Analysis and Identification of Mistakes and Missed Notes in Piano Recordings Based on Fourier Transform

Xiyuan Li

Beijing 101 Middle School, Beijing, China

315274994@qq.com

Abstract. This study aims to explore the theoretical application of the Fourier transform in analyzing piano audio signals and to identify potential methods for detecting mistakes and missed notes. The Fourier transform is a powerful tool for revealing the frequency characteristics of audio signals, making it suitable for extracting frequency domain features from piano recordings, which can be used to compare actual performance with expected spectral features to detect inaccuracies.

Keywords: Fourier transform, piano audio signal, frequency analysis, mistake identification, missed note detection

1 Introduction

1.1 Research Background

The digital age and increased computational power have accelerated audio signal processing research. This field utilizes computational methods to analyze, manipulate, and modify musical signals for various applications. Specialized techniques harness musical elements like rhythm, harmony, timbre, and melody. The diversity and unique properties of musical signals underpin processing techniques and algorithms[1]. Signal processing techniques have revolutionized music generation and analysis.

Piano audio signals exhibit a broad frequency range, dynamics, and harmonic content, with complex temporal features. These signals differ from string sounds due to the driving point admittance effect[2], posing theoretical challenges like overlapping notes, inharmonicity, onset detection issues, harmonic masking, and nonlinearities. Advanced signal processing, like time-frequency analysis, can help overcome these challenges.

Fourier Transform is a mathematical tool used to analyze piano audio signals. It breaks down time-domain signals into frequencies, aiding in identifying dominant frequencies and distinguishing sounds. Commonly used in music signal processing, it converts signals from time to frequency domains, revealing spectral characteristics and harmonic structures. This approach is valuable in analyzing piano recordings, extracting features like spectral envelopes, and identifying mistakes or missed notes by comparing actual and expected spectral features[3].

S. Kaki et al. (eds.), *Proceedings of the 2024 6th International Conference on Literature, Art and Human Development (ICLAHD 2024)*, Advances in Social Science, Education and Humanities Research 885,
https://doi.org/10.2991/978-2-38476-319-1_27

1.2 Research Objective

The theoretical goal of this study is to explore how the Fourier transform can be used to analyze piano audio signals and theoretically identify mistakes and missed notes.

1.3 Literature Review

Fourier Transform analyzes piano audio into frequencies, revealing acoustic properties. Hong[3] used it in MATLAB to visualize piano harmonics. Lenssen & Needell[4] developed DFT & STFT for chord analysis, capturing frequency changes over time. Zhao et al.[5] improved scale recognition with FFT-based matched filtering. Eggink et al.[2] employed FFT for spectral analysis, focusing on low-frequency harmonics. Joji et al.[6] created a Python algorithm using FFT for piano note detection, reducing noise and assessing performance accuracy. The algorithm assesses the frequency and rhythm of the piano audio signals to determine a correctness score for the piano play. By converting the piano audio signal from the time domain to the frequency domain, the FFT algorithm can be used to accurately identify individual piano notes.

2 Theoretical Foundation

2.1 Mathematical Principles of Fourier Transform

The Fourier transform is a mathematical tool used in audio signal processing. It converts the time-domain input signal into a frequency-domain representation for analysis. The connection between frequency and time domains is easily observable in a vibrating string. This vibrational motion can be modeled using differential equations with a function $f(t)$, where t is time. This is time-domain representation of the string motion. Audio signals are recorded in the time domain.

