Signal Processing Approaches to Musical Tuning System Detection in Audio

Signal Processing Approaches to Musical Tuning System Detection in Audio

Ethan B. Cobb

A thesis presented to the
Department of Applied Mathematics
in partial fulfillment of the honors requirements
for the degree of Bachelor of Arts

Faculty Advisors: Michael P. Brenner, Flavio P. Calmon

Harvard University School of Engineering and Applied Sciences
Cambridge, MA
May 2022

Contents

List of Figures

List of Tables

Abstract

There are three main intonation systems musicians use—just, equal temperament, and Pythagorean. The use of any one of these particular intonation systems depends on a number of factors, most importantly musical and harmonic context. While work has been done in analyzing intonation tendencies in vocal and instrumental performance as well as temperament estimation of fixed-pitch instruments such as the harpsichord, no work has been done in detecting instances of intonation systems in audio from a signal processing perspective. For my thesis, I propose several algorithms and signal processing techniques to detect and identify intonation systems in recordings of the Third Bach Cello Suite Prelude, which I chose for its musical and harmonic complexity. I first obtained timestamps and fundamental frequency estimates utilizing CREPE, a pretrained deep convolutional neural network used for monophonic frequency estimation. Then, I implemented an algorithm that obtains an estimated classification for each note by calculating an associated probability for each tuning system and applying a centered moving average. Finally, a set of sequences of intonation systems were identified by splitting the intonation system-labeled time series at instances where consecutive labels differ and merging any consecutive equivalently-labeled sequences within a presupplied time threshold of each other. Clear overall trends in the use of just and Pythagorean intonation emerged from running the model on twelve different recordings of the Prelude. This research offers musicians a tool for understanding and assessing their intonation by providing an objective measure of intonation. It also provides a way to gain insight into the history of intonation in performance.

A link to the Github repo can be found at `https://github.com/ecobb/thesis`

Acknowledgements

This thesis would not have been possible without the support of several people. First and foremost, I would like to thank my thesis advisors, Professor Michael Brenner and Professor Flavio Calmon. Professor Brenner generously offered to advise me on this idea that I had been wondering about for years, and I am grateful for his unconditional support and for serving as my concentration advisor during my time at Harvard. He was always willing to meet over zoom at ungodly hours and I will never forget discussing with him advanced technical concepts as he tended to his baby. Taking ES-156 with Professor Calmon was one of the greatest course experiences I've had at Harvard - I knew that if I were to write a thesis, I would want it to be with him. His humility and kindness as a person in addition to his technical knowledge is unparalleled.

This thesis especially would not have been achievable without the help of Jeff Li, Google software engineer and previous student of my cello teacher, Richard Aaron. Jeff's technical advising and vast knowledge of intonation systems were critical in combining my mathematical and musical ideas into a coherent problem formulation which granted me the results I hope to achieve. I would like to thank Hsiang Hsu as well for meeting with me several times throughout the semester offering fantastic guidance not only concerning the technical details of the algorithms but about the research process in general. I am grateful to Jimmy Qin for helping me solidify my technical approach by steering me in the direction of a probabilistic approach and to Nathan Le for our numerous discussions about this fascinating subject.

Finally, I would like to thank my friends and family for their support during this journey. I am forever grateful to my parents for encouraging me and providing the means to pursue a liberal arts education, let alone at such an amazing institution that is Harvard. I would like to particularly thank Willie Swett for unintentionally motivating me to continue working diligently during our spring break HRO residency in Cremona. I am especially grateful to my lifelong friend and mentor Sue Poliacik for taking me under her wing when I first explored this idea as part of my senior project at Riverdale Country School. When I left Riverdale, I knew I wanted to circle back to this topic someday when I had acquired greater mathematical and computational knowledge. I am proud to say that I have finally done so.

Chapter 1

Introduction

1.1 Background

The subject of intonation and tuning systems in classical music is of central importance. It is most often discussed from a qualitative perspective, lending itself to individual subjective impressions of how “in tune” or “out of tune” a note, series of notes, or musical passage sounds. As such, the placement of notes on a frequency level (beyond many musical context-related cues) is largely dependent on individual pitch preferences. There is a general lack of understanding or even awareness about the mathematical and physical foundations of intonation and tuning in classical music. Whether musicians consciously decide or not, they typically use a combination of three main tuning systems: just intonation, Pythagorean intonation, and equal temperament [1].

1.1.1 History of Tuning Systems and Harmony

Theories of temperament and intonation have roots in Ancient Greece. The Pythagorean tradition contributed greatly to notions of musical intervals and consonance and noted that when two strings (or other “sounding bodies”) were played simultaneously and their lengths were in proportion to one another, they would produce harmonious tones that would correspond to particular musical intervals and sound pleasing to the ear [2]. Equivalently, the musical intervals we would deem consonant or pleasing would be those intervals that correspond to small integer ratios of the form $\frac{n+1}{n}$ with $n \in \mathbb{Z}$:

Musical Interval	**Ratio**
Unison	1:1
Octave	2:1
Fifth	3:2
Fourth	4:3

Table 1.1: Pythagorean Ratios

Thus for Pythagoreans, musical harmony and mathematics were two sides of the same coin and musical harmony was a demonstration of simple mathematics. Correspondingly, the mathematical laws of the universe manifest in the palpable phenomenon that is musical harmony. This correspondence between the metaphysical and the palpable is the basis of the Pythagorean notion of the "harmony of the spheres" [2].

