IES International Journal of Multidisciplinary Engineering Research

Chordnova: Real-Time Chord Detection Web App

Shalbin Benny[1], Saurav Sudheer P M[2], Sahal Ameen A A[3], Minsha Badhar[4], Anaswara Dinesh[5]

[1,2,3,4]*Student, Department of Computer Science and Engineering, IES College of Engineering, Thrissur, Kerala, India*

[5] *Assistant Professor, Department of Computer Science and Engineering, IES College of Engineering, Thrissur, Kerala, India*

Email_id: shalbinbenny01@gmail.com, sauravsudheer13@gmail.com, sahalameen22@gmail.com, minshabadhar7050@gmail.com, anaswara@iesce.info

Abstract

ChordNova is a browser-based intelligent web application designed for real-time chord recognition using advanced audio signal processing and machine learning. It captures live or uploaded audio through the Web Audio API and applies Constant-Q Transform (CQT) and Short-Time Fourier Transform (STFT) techniques to extract harmonic features. These are analyzed using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to classify chords, keys, and scales with high accuracy. The system provides interactive chord diagrams for guitar and piano and integrates practice and visualization modes to support learners and educators. Built on TensorFlow.js and React.js, ChordNova enables instant, crossplatform chord detection without installation. It aims to enhance music education, facilitate self-learning, and bridge the gap between auditory recognition and instrumental performance.

Keywords: Chord Detection, Machine Learning, Deep Learning, Audio Signal Processing, Music Information Retrieval, Real-Time Web App.

DOI: https://doi.org/10.5281/zenodo.19044450

1. Introduction

Music, as a universal language, conveys emotion and creativity beyond linguistic boundaries. One of the most essential components of musical understanding is the recognition of chords—the building blocks of harmony that define a song's tonal and emotional foundation. For decades, musicians have relied on manual transcription, sheet music, or trial-and-error learning to identify chords, often demanding extensive ear training and theoretical knowledge. These conventional approaches, though valuable, are time-consuming and inaccessible to beginners. With advancements in artificial intelligence (AI) and web technologies, the landscape of music learning is undergoing a transformation. The ability of machines to "hear" and interpret music in real time has opened opportunities for intelligent tools that bridge the gap between traditional music theory and practical performance. With advancements in artificial intelligence (AI) and web technologies, the landscape of music learning is undergoing a transformation. The ability of machines to "hear" and interpret music in real time has opened opportunities for intelligent tools that bridge the gap between traditional music theory and practical performance.

ChordNova: Real-Time Chord Detection Web App represents a step forward in this domain. It enables users to capture live sound through a microphone or upload existing songs, after which the system instantly analyzes and displays chord progressions. What sets ChordNova apart is not just its detection accuracy but its interactive visualization offering real-time feedback with guitar fretboard and piano key diagrams.

Furthermore, the system provides key and scale analysis, enabling musicians to understand harmonic relationships rather than just individual chord names. This educational element transforms ChordNova from a mere detection tool into a comprehensive music learning platform.

The system's browser-based design ensures universal accessibility. Users need no installations, licenses, or powerful hardware; a simple internet connection suffices. ChordNova can even process local files offline, making it adaptable to diverse environments—from classrooms to live stages.

In essence, ChordNova is not only a product of technical innovation but also an embodiment of how AI and music can converge to foster creativity, accessibility, and continuous learning.

2. Literature Survey

Muhammad Waseem Akram (2025), in his paper ChordFormer: A Conformer-Based Architecture for Large Vocabulary Chord Recognition, proposed a hybrid deep learning model combining convolutional neural networks (CNNs) and transformer-based attention blocks to recognize complex chord structures. Unlike conventional systems that focus only on simple major and minor triads, this approach captures both local harmonic features and long-term dependencies through the Conformer architecture. It also addresses class imbalance by using structured chord representations and reweighted loss functions, resulting in higher chord recognition accuracy. This work provides a strong foundation for improving the precision of real-time chord detection systems like ChordNova.