If a piano string is plucked in such a way that it vibrates only at the fundamental harmonic, the standing waves of the vibrating string represent the time domain by a single sinusoid of frequency v_0. Thus, the frequency representation function of a vibrating string is $F\ (f_0)$. In real-world, systems have more than one frequency. This can be accounted for by constructing the frequency-domain representation of the frequency domain by an infinite series of the harmonics representing the motion [4]. If $f(t)$ is a periodic signal with period T, the fundamental frequency of the signal is given by:

$$\boldsymbol{f_0 = \frac{1}{T}} \tag{1}$$

The corresponding angular frequency is given by:

$$\boldsymbol{\omega_0 = 2\pi f_0} \tag{2}$$

The Fourier series expresses the periodic signal $f(t)$ as the sum of harmonically related sinusoids or complex exponentials. The general form of the Fourier series is expressed using complex exponentials:

$$\boldsymbol{f(t) = \sum_{-\infty}^{\infty} c_n\, e^{jn\omega_0 t}} \tag{3}$$

In equation (3), c_n represents following complex Fourier coefficients:

$$\boldsymbol{c_n = \frac{1}{T}\int_{T=0}^{T=\infty} f(t)e^{-jn\omega_0 t}dt} \tag{4}$$

The Fourier transform stems from the Fourier series, and is used to obtain the frequency-domain representation of a time-domain function. The inverse of the Fourier transform returns the time-domain function from a frequency-domain function.

2.1.1 The Continuous Fourier Transform.

The relationship between angular frequency (ω_k) and ordinary frequency (v) is defined by the equation:

$$\boldsymbol{\omega_k \equiv 2\pi k v} \tag{5}$$

Then, the Fourier transform defines the relationship between the time-domain function f and its corresponding frequency-domain function F as follows:

$$\boldsymbol{F(\omega_k) \equiv \int_{-\infty}^{\infty} f(t)e^{-2\pi i k t}dt, \qquad k \in (-\infty, \infty)} \tag{6}$$

The above complex exponential has sinusoidal components within it: $e^{i\omega t} = \cos(\omega t) + i sin(\omega t)$.

We can derive the inverse Fourier transform from (6):

$$\boldsymbol{f(t) = \int_{-\infty}^{\infty} F(\omega k)e^{2\pi i k t}dk, \qquad k \in (-\infty, \infty)} \tag{7}$$

Digital applications of the continuous Fourier transform require the discrete form, known as the discrete Fourier transform (DFT).

2.1.2 The discrete Fourier transform (DFT).

The DFT for vector $f \in \mathbb{C}^N$ is defined by:

$$\boldsymbol{F_k \equiv \sum_{n=-\frac{N}{2}+1}^{\frac{N}{2}} f_n e^{-i2\pi nk/N} \qquad k = 0, 1, \ldots, N-1} \tag{8}$$

In equation (2) above, f_n denotes the n^{th} entry of the vector f. The inverse of the DFT is:

$$\boldsymbol{f_k = \sum_{n=-\frac{N}{2}+1}^{\frac{N}{2}} F_n e^{i2\pi nk/N} \qquad k = 0, 1, \ldots, N-1} \tag{9}$$

The inverse Fourier transform decomposes the audio signal into the elements of its frequency in a way that enables easy extraction of spectral data and perception of individual notes by the human ear [4]. For chord detection and audio processing, sinusoids are defined as a function of the form:

$$\boldsymbol{x(t) = Asin\ (2\pi vt + \emptyset)} \tag{10}$$

Where:
A – amplitude
v – radian frequency (rad/sec)
2πv – frequency (Hz)
t – time (s)
∅ - initial phase (radians)
2πvt + ∅ - instantaneous phase (radians)

Fourier transforms are based on the complex properties of sinusoids which are defined by Euler's identities:

$$\boldsymbol{e^{i\theta} = \cos(\theta) + i \sin\ (\theta)} \tag{11}$$

$$\boldsymbol{e^{\pm i2\pi vx} = \cos(2\pi vx) \pm i\ sin(2\pi vx)} \tag{12}$$

Equation (12) is the form most relevant to audio signal processing.

