Computational Analysis of Jazz Improvisation: Integrating CREPE Pitch Detection with Language Model Feedback

Ayden Kanter

Abstract

Jazz improvisation is inherently subjective, which makes it difficult for musicians to receive consistent, actionable, and quantitative assessments of their performances. This paper presents a method of analyzing monophonic jazz solos computationally through the combination of pitch estimation and large language model (LLM) reasoning. After transcribing the uploaded solo by using the Convolutional Representation for Pitch Estimation (CREPE) model, the note sequence is placed alongside the provided chord timeline. This alignment allows the LLM to generate structured feedback on phrase structure, harmonic alignment, and motif development, as well as provide actionable advice to improve. A prototype web application using this pipeline was evaluated by eight musicians of varying experience and skill. While experienced jazz soloists rated the feedback as highly actionable, beginning musicians often found it difficult to decipher the musical jargon, thus indicating the necessity of simple and easy-to-understand language. Currently, this framework does not currently assess dynamics, articulation, or other expressive characteristics. However, it demonstrates the potential of integrating pitch estimation and LLMs to generate structured feedback on jazz improvisations, providing a computational system for improvisational analysis that has future applications in AI-assisted music education.

Introduction

Jazz improvisation is difficult to quantify, and while computational approaches have explored improvisation generation, interaction, and evaluation, fewer systems focus on providing structured feedback on recorded human performances. Musicians often rely on private teachers, who would likely each give different feedback. Furthermore, instead of relying on quantitative evaluations, students depend on repetitive trial-and-error to find ideas that are pleasing to the ear, or they turn to previous professional jazz solos from which they learn note-by-note. In essence, the steps towards improvement in jazz improvisation are arduous and heavily rely on subjective feedback. While this subjectivity is crucial in ensuring that players develop their own "sound", quantitative evaluations can be particularly useful at beginning and intermediate stages, serving as foundations on which to build their knowledge. This paper introduces a multi-stage process, where a deep learning pitch extraction model converts frequencies into discrete notes, and then an LLM maps pitches alongside a chord timeline. This allows for a more accurate analysis of the musician's use of tension and resolution as well as the note choice.

While programmed MIDI notes are less difficult to represent due to their anticipated behavior, the spontaneity and unpredictability of jazz improvisations cause note extraction to be tedious. This paper utilizes a CREPE model, which is a deep convolutional neural network devised as an improved alternative to pYIN algorithms, which often fail to accurately identify notes at a high speed (Kim et al). CREPE is now available as a free-to-use Python module, which is how pitches were quantified in this research.

By using a chord structure, time signature, and tempo that is entered by the user, a chord timeline is created, which represents each chord along with its start and end time. Alongside the extracted pitches, the LLM can analyze motif development and note choice to be accurately analyzed. This tool is available as a web-based prototype.

The contributions of this paper are threefold: a method for transforming a recorded improvisation into quantitative representations of notes; a method of organizing notes alongside the respective chords over which they were played; and a way of generating structured feedback using LLMs. In total, this research provides an initial system to computationally analyze harmonic relationships and generate actionable feedback for jazz improvisation, while also acknowledging the limitations currently present in handling this complex musical form.

Related Work

Automatic Music Transcription Tools (AMTTs) have been prototyped in the past, but they have often been unsuccessful. Ethan Hein, a music professor at New York University, acknowledges the ability of AMTTs to accurately detect pitches, but also highlights the frequently incorrect rhythm detection and “nonsensical time signatures” often transcribed by these softwares (Hein). In addition, certain instruments have a timbre that is more conducive to transcription, such as the piano, while others (e.g. saxophone, flute) have shown more difficulty (Riley and Dixon).

When comparing Probabilistic YIN (pYIN) and CREPE algorithms, CREPE vastly outperforms the accuracy of pYIN at 50, 25, and 10 cent thresholds, with an 8.7% improvement under a 10 cent threshold (Kim et al.). This technological advancement is crucial for ensuring precision in transcribing jazz improvisations.