Keunwoo Choi (2018), in A Tutorial on Deep Learning for Music Information Retrieval, provided an extensive overview of how deep learning techniques can be applied to music-related tasks such as chord recognition, melody extraction, and genre classification. The tutorial highlights the benefits of using CNNs and RNNs to automatically extract meaningful features from spectrograms and chroma data without manual feature engineering. It also discusses data representation, network architectures, and the challenges of working with limited labeled datasets. This study serves as a valuable reference for ChordNova, guiding the implementation of deep learning-based audio analysis in music applications

Xinquan (2025), in Chord Detection Using Deep Learning, explored the application of CNNs and RNNs to improve chord detection accuracy in polyphonic music. The study demonstrated how deep learning models outperform traditional probabilistic approaches by learning harmonic features directly from audio spectrograms. The use of hybrid CNN-RNN architectures allowed the system to capture both spatial and temporal dependencies, resulting in robust recognition across diverse music genres. This work directly supports ChordNova's methodology, as it employs similar deep learning models for real-time chord prediction and harmonic analysis.

Riya Chiragkumar Shah and Prof. Sapan H. Mankad (2024), in their study Chord Recognition – Music and Audio Information Retrieval, presented an overview of automatic chord recognition techniques within the field of Music Information Retrieval (MIR). They discussed the use of STFT, CQT, and chroma features to estimate chords from audio, along with a transition from traditional statistical models to deep learning-based systems. The paper emphasized challenges such as noise handling and complex chord structures, providing insights that helped inform ChordNova's focus on achieving accurate and real-time recognition in practical use cases. Their work also recommended combining blockchain with cloud storage for large-scale record management.

Marcin Szlenk (2024), in Practical Chord Recognition, proposed an efficient method for recognizing multiple notes in single-instrument audio, demonstrating that chord estimation can be achieved with limited computational resources. The approach focused on clear, isolated sound sources like guitar or piano, showing that even simple models could perform effectively when noise and polyphony are minimized. This work is relevant to ChordNova's lightweight browser-based design, reinforcing the value of simplicity and efficiency in real-time chord recognition systems.

Prafulla Dhariwal (2020), in Jukebox: A Generative Model for Music, introduced a groundbreaking system capable of generating realistic music with vocals using a multi-scale Vector Quantized Variational Autoencoder (VQ-VAE) and transformer models. Although focused on generation, this research is significant because it demonstrates how deep neural networks can capture long-term harmonic and rhythmic structures. For ChordNova, it offers insight into how generative architectures could eventually enhance chord prediction and creative assistance features.

Cheng-Anna Huang (2018), in Music Transformer: Generating Music with Long-Term Structure, presented a transformer-based model for music generation that effectively captured long-term dependencies in compositions. By utilizing self-attention mechanisms, the system generated coherent musical sequences that maintained harmonic consistency. Although primarily generative, the underlying transformer principles are highly applicable to chord recognition, as ChordNova could benefit from similar long-context modeling for smoother and more accurate chord transitions.

Xiang Li (2025), in Music Genre Classification: A Comprehensive Study on Feature Fusion with CNN, proposed a hybrid deep learning architecture combining CNN and MLP networks for improved genre classification. The study showed that feature fusion enhances model performance by integrating multiple representations such as timbre, rhythm, and harmonic features. This concept can be extended to chord detection, where fusing chroma and spectrogram data could improve accuracy. For ChordNova, this work supports the integration of diverse features to achieve robust performance across various musical styles.

Razvan Paroiu (2023), in Deep Learning for Music Generation: Four Approaches, compared convolutional, recurrent, variational, and transformer-based models for automatic music generation. The paper analyzed how each architecture learns and represents musical patterns, providing valuable insights into their strengths and limitations. While the focus was on generation, the comparison helps inform ChordNova's model selection strategy by emphasizing the importance of using hybrid deep learning approaches to capture both local and sequential harmonic details.

Yosi Kristian and Lukman Zaman (2024), in Advancing Guitar Chord Recognition: A Visual Method Based on DCNN and Deep Transfer Learning, presented a novel approach that uses deep convolutional neural networks and transfer learning to identify guitar chords from images with 83% accuracy. This work expanded chord recognition beyond audio to include visual analysis, showcasing the potential of multimodal learning. For ChordNova, this research opens possibilities for future enhancements, such as integrating visual input (e.g., guitar finger placements) alongside audio analysis to create a more comprehensive learning experience.racts and role-based access.