2.2 Spectral Characteristics of Piano Audio Signals

The harmonic structure and fundamental frequency (f_0) of piano notes are unique and complex due to their harmonic content. The fundamental frequency is the lowest frequency of the audio signal corresponding to the pitch of each note played. In addition, each note produces a series of harmonics (overtones), represented by the integer multiples of the fundamental frequency ($2f_0$, $3f_0$, $4f_0$, $5f_0$, etc.). The inharmonicity of piano strings, caused by string stiffness, gives the piano its unique "bright" sound. Each note's fundamental frequency and harmonic structure serve as identifying features. Inharmonicity algorithms, when modeled accurately, can identify notes by considering deviations from perfect harmonic intervals.

The amplitude envelope and temporal characteristics of piano audio signals reflect their unique identification features. The sharp attack, decay rate (influenced by pitch and physical properties), sustain and release phases offer distinguishing temporal traits that can aid in identifying and distinguishing piano notes. The sustain phase duration also informs note sequencing or holding[7].

Piano audio signals exhibit a rich spectral content, with harmonics decreasing in amplitude as frequency increases. Spectral peaks match known harmonic series for note identification. Spectral balance affects brightness and warmth. Higher notes have fewer harmonics than lower notes, making them distinguishable. Dynamic range and playing force influence spectral characteristics, allowing for note differentiation.

Spectral characteristics like string and soundboard resonances, sympathetic vibrations, and mechanical noise offer nuanced information for identifying and distinguishing piano notes. Resonances and sympathetic vibrations create subtle spectral variations, while mechanical noises aid in confirming note onset.

3 Theoretical Analysis

3.1 Theoretical Application of The Fourier Transforms to Piano Audio Signals

Piano audio signals are sampled digitally, and, therefore, DFT is the most common form of Fourier transform used:

$$\boldsymbol{X}(\boldsymbol{k}) = \sum_{n=0}^{N-1} \boldsymbol{x}(\boldsymbol{n})\boldsymbol{e}^{-j\frac{2\pi kn}{N}} \quad (13)$$

Where:

N – number of samples

k – the frequency components

When applying the Fourier transform to piano audio signals, we obtain the frequency, amplitude, and phase information of the signals. These components define the characteristics of the piano note. The frequency of the signal represents the pitches and associated harmonics. The intensity of each frequency component is represented by amplitude, which influences the loudness and timbre perceived by a listener. Phase is defined as the initial angle of each sinusoidal component at the start of the audio signal. Although less perceptible by the human ear, phase angle shapes the waveform of a sound wave. The DFT can be used to analyze Fourier transform results of different piano notes and their frequency domain representations:

Case 1: Low A (A2) piano note.

The low A (A2) note has a fundamental frequency (f_0) of ~ 110 Hz. The note generates a series of harmonics at integer multiples of f_0:

1st harmonic is f_0 = 110 Hz; 2nd harmonic is 2×110 Hz = 220 Hz; 3rd harmonic is 3×110 Hz = 330 Hz. And so on.

The 1st harmonic is the fundamental component that gives a peak at 110 Hz and represents the main pitch of the note. The 2nd and 3rd harmonics give peaks at frequencies of 220 Hz and 330 Hz, respectively. The Fourier transform of these notes reveals strong lower harmonics that give a richer and warmer sound. The amplitude spectrum tapers off gradually, with the first few harmonics having strong peaks.

Case 2: Middle C (C4) piano note.

The middle C4 note has a fundamental frequency of ~ 261.63 Hz. The note generates a series of harmonics at integer multiples of f_0:

1st harmonic is f_0 = 261.63 Hz; 2nd harmonic is 2×261.63 Hz = 523.26 Hz; 3rd harmonic is 3×261.63 Hz = 784.89 Hz. And so on.

The 1st harmonic is the fundamental component that gives a peak at 261.63 Hz and represents the main pitch of the note. The 2nd and 3rd harmonics give peaks at frequencies of 523.26 Hz and 784.89 Hz, respectively. As the frequency of these harmonics decreases with the increasing integer multiples, their amplitudes decrease, shaping the timbre.