With regards to artificial intelligence, a paper published in *Transactions on Artificial Intelligence* analyzed its uses in music education, which included creating lesson plans, analyzing music performance, and providing personalized practice feedback (Yu et al. 39, 44). However, this research also acknowledged AI’s incapability of examining expressiveness in music, as well as the lack of transparency in how LLMs evaluate music.

AI has been applied to jazz for several decades. GenJam, short for genetic jammer, uses genetic algorithms to generate coherent jazz improvisations through evolutionary optimization (Biles 131). Band-Out-of-a-Box, commonly referred to as “BoB”, presents a software tool that extends this to interactive collaboration, where the system interactively trades short solos with the musician, adjusting its improvisation based on the user’s personal style and musical choices (Thom 133). Another influential system, The Continuator, created by François Pachet, explores interactive improvisation by learning a performer’s musical style in real time and generating stylistically consistent continuations (Pachet 333). These systems demonstrate that LLMs can model important aspects of jazz harmony and improvisational structure through the generation of music. Impro-Visor, an application developed by Robert Keller, further bridges the gap between AI, music education, and jazz improvisation, allowing users to generate solos through the use of language models while simultaneously learning about music theory and soloing techniques (Keller 333). Collectively, these systems established that LLMs can represent stylistic, harmonic, and interactive aspects of jazz improvisation.

Some researchers argue that the creativity involved with jazz improvisation is computable. Even though soloing is sometimes viewed as a purely spontaneous process, it often relies on harmonic, rhythmic, and stylistic constraints, which are necessary to comprehend if one wishes to learn to improvise. These elements of music, namely pitches, chords, and durations, are representable computationally (Johnson-Laird 420). Wiggins formalizes this notion, establishing computational creativity as a search process within structured constraints, where creative outputs emerge from systematic examination rather than unrestricted spontaneity (Wiggins 213-214). Building on this idea, research by Jordanous into computational creativity proposed an embeddings-based framework for representing emotional content in jazz improvisation. The resulting “emovectors” provide a quantitative representation of emotional expression in jazz solos, which the study hypothesizes may be a useful proxy for perceived creativity (Jordanous 5003). These studies demonstrate that computational methods can computationally analyze meaningful musical characteristics that have traditionally been regarded as subjective.

While past research has delved into modeling creativity, emotion, and AI-generated improvisation, less attention has been devoted to systems that analyze human improvisation to generate structured feedback for improvement. This paper builds upon advancements in pitch estimation and AI to evaluate musical characteristics such as harmonic alignment and motif development, providing structured feedback to assist musicians in improving their improvisational skills.

Prior research has investigated how computational methods can model aspects of jazz improvisation, including harmonic structure, stylistic interaction, emotional content, and creativity. However, this research focused on generation or mathematical representation rather than providing feedback on recorded improvisation. Research on AI in music education has identified that personalized feedback could have promising applications, while also mentioning limitations of AI in evaluating musical expressiveness (Yu et al. 39, 44). To build on these areas of research, this paper combines pitch estimation, symbolic harmonic representation, and language model feedback to evaluate jazz improvisations. Specifically, the proposed framework focuses on phrase structure, motif development, and harmonic alignment to generate structured feedback for musicians.

Methodology

To analyze jazz improvisation, this system uses CREPE to extract pitches from an uploaded recording. Then, the waveform is visualized, allowing the user to mark the time of the first measure of their solo.

CREPE predicts a pitch with a confidence level between 0.0 and 1.0. Predictions are discarded if below a certain confidence threshold. After confidence filtering, continuous pitch estimates are segmented into discrete note events. Each note is represented by its rounded MIDI pitch, onset time, and duration.

With respect to the chord timeline construction, the user inputs tempo, time signature, first-measure start time, and the sequence of chords. From this, the system is able to generate a list of chords and start times that repeat to cover the full audio duration if the recording lasts longer than one sequence.