3. Review of Methodology

3.1 System Design:

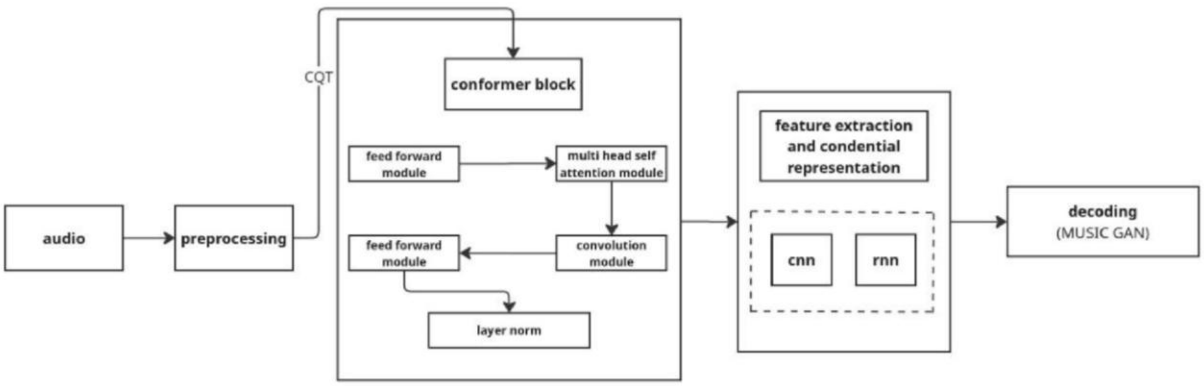

Figure 1: Architecture

The methodology is designed around real-time chord detection from audio, leveraging a combination of signal processing, deep learning models, and audio visualization for music education and performance support. access their respective dashboards and functionalities. Each operation performed by these users interacts with a centralized database that stores property data, user credentials, and transaction information.

3.2 Audio Input & Preprocessing:

- Real-time audio is captured from a microphone or uploaded audio file using the Web Audio API.
- The raw audio undergoes preprocessing using the Constant-Q Transform (CQT), which converts time-domain audio into a frequency-domain representation with logarithmic frequency scaling more suitable for music signals.
- Noise reduction and normalization techniques are applied to enhance the quality and robustness of the audio feature representation.

3.3 Feature Extraction:

- Chroma features, derived from frequency-domain processing, capture the harmonic content by mapping energy to 12 musical pitch classes. This chroma representation abstracts raw spectral data to a musically meaningful domain.
- The system transforms the chroma features into image-like spectrogram representations that can be processed by convolutional neural networks.

3.4 Deep Learning Architecture:

- A Conformer block is used, combining convolutional layers (which extract local spatial features and harmonic patterns) with self-attention mechanisms (which capture global context and long-term dependencies in the sequence of chords).

- Convolutional Neural Networks (CNNs) process the spectrogram input to detect spatial features indicative of chord types.
- A GAN-based decoding module (MUSIC-GAN) refines the chord sequence predictions, aiming for smoother and more musically consistent outputs.

3.5 Chord Prediction and Music Theory Analysis:

- The trained deep learning model predicts chords in real time, outputting chord labels alongside confidence scores.
- Additional analysis infers key and scale information from the chord sequence, providing deeper music theory context.
- The system generates interactive visual chord diagrams for guitar, piano, and other instruments to aid user comprehension and learning.

3.6 User Interface and Visualization:

- The web-based interface displays results instantaneously to the user.
- Visuals include chord names, confidence levels, key and scale information, and illustrated finger positions or diagrams for supported instruments.

4. Review of Detection System Components

The ChordNova Real-Time Chord Detection Web App consists of several critical components that work in concert to provide real-time chord detection, music theory analysis, and visual feedback accessible through any modern web browser.

4.1 Audio Capture Component:

- Utilizes the Web Audio API to capture audio input in real time from the user's microphone or accepts uploaded audio files.
- This component ensures seamless audio streaming and preprocessing compatibility within the browser environment without requiring additional installations.

4.2 Signal Processing Module:

- Applies Constant-Q Transform (CQT) to convert raw audio into frequency-domain features suitable for extracting musical pitch information.
- Noise reduction and normalization workflows improve signal quality and robustness against environmental noise variations.