Case 3: High A (A5) piano note.

The fundamental frequency of the A(A5) piano note is 880 Hz. This high frequency means that there are fewer harmonics within the audible range.

1st harmonic is f_0 = 880 Hz; 2nd harmonic is 2×261.63 Hz = 1760 Hz; 3rd harmonic is 3×261.63 Hz = 2640 Hz. And so on.

Higher harmonics typically have lower amplitudes, giving a thinner sound compared to lower notes.

3.1.1 Deriving the Frequency Domain Representation.

The frequency domain representations for these notes can be derived by applying the Fourier Transform.

Case 1: Low A (A2) piano tone.

Fundamental frequency, f_0~110 Hz. The time domain representation of A(A2) note can be modeled as the summation of the fundamental frequency and its harmonics:

$$x_{A2}(t) = A_0 \cos(2\pi f_0 t + \emptyset_0) + \sum_{n-1}^{N} A_n \cos(2\pi n f_0 t + \emptyset_n) \tag{14}$$

Considering the first three harmonics:

$$x_{A2}(t) = A_0 \cos(2\pi.110t + \emptyset_0) + A_1 \cos(2\pi.220t + \emptyset_1) + A_2 \cos(2\pi.330t + \emptyset_2) \tag{15}$$

This is the time-domain representation of the low A (A2) piano note.

The frequency domain representation is obtained by applying the Fourier transform:

$$X_{A2}(F) = A_0 e^{j\emptyset 0}\delta(f - 110) + A_1 e^{j\emptyset 0}\delta(f - 220) + A_2 e^{j\emptyset 0}\delta(f - 330) \tag{16}$$

At 110 Hz, amplitude = A_0 and phase = $\emptyset_0$; At 220 Hz, amplitude = A_1 and phase = $\emptyset_1$; At 330 Hz, amplitude = A_2 and phase = $\emptyset_2$

Case 2: Middle C (C4) piano note.

The time-domain representation of the note can be modeled as:

$$x_{C4}(t) = A_0 \cos(2\pi f_0 t + \emptyset_0) + \sum_{n-1}^{N} A_n \cos(2\pi n f_0 t + \emptyset_n) \tag{17}$$

Considering the first three harmonics:

$$x_{C4}(t) = A_0 \cos(2\pi.261.63t + \emptyset_0) + A_1 \cos(2\pi.523.26t + \emptyset_1) + A_2 \cos(2\pi.784.89t + \emptyset_2) \tag{18}$$

Applying the Fourier transform:

$$X_{C4}(F) = A_0 e^{j\emptyset 0}\delta(f - 261.63) + A_1 e^{j\emptyset 0}\delta(f - 523.26) + A_2 e^{j\emptyset 0}\delta(f - 784.89) \tag{19}$$

At 261.3 Hz, amplitude = A_0 and phase = $\emptyset_0$; At 523.26 Hz, amplitude = A_1 and phase = $\emptyset_1$; At 784.89 Hz, amplitude = A_2 and phase = $\emptyset_2$

Case 3: High A (A5) piano note.

The time-domain representation of the note can be modeled as:

$$x_{A5}(t) = A_0 \cos(2\pi f_0 t + \emptyset_0) + \sum_{n-1}^{N} A_n \cos(2\pi n f_0 t + \emptyset_n) \tag{20}$$

Considering the first three harmonics:

$$x_{A5}(t) = A_0 \cos(2\pi.880t + \emptyset_0) + A_1 \cos(2\pi.1760t + \emptyset_1) + A_2 \cos(2\pi.2640t + \emptyset_2) \quad (21)$$

Applying the Fourier transform:

$$X_{A5}(F) = A_0 e^{j\emptyset 0}\delta(f - 880) + A_1 e^{j\emptyset 0}\delta(f - 1760) + A_2 e^{j\emptyset 0}\delta(f - 2640) \quad (22)$$

At 880 Hz, amplitude = A_0 and phase = $\emptyset_0$. At 1760 Hz, amplitude = A_1 and phase = $\emptyset_1$. At 2640 Hz, amplitude = A_2 and phase = $\emptyset_2$.