Together, the note sequence and chord timeline provide a symbolic representation of the improvisation that enables harmonic and melodic relationships to be analyzed computationally. This representation enables harmonic analysis and the detection of tension and resolution.

The symbolic note and chord data are sent to an LLM, along with the first-measure start time, instrument, tempo, and time signature. The model is then prompted to give structured feedback that can be used to improve one's improvisation.

The prompt was crafted to organize feedback into four sections: phrase structure, motif development, harmonic alignment, and actionable feedback for improvement. By organizing this evaluation into discrete sections, it helped to ensure that the generated output was interpretable analysis, rather than vague assessments of ability, an issue that was underscored in previous research mentioned above (Yu et al. 44). The overall analysis pipeline is shown in Figure 1. Figure 2 illustrates example feedback generated for an intermediate piano solo over a 12-bar standard Bb jazz blues chord progression.

Figure 1. System analysis pipeline for converting audio recording into structured language model feedback.

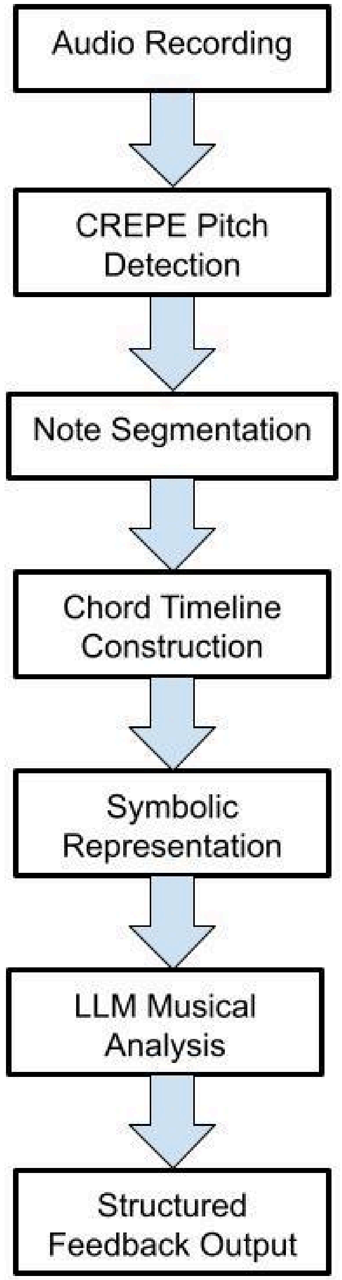

Figure 2. Example feedback produced by the proposed framework given to an intermediate jazz pianist.

Instrument: Piano
Tempo: 160
Time Signature: 4/4
First Measure: 1.695231s

Phrase Structure

The solo demonstrates a generally clear phrase structure, with groupings of notes that seem to delineate musical ideas. Phrases tend to be around 4-8 beats long, creating a conversational feel. However, some phrases feel a bit abrupt, lacking a strong sense of resolution or melodic completion.

Harmonic Alignment

Overall, the solo shows a decent understanding of the underlying harmony, but there are moments where the note choices don't fully align with the chord changes. Over the Bb7 chords, there's a tendency to stay within a narrow range of notes, sometimes neglecting the broader possibilities of the Bb7 scale (which includes alterations). During the F7 chord, the solo could benefit from incorporating more of the F7 altered scale to create more tension and release. The Eb7 chords are generally handled well, but occasionally, notes outside the Eb7 dominant scale create a dissonant clash without a clear resolution.

Motif Building

The solo exhibits some interesting motivic development, particularly the recurring use of the G-Ab-Bb pitch sequence. This creates a recognizable melodic thread throughout the solo. However, these motifs could be more consistently developed and varied to create a stronger sense of narrative. While there are moments of melodic contour, the solo sometimes lacks a clear direction or goal, resulting in a somewhat meandering feel. Exploring rhythmic variations of these motifs would also add depth and interest.