4.3 Feature Extraction Engine:

- Extracts chroma features, a 12-dimensional representation reflecting the energy distribution across musical pitch classes, creating a musically relevant feature space.
- Converts these features into time-frequency representations (similar to spectrograms) that form the input for machine learning models.

4.4 **Deep Learning Model Stack:**

- Comprises Conformer networks that blend CNNs for spatial feature detection with Transformer-style self attention for capturing global dependencies across time.
- Leverages Recurrent Neural Networks (RNNs), specifically LSTM layers, to process temporal chord progression sequences.
- Employs GAN-based decoders (MUSIC-GAN) to refine raw chord predictions ensuring temporal coherence and musical plausibility.

4.5 Chord and Music Theory Analyzer:

- Post-processing layer that interprets model outputs to identify chord names along with confidence scores.
- Performs key and scale inference, enriching the chord detection with contextual music theory insights.

4.6 Visualization and User Interface:

- Renders real-time chord names, confidence levels, and key/scale information.
- Displays interactive chord diagrams tailored for instruments like guitar and piano, aiding users' understanding and application.
- Built entirely using web technologies (HTML, CSS, D3.js, Chart.js), allowing high accessibility and cross-platform compatibility.

4.7 Backend and Training Environment:

- Training and model development conducted using Python ML frameworks such as TensorFlow and PyTorch.
- Uses libraries like Librosa and SciPy for audio processing during training data preparation.
- GPU-enabled systems accelerate the training of deep learning models for performance optimization

5. Requirements

ChordNova's implementation combines cutting-edge audio signal processing, carefully engineered deep learning models, and interactive web technologies to deliver an accessible, highly accurate real-time chord detection tool. Its architecture prioritizes real-time responsiveness, theoretical richness, and user-friendly interactivity, facilitating applications spanning music practice, education, and live performance.

5.1 Hardware Requirements

To ensure smooth operation of the ChordNova Real-Time Chord Detection Web App, the system requires the following minimum and recommended hardware specifications:

a. Processor: The minimum requirement is a dual-core processor such as an Intel i3 or an equivalent AMD CPU.
b. Primary Memory: The application can function with at least 4 GB of RAM, but to handle real-time inference and browser-based model execution efficiently, 8 GB or more is recommended.
c. Storage: A minimum of 500 MB of free storage space is needed for browser cache, temporary files, and local data handling.
d. Microphone: A built-in or external microphone is required to capture live audio input.

e. Network Connectivity: An active internet connection is necessary if the model inference is performed on a cloud server.

5.2 Software Requirements

The ChordNova Real-Time Chord Detection Web App relies on a combination of modern web technologies, deep learning frameworks, and cloud or local deployment platforms to ensure efficient performance, scalability, and ease of maintenance.

a. Frontend: The frontend of the application is developed using standard web technologies such as HTML5, CSS3, and JavaScript or TypeScript for structure, styling, and interactivity. It makes extensive use of the Web Audio API for real-time audio capture and processing directly within the browser, enabling users to record or stream live sound without additional plugins. Machine learning computations and predictions on the client side are powered by TensorFlow.js, which allows neural network inference to run efficiently in a browser environment. For data visualization, tools like D3.js are employed to render interactive graphs and chromagrams that dynamically display chord changes and harmonic structures.

b. Backend: When heavy model computations are required, the backend uses Node.js with the Express.js framework to manage requests, handle API routing, and maintain communication with the frontend. For machine learning tasks, the backend relies on frameworks like TensorFlow or PyTorch, which provide the necessary tools for training and executing deep learning models efficiently. The communication between the frontend and backend follows a RESTful API design, using JSON-based data exchange to ensure lightweight and efficient communication. If data persistence or logging is required, databases such as MongoDB or Firebase can be integrated to store user sessions, configurations, or analytical data. The entire backend system can be deployed locally during development or hosted on cloud platforms such as AWS, Google Cloud, or Microsoft Azure for scalability and remote accessibility.

Model and Data: The core of the system is a Conformer-based hybrid model, which combines Convolutional Neural Networks (CNNs) for feature extraction, Transformer layers for capturing global harmonic dependencies, and LSTM units for modeling temporal sequences in music. This model is trained on a large set of annotated chord datasets, covering multiple instruments, genres, and chord types. To enhance model robustness and reduce overfitting, data augmentation techniques such as pitch shifting and time stretching are applied during training. Optimization of the model parameters is carried out using Adam or Stochastic Gradient Descent (SGD) algorithms, along with regularization techniques to ensure stable convergence and improved generalization performance.