The Fourier transform decomposes piano audio signals into frequency components, showing the harmonic structure properties of each note. In each of the above three cases, the representation of each note is defined by the frequencies of the delta functions, the amplitudes, and the phases. Collectively, these audio signal properties the harmonic richness, waveform shape, and the perceived loudness of the notes. The unique spectral patterns of the notes allow for their distinction. Figure 1 shows the plots of the magnitude of the frequency components, where the peaks of fundamental frequency and harmonics of notes A2, C4, and A5 can be visualized.

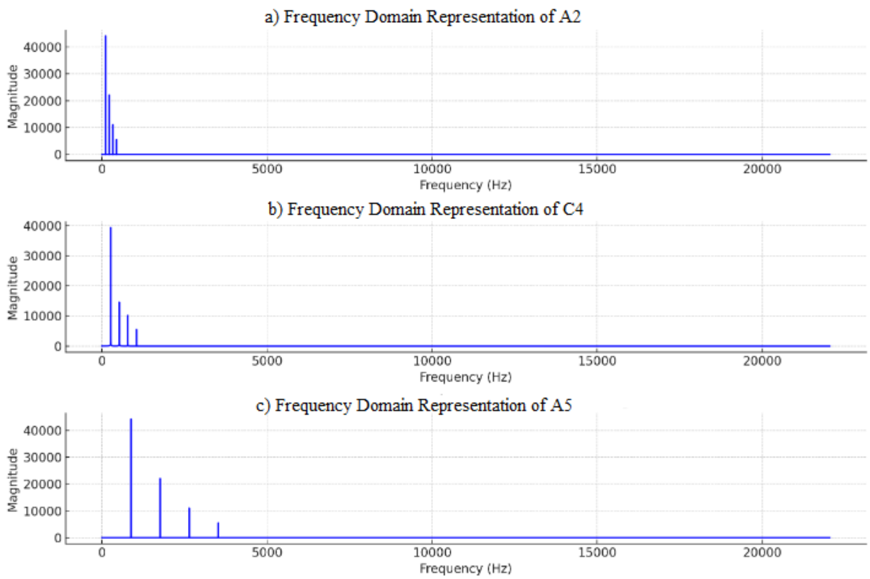

Fig. 1. Frequency domain representations of notes (a) A2, (b) C4, and (c) A5 based on DFT analysis of the notes. The DFT is computed using the FFT algorithm, after applying a Hamming window to the audio signal of a piano note.

3.2 Theoretical Detection of Mistakes and Missed Notes

Fourier Transform Analysis (FFT) converts piano audio from time to frequency domain, segmenting it into precise timeframes. It identifies notes by matching dominant frequencies to predefined spectra of a well-tuned piano. FFT can detect mistakes like extra or wrong notes by analyzing unexpected frequencies and missed notes by identifying absent expected frequencies. This conversion and analysis enable precise note identification and mistake detection.

3.3 Model Derivation and Validation

Establishing a theoretical model for detecting mistakes and missed notes in piano audio signals requires a good understanding of the characteristics of the signals and how the Fourier transform can be used to analyze them. The model is established to detect discrepancies between the actual and expected signals, as an indicator of mistakes or missed notes.

A piano audio signal is a time-domain signal comprising multiple harmonic frequencies that correspond to the notes played. Each of the notes can be modeled as a sum of sinusoidal functions.