Actionable Feedback

The solo has a strong foundation and demonstrates a good ear for phrasing. The use of chromatic passing tones is effective in creating interest, but sometimes they lack a clear harmonic purpose. Here are some suggestions to elevate the improvisation:
1. **Expand Chord Scale Vocabulary:** When improvising over the Bb7 chord, consciously incorporate notes from the altered scale (b9, #9, #11, b13) to add color and tension. For example, try incorporating a Db (b9) or E (b13) over the Bb7.
2. **Motific Development with Rhythmic Variation:** Take the recurring G-Ab-Bb sequence and experiment with different rhythms and articulations. Try playing it as a triplet, a dotted rhythm, or with a syncopated feel. This will add rhythmic interest and reinforce the motif.
3. **Target Chord Tones on Strong Beats:** Ensure that important chord tones (root, 3rd, 5th, 7th) land on strong beats (1 and 3) to create a stronger harmonic foundation. This will help ground the improvisation and make it more cohesive.

Implementation

The implementation details detailed below correspond to the stages shown in Figure 2. CREPE is used to extract pitches from an uploaded recording in WAV/MP3 format. Then, audio is loaded using librosa and WaveSurfer.js is used to visualize the waveform. To ensure that processing time does not cause the website to timeout, total recording length is limited to 45 seconds.

Audio is slowed down using librosa.effects.time_stretch. CREPE is configured with a timestep of 10 ms. Frequencies are smoothed with a median filter, and predictions are discarded if below a confidence threshold of 0.80. Each note is stored as a JSON object in the format { "p": pitch, "s": start_time, "d": duration}: “p”, “s”, and “d” are used as abbreviations to reduce prompt size when delivering note data to the LLM. “p” represents the rounded MIDI pitch number. Frequencies with a duration under 0.05 seconds, equivalent to 16th notes at >300 beats per minute, are removed.

With the tempo, time signature, first-measure start time, and chord sequence, the system generates a list of chords in the JSON format { "c": chord, "s": start_time}. If two chords are in the same measure, the measure is divided evenly. Additionally, the chord progression repeats to cover the full audio duration if the recording lasts longer than one sequence.

The chosen model for this experiment is Gemma 3 12B, selected for its lightweight size and large 128k token context window, which is appropriate for cross-referencing improvised phrases at different moments in the solo.

The LLM is instructed to give 50% praise and 50% critique, no matter how adept the soloist may be. In ensuring that the evaluation is readable, MIDI note numbers are converted to note names (e.g. C, Eb) to be referenced in analysis, and all timestamps are rounded to the nearest tenth of a second.

Users may download the generated evaluation as a plain-text (.txt) file for future reference.

The implementation details mentioned above are available as an open-source GitHub repository at: github.com/Kanter1156/LLMJazzImprovEvaluation. The repository contains the necessary information needed to reproduce the audio transcription pipeline, chord timeline construction, and the LLM prompt.

Results

The prototype was evaluated by eight musicians with varying levels of jazz experience, ranging from less than three months to more than ten years. Each individual uploaded a recorded solo to the system and received personalized LLM feedback. Participants rated the actionability of the generated feedback on a five-point scale. All participants with more than three years of jazz experience rated the feedback 5/5 for actionability, while the remaining participants gave an average rating of 3.7/5. Despite receiving feedback generated using the same prompt structure, participants with greater jazz experience reported higher perceived actionability.

Six of the eight participants (75%) indicated that the system would be best suited for beginner-to-intermediate musicians. Participants with greater experience also identified limitations in the system's interpretation of advanced chromaticism, particularly when dissonant notes were intentionally used to create tension.

Due to the small sample size, these findings should be regarded as preliminary and are primarily useful for identifying trends for future evaluation.

Discussion

The results of this study suggest that combining pitch estimation with language models is a feasible approach for generating structured feedback on jazz improvisation. While previous research has largely focused on automatic music transcription or AI-generated improvisation, this work instead demonstrates how symbolic representations of recorded performances can be analyzed to provide structured and actionable feedback. By aligning extracted notes with a harmonic timeline, the system enables the LLM to reason about phrase structure, motif development, and harmonic alignment rather than evaluating isolated notes.