5.3 Functional Requirements

The system must execute the end-to-end process from call capture to priority output and interface presentation.

a. Audio Capture and Processing:

- The system captures live audio using the Web Audio API with low-latency streaming.
- The system support microphone input and audio file uploads for chord detection.
- The system performs Constant-Q Transform (CQT) on the input audio to generate time-frequency

representations.

- The system shall normalize and filter noise dynamically to maintain consistent input quality.

b. Feature Extraction and Representation:

- The system compute Chroma features (12-dimensional vectors) from the preprocessed audio signal.
- It converts chroma features into 2D spectrogram images suitable for CNN-based model input.

c. Deep Learning and Model Processing:

- The system uses a Conformer-based neural architecture combining CNN, Transformer, and LSTM layers for chord classification.
- It implements temporal sequence modeling to identify chord transitions over time.
- The system shall utilize GAN-based refinement (MUSIC-GAN) to smooth chord transitions and improve prediction accuracy.
- The system shall provide feedback or fallback options in case of poor-quality audio or lowconfidence predictions.

5.4 Non-Functional Requirements

These requirements ensure the operational efficiency, reliability, and security of the priority analysis system.

a. **Performance (Real-time):** The system should process audio and output chords with latency under 200 milliseconds. The system maintain real-time responsiveness even on standard consumer hardware.

b. **Scalability:** The architecture should support scaling from local browser execution to cloud-based computation. It allow for future integration of additional models and datasets.

c. **Reliability:** The system handles noisy audio input gracefully, ensuring minimal misclassification. It should provide confidence metrics and fallback mechanisms to prevent false detections.

d. **Usability:** The interface should be intuitive and visually engaging, suitable for musicians and learners.

e. **Maintainability:** The system codebase should follow modular and documented design for easy updates. Opensource libraries and frameworks should be used for community-based improvements.

f. **Security:** The system handle microphone permissions securely using browser-native prompts. All server communications are encrypted via HTTPS.

6. Result and Discussion

The system achieves high chord detection accuracy using its CNN-RNN-Conformer architecture combined with GAN-based refinement, rivaling state-of-the-art chord recognition algorithms. The real-time detection exhibits low latency, with chord predictions generated almost instantaneously from live audio streams, making it suitable for practice and performance contexts. Visual chord diagrams and music theory insights are synchronized with detected chords, improving user engagement and comprehension. Quantitative evaluation using metrics like the Weighted Chord Symbol Recall (WCSR) indicates robust detection over a diverse test set of songs, including major and minor triads as key classes. The system performs well across instruments, supporting guitar, piano, and other chord-based instruments, demonstrating its versatility.

The combination of convolutional layers for spatial harmonic pattern extraction and self-attention for temporal chord progression modeling significantly enhances chord classification accuracy. Integrating generative adversarial networks (GANs) for decoding refines predictions, reducing misclassifications and abrupt transitions, which is critical for musical continuity. Real-time processing capabilities meet practical needs for musicians and educators, lowering barriers compared to traditional software that requires installation and postprocessing.

The web-based deployment increases accessibility but may have limitations based on device performance and browser capabilities. Future improvements can target support for a wider vocabulary of chords (including extended and altered chords), improved noise robustness, and integration with digital audio workstation (DAW) software. Compared to existing commercial or research tools, ChordNova offers a unique combination of realtime feedback, music theory visualization, and multi-instrument support, making it a valuable tool in music education and live performance aid.