Let $x(t)$ represent a piano audio signal and i represent the note played. Then,

$$x_i(t) = A_i . \sin\left(2\pi f_i t + \emptyset_i\right) \tag{23}$$

Where:

A_i – amplitude of the i-th note; f_i – fundamental frequency of the i-th note; $\emptyset_i$ – phase of the i-th note

The sum of individual notes represents the overall piano signal x(t):

$$x(t) = \sum_{i=1}^{N} x_i(t) = \sum_{i=1}^{N} A_i . \sin\left(2\pi f_i t + \emptyset_i\right) \tag{24}$$

When the Fourier transform $X(f)$ of the signal $x(t)$ is applied:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t} dt \tag{25}$$

Based on the linearity of the Fourier transform:

$$X(f) = \sum_{i=1}^{N} A_i \left[\int_{-\infty}^{\infty} \sin(2\pi f_i t + \emptyset_i)\, e^{-j2\pi f t} dt\right] \tag{26}$$

Using Euler's formula (27):

$$\sin(2\pi f_i t + \emptyset_i) = \frac{e^{j(2\pi f_i t + \emptyset_i)} - e^{j(2\pi f_i t + \emptyset_i)}}{2j} \tag{27}$$

We get:

$$X(f) = \sum_{i=1}^{N} \frac{A_i}{2j} \left[\int_{-\infty}^{\infty} \left(e^{j(2\pi f_i t + \emptyset_i)} - e^{j(2\pi f_i t + \emptyset_i)}\right) e^{-j2\pi f t} dt\right] \tag{28}$$

Each component of the Fourier transform will show peaks at frequencies that correspond to the notes played. The Fourier transform algorithm can be simplified and optimized to reduce computational complexity, especially when dealing with high-resolution and long-term recordings, and improve the efficiency and practicality of the algorithm. The simplified version of equation (28) is:

$$X(f) = \sum_{i=1}^{N} \frac{A_i}{2j} \left[\int_{-\infty}^{\infty} (2j sin(2\pi f_i t + \varphi_i)) e^{-j2\pi f t} dt \right] \quad (29)$$

The FT can be optimized by using FFT to compute the DFT. This allows for the analysis of large windows of audio signal much faster when dealing with high-resolution and long-term recordings. For a discretized signal x[n] sampled at N points, the FFT is expressed as:

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi kn}/N \quad (30)$$

In equation (30), k represents the frequency index. Other optimization techniques for the FT include windowing for Short-Time Fourier Transform (STFT), downsampling or decimating the signal for long-term recording, using multiresolution analysis with wavelet transform, and pruning insignificant frequencies and associated harmonics.

Mistakes or missed notes can be detected by comparing the Fourier transform of the actual signal and that of the expected signal. The difference function can be defined as follows:

$$D(f) = \left| X_{actual}(f) - X_{expected}(f) \right| \quad (31)$$

Where:

D(f) – the difference function; $X_{actual}(f)$ – Fourier transform of the actual signal; $X_{expected}(f)$ – Fourier transform of expected signal.

If $D(f) \neq 0$ at a specific frequency, then the difference is explained by discrepancies between the expected and actual signals, which can either be due to mistakes or missed notes. The degree of mistake or missed note can be determined by calculating the energy difference shown in equation (32):

$$E = \int_{0}^{\infty} \left| X_{actual}(f) - X_{expected}(f) \right|^2 df \quad (32)$$

The higher the value of E, the greater the deviation from the expected performance. Stability considerations for this model include noise sensitivity, harmonic overlaps, and model complexity.

The model applies to multiple instruments in controlled environments and live performances but may struggle with isolating piano signals amidst others. Source separation techniques can help. It excels in low-noise, high-quality recordings[8]. Live performances' acoustics and noise may hinder stability, but noise reduction and adaptive filtering can help. Validation through diverse testing conditions and performance metrics like accuracy, precision, and confusion matrix is crucial. Comparing with existing models assesses statistical significance.