An important finding from the user evaluation was the difference in rated actionability between experienced and beginning musicians. Although the prompt structure was identical across participants, musicians with more than three years of jazz experience consistently rated the feedback as more actionable. This suggests that the usefulness of AI-generated musical feedback depends not only on the content and quality of analysis, but also on the user's ability to interpret musical terminology. Future AI-based educational systems should consider synchronizing the complexity of the feedback with the experience level of the performer.

The proposed framework also demonstrates the benefits of separating music transcription from musical reasoning. CREPE is responsible only for converting audio into a symbolic representation of notes, while the LLM analyzes these notes within their harmonic context. This system allows for improvements in either transcription algorithms or LLMs to be incorporated independently without requiring fundamental changes to the overall system. As both technologies continue to advance, the quality of automated improvisation analysis is likewise expected to improve.

Rather than replacing private instruction, this system is intended to function as a supplementary practice tool. Immediate, structured feedback allows musicians to reflect on their improvisations, identifying recurring musical patterns that may otherwise go unnoticed. In particular, beginner and intermediate musicians with limited access to individualized feedback may benefit from a system that provides structured and actionable advice to improve. All in all, these findings suggest that integrating pitch estimation with LLMs represents a promising direction for AI-assisted music education.

Limitations

As this system was designed to analyze single-note solos, it is not able to extract polyphonic sections. This poses a difficulty for pianists and guitarists, who frequently play multiple notes simultaneously to add a new dimension to their improvisations.

Furthermore, background noise can be problematic when extracting notes from recordings. Each tester was instructed to use a quiet backing track or metronome, preferably in headphones, to ensure that excess noise did not detract from solo analysis. This raises the question if CREPE will ever be able to accurately detect pitches from live recordings and “zero-in” on the timbre of a specific instrument. However, for now, it is limited to only function properly in quiet environments.

This model also relies on user chord input, which is often tedious when accounting for the quality of each chord. A standard jazz song is often 32 measures in length, and in many cases, can be longer, which makes inputting chords very inefficient.

Even though this tool solves the problem of improvisation analysis at a harmonic level, it still lacks dynamic / articulation feedback. As expressivity is a key component in jazz solos, this will be a necessary feature to incorporate in the future.

Lastly, there remain some limitations in analyzing advanced jazz solos. Adept jazz soloists often utilize the technique of “playing out”, where the musician intentionally plays dissonant notes that do not harmonically “fit” over the chords. The LLM has shown difficulty in analyzing this, due to the misinterpretation of wrong notes, rather than the true understanding of the intention behind each note choice. In a broader sense, the use of altered notes to create tension is conceptually understood, but the practical applications are not fully known by LLMs.

Future Work

To make chord inputting more efficient, this model would benefit from a tool such as the Python package ‘pyRealParser’, which allows chord progressions to be extracted from iRealPro links; this would enable a less tedious way to enter chords.

Moreover, this model could be refined by training an LLM on professional jazz solos. Whereas the current language model relies on generalized patterns on jazz from training, knowledge of common harmonic language and phrase structure would likely be very beneficial.

Lastly, it is important to develop a method to examine the expressivity of improvisation, which involves the dynamics and articulation of each note and phrase, thus indicating how the soloist builds energy throughout their performance. The librosa Python package can be used in the future to calculate rms_mean and rms_peak, which would delineate the crescendos within a solo. Measuring articulation becomes conceivable if note duration is compared to expected subdivision. For example, if the frequency lasts for a shorter time, it can be marked as staccato. An excess of short or long notes can more accurately inform the LLM evaluation of one’s solo.

