional videos through audio processing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[41] Yamaha Corporation. How to play the electric guitar : There are many ways to play the electric bass, 2025. URL https://www.yamaha.com/en/musical_instrument_guide/electric_guitar/play/play005.html.

[42] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*, 2020.