4 Theoretical Discussion

4.1 Limitations of Theoretical Analysis

Some potential limitations of using the Fourier transform in piano signal analysis have been mentioned in section 3.3. They include noise interference, signal non-linearity, harmonic overlaps, and model complexity. The proposed theoretical model may be sensitive to noise accompanying the audio signal. Noise can introduce unexpected frequencies, which may lead to incorrect detection of mistakes. Filtering techniques such as band-pass filtering can be used to improve stability by isolating the fundamental frequencies. Signal non-linearity causes distortion so that the output waveform is different from the input waveform. The harmonics in piano notes may also overlap in the Fourier spectrum. This phenomenon may complicate the detection process, especially when the frequency of the notes falls within a small range. This problem can be mitigated using techniques such as wavelet transforms or Short-Time Fourier Transform (STFT), which provide time-frequency localization. The stability of the model also depends on its complexity and accuracy. If it does not accurately capture the audio characteristics, then it can be difficult to detect mistakes or missed notes.

4.2 Practical Applications of Theoretical Results

The theoretical results of Fourier transform analysis can be applied to practical piano teaching and performance evaluation. Successful models for accurate note identification can be developed by utilizing FFT capabilities. The theoretical results also highlight the potential to integrate signal processing into the student learning process, for example, through real-time feedback to students and real-time assessment of piano performances. This may potentially promote automated student assessment [6]. Beyond the academic applications, the theoretical model also extends into real-world applications. Understanding the harmonic composition and how it contributes to timbre and other audio characteristics can inform the design of better musical pianos or the refinement of existing ones to achieve high-output audio qualities. The FFT analysis is directly applicable in the development of audio signal processing software with effects, synthesis, and noise reduction capabilities. Sound engineers and music producers can leverage theoretical models to effectively manipulate recordings, to achieve the desired qualities of music [3].

5 Conclusion

Fourier Transform analyzes piano audio signals to identify notes by converting them from time to frequency domain. It aids in detecting mistakes and missed notes, providing insights into music recording and performance accuracy.

The application of the Fourier transform in audio signal processing provides a foundation for future research in music theory. Future research directions include the development of more complex audio signal processing methods for future theoretical exploration and potential application in the realm of the design of musical instruments, sound engineering, and music production. Promising technologies include machine learning models such as deep learning, and advanced audio feature extraction techniques such as constant-Q transforms. Combining deep learning techniques with Fourier transform to improve the accuracy and reliability of error detection in piano audio signals. Adaptive learning systems could also potentially provide personalized feedback and ensure high-precision audio. Another possibility for future research is the standardization of data sets and performance evaluation metrics for the detection of mistakes and missed notes in piano audio recordings.

References

1. Jamshidi, F., Marghitu, D., & Chapman, R. (2021, July). Developing an online music teaching and practicing platform via machine learning: a review paper. In International conference on human-computer interaction (pp. 95-108). Cham: Springer International Publishing.
2. Dalmont, J. P. 2021. Piano bass strings with reduced inharmonicity: theory and experiments. Acta Acustica, 5, 9.
3. X. Hong 2024, Analyzing Musical Tones with Fourier Transformation," Theoretical and Natural Science, vol. 41, pp. 1–5.
4. N. Lenssen and D. Needell 2014, An Introduction to Fourier Analysis with Applications to Music," jhummath, vol. 4, no. 1, pp. 72–91.
5. Z. Zhao, A. Zhao, J. Hui, B. Hou, R. Sotudeh, and F. Niu 2017, A Frequency-Domain Adaptive Matched Filter for Active Sonar Detection, Sensors (Basel), vol. 17, no. 7, p. 1565.
6. N. M. Joji, P. Namjoshi, J. Joeboy, and S. Khandare 2023, Sounding the Keys: A Comprehensive FFT-Based Approach for Piano Note Detection, vol. 10, no. 12, pp. 195-198.
7. D. Modi 2021, Harmonium and Piano Frequencies, vol. 08, no. 02, pp. 808–817.
8. Kob, M., Amengual Garí, S. V., & Schärer Kalkandjiev, Z. (2020). Room effect on musicians' performance. The Technology of Binaural Understanding, 223-249.