Conclusion

Jazz improvisation remains complex and subjective. However, note extraction models in tandem with LLMs have the potential of generating actionable feedback on which soloists can improve. This work presents a multi-stage system, which integrates pitch detection and chord input alongside one another to allow the LLM to provide structured feedback. By aligning identified notes to be viewed next to the active chord, detailed analysis of harmonic content and phrase structure becomes possible.

The results support the information mentioned in the introduction, displaying that this system of LLM jazz solo analysis would be most appropriate for beginner to intermediate musicians. Nevertheless, there exists a disconnect between perceived actionability of feedback and the audience who would benefit most from this tool. It is important that the LLM is prompted to give actionable advice with easy-to-understand terms that all musicians can comprehend and apply.

However, multiple limitations remain. The system is still unable to process polyphonic solos; background noise can interfere with pitch detection; expressivity currently cannot be assessed; and advanced improvisation techniques using dissonance are misrecognized as “wrong” sequences of notes, rather than acknowledged for their ability to create tension.

Future work will acknowledge these limitations by incorporating dynamic and articulation analysis to measure expressivity. Moreover, a new LLM can be trained on past jazz solos to improve the accuracy of assessment, and knowledge of the timbres of various instruments can help separate important frequencies from background noise. Lastly, the user interface can be streamlined: chord input can be made less tedious by using pyRealParser, or a similar tool, to obtain chord progression data from a link, rather than from user-entered data. By integrating pitch detection and artificial intelligence, this model represents a step towards a more comprehensive and structured computational analysis of jazz improvisation.

Works Cited

Biles, John A. “GenJam: A Genetic Algorithm for Generating Jazz Solos.” *International Computer Music Conference (ICMC)*, July 1994, pp. 131–137.

Hein, Ethan. ““Humans Will Be Doing All the Serious Music Transcription for the Foreseeable Future”: Songscription Review.” *MusicRadar*, 8 Dec. 2025, musicradar.com/music-tech/humans-will-be-doing-all-the-serious-music-transcription-for-the-foreseeable-future-songscription-review. Accessed 1 May 2026.

Johnson-Laird, Philip N. “How Jazz Musicians Improvise.” *Music Perception*, vol. 19, no. 3, 2002, pp. 415–442, doi.org/10.1525/mp.2002.19.3.415.

Jordanous, Anna. “Emovectors: Assessing Emotional Content in Jazz Improvisations for Creativity Evaluation.” *2025 IEEE International Conference on Big Data*, 2025, pp. 5003–5006, doi.org/10.1109/BigData66926.2025.11402279.

Keller, Robert M., and David R. Morrison. “A Grammatical Approach to Automatic Improvisation.” *Proceedings of the Sound and Music Computing Conference*, July 2007, pp. 330–337.

Kim, Jong Wook, et al. “CREPE: A Convolutional Representation for Pitch Estimation.” *ArXiv*, 2018, arxiv.org/abs/1802.06182.

Pachet, François. “The Continuator: Musical Interaction with Style.” *Journal of New Music Research*, vol. 32, no. 3, 1 Sept. 2003, pp. 333–341, doi.org/10.1076/jnmr.32.3.333.16861.

Riley, Xavier, and Simon Dixon. “Reconstructing the Charlie Parker Omnibook Using an Audio-To-Score Automatic Transcription Pipeline.” *ArXiv*, 2024, arxiv.org/html/2405.16687v1.

Thom, Belinda. “Interactive Improvisational Music Companionship: A User-Modeling Approach.” *User Modeling and User-Adapted Interaction*, vol. 13, no. 1, Feb. 2003, pp. 133–177, doi.org/10.1023/a:1024014923940.

Wiggins, Geraint A. “Searching for Computational Creativity.” *New Generation Computing*, vol. 24, no. 3, Sept. 2006, pp. 209–222, doi.org/10.1007/bf03037332.

Yu, Guo, et al. “Recent Advances in Artificial Intelligence For Music Education.” *Transactions on Artificial Intelligence*, vol. 2, no. 1, 27 Feb. 2026, pp. 39–53, doi.org/10.53941/tai.2026.100004.