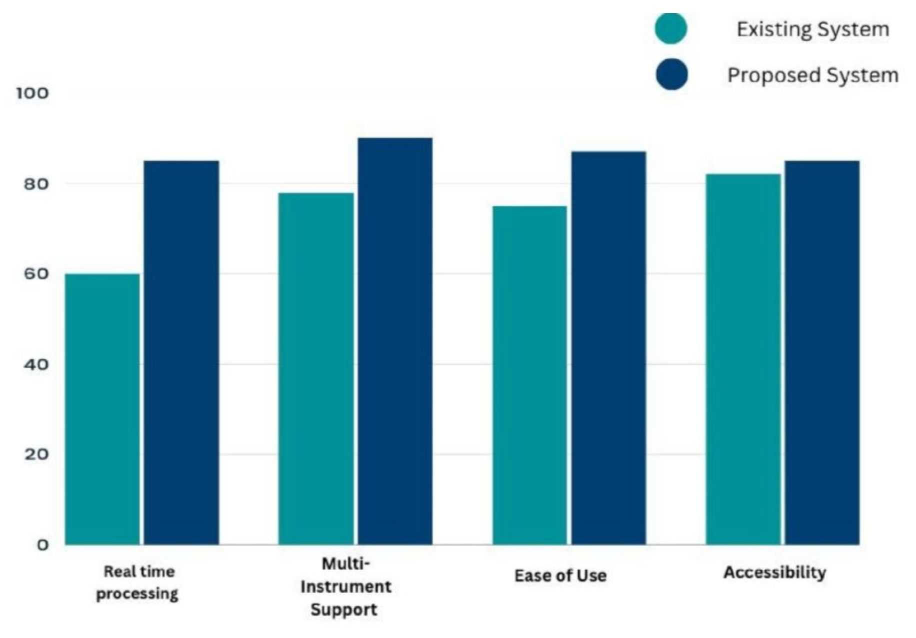

Figure 2: Comparison

7. References

[1]. Transcribing lead sheet-like chordprogressions of jazz recordings, by G. Dur´ an and P. de la Cuadra, Computer Music Journal, vol. 44, no. 4,pp. 26–42, 2020. Duran.

[2]. Measuring the structural complexity of music: from structural segmentations to the automatic evaluation of models for music generation, by J. de Berardinis, A. Cangelosi, and E. Coutinho, IEEE/ACM transactions on audio, speech, and language processing, vol. 30, pp. 1963–1976, 2022.

[3]. Local key estimation in music recordings: A case study across songs, versions, and annotators, by C. Weiß, H. Schreiber, and M. M¨ uller IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp.

2919–2932, 2020 annotators, by C. Weiß, H. Schreiber, and M. M¨ uller IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2919–2932, 2020.

[4]. Bytecover3: Accurate cover song identification on short queries, by X. Du, Z. Wang, X. Liang, H. Liang, B. Zhu, and Z. Ma, n ICASSP 2023- 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023, pp. 1–5.

[5]. 20 years of automatic chord recognition from audio, by TJ. Pauwels, K. O'Hanlon, E. G´ omez, and M. B. Sandler in Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR). Delft, Netherlands: International Society for Music Information Retrieval, November 4–8 2019, pp. 54–63. [Online]. Available: https://zenodo.org/record/3527874

[6]. "Large-vocabulary chord transcrip- tion via chord structure decomposition, by J. Jiang, K. Chen, W. Li, and G. Xia, in ISMIR, 2019, pp. 644–651. [7] L. O. Rowe and G. Tzanetakis, "Curriculum learning for imbalanced classification in large vocabulary automatic chord recognition." in IS- MIR, 2021, pp. 586–593.

[7]. Choco: a chord corpus and a data transformation workflow for musical harmony knowledge graphs, by J. de Berardinis, A. Mero˜ no-Pe˜ nuela, A. Poltronieri, and V. Presutti, cientific Data, vol. 10, no. 1, p. 641, 2023.

[8]. Chord segmentation and recognition using em- trained hidden markov models, by A. Sheh and D. P. Ellis in International Symposium on Music Information Retrieval, 2003.

[9]. Curriculum learning for imbalanced classification in large vocabulary automatic chord recognition, by L. O. Rowe and G. Tzanetakis in IS- MIR, 2021, pp. 586–593.

[10]. Hmm- based approach for automatic chord detection using refined acoustic features, by Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono, and S. Sagayamain 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010, pp. 5518–5521.

[11]. Use of hidden markov models and factored language models for automatic chord recognition, by M. Khadkevich and M. Omologo in ISMIR, 2009, pp. 561–566.

[12]. Feature learning for chord recognition: The deep chroma extractor, by F. Korzeniowski and G. Widmer arXiv preprint arXiv:1612.05065, 2016.