Another contributor to the Greek tradition that greatly impacted notions of harmony and music theory was Aristoxenus, a student of Pythagoras.[1] Aristoxenus did not dispute the logic of Pythagoras' mathematical formulae or metaphysical notions but asserted that music could not be reduced to a rational mathematical framework or a manifestation of the cosmic harmony. To do so would contradict music's true role as an inherently *human* phenomenon. Numerical relationships may exist among musical tones but it is the human experience of listening and perception that gives meaning to music. According to Aristoxenus, "The mere sense-discrimination of magnitudes is no part of the general comprehension of music...Mere knowledge of magnitudes does not enlighten one as to the functions of the tetrachords, or of the notes, or of the differences of the genera, or, briefly, the differences of simple and

[1] Aristoxenus is one of the first music theorists from whom we actually have writings. We have little evidence that many of the mathematical and musical contributions attributed to Pythagoras were indeed his.

compound intervals, or the distinction between modulating and non-modulating scales, or the modes of melodic construction,or indeed anything else of the kind" [3].

While largely not concerned with notions of tuning, Aristoxenus did contest Pythagorean ratios and the concept of the division of the octave. He claimed that six tones divided the octave; this construction did not agree with Pythagorean ratios ($\left(\frac{9}{8}\right)^6 \neq 2$) [4]. Pythagoreans and Aristoxenians also disagreed about whether a whole tone could be divided into two equal semitones. Pythagoreans said it could not since $\sqrt{\frac{9}{8}}$ is irrational while Aristoxenes believed that it could be divided into various fractional divisions, which was a kind of foreshadowing of equal temperament [4].

A natural flaw in the Pythagorean system results from the fact that if one were to ascend 12 perfect fifths from a starting note, one should end up at the same note compared to if one were to ascend by seven octaves. This is of course mathematically impossible:

$$\left(\frac{3}{2}\right)^{12} \neq 2^7$$

The difference between these two theoretical values is called the Pythagorean comma. As I will explain in greater mathematical detail in section 1.2.2, when the Pythagorean system is used to construct a 12-tone scale, it inevitably creates an extremely harsh sounding interval between the second step of the scale and the sixth note of the scale, called a "wolf" interval. For instance, in the key of D, this interval would correspond to the diminished sixth interval between Eb and G#, almost a quarter of a semitone flatter than the just intonation $\frac{3}{2}$ ratio. While it has always posed an audible dissonance, the issue was mostly ignored for many centuries since it did not affect musical practice significantly; musical harmony in western music was not introduced until the middle of the medieval era in the 12th century. Perfect intervals such as octaves, fourths, and fifths were the primary intervals used for their inherent purity and the Pythagorean system posed no conflict.[2]

[2] Other musical cultures followed different trajectories in their notions of harmony, consonance, and dissonance; the Pythagorean comma was widely known, for instance in Chinese music theory.

Debates surrounding the quality of major thirds began during the Renaissance period [5]. The Pythagorean major third with a ratio of $\frac{81}{64}$ is higher than its just equivalent of $\frac{5}{4}$ (to be discussed in section 1.2.1) and this difference is called the "syntonic comma." This gave rise to different meantone temperament systems where the "comma" difference was split up between intervals in various ways. It was not until the 19th century when upright pianos became mass-produced and were tuned by professional tuners that 12-tone equal temperament (12-TET) was widely adopted.

1.2 Mathematical Definitions of Tuning Systems

1.2.1 Just Intonation

Just or harmonic intonation is the system of tuning based on the physical phenomenon known as the harmonic series, a sequence of notes generated from a fundamental frequency in which the frequency ratios are all whole integer ratios. This is shown by the figure below:

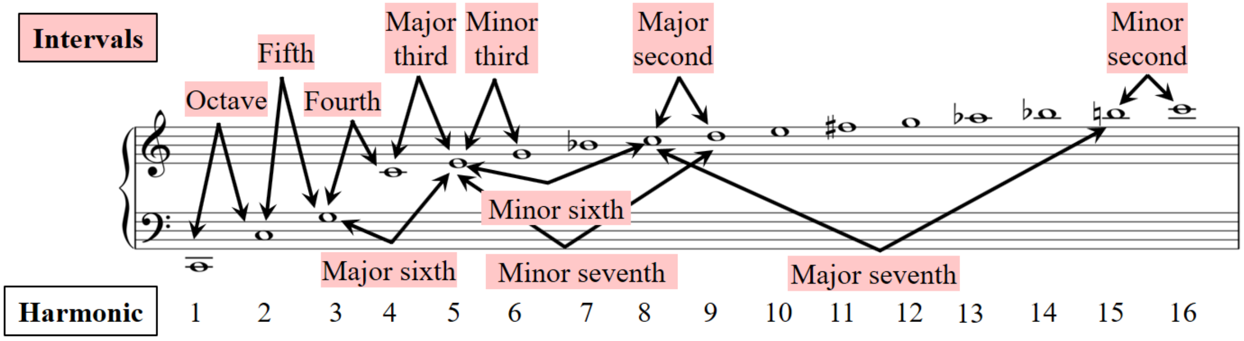


Figure 5.4 from [Müller, FMP, Springer 2015]

Figure 1.1: Example of harmonic series created from starting note C2

Just intonation in string playing is often used in the tuning of chords and double stops (two notes played at the same time). The perfect intervals such as the fourth and fifth that result from the 4/3 and 3/2 ratios, in relation to the fundamental frequency, contain a pure, resonant sound that makes these intervals satisfying to the ear. Due to the fact that 12 perfect fifths does not equal 7 octaves, just intonation does not allow for easy movement between keys and thus was abandoned in favor of more versatile tuning systems.

1.2.2 Pythagorean Intonation

In Pythagorean intonation, all of the frequency ratios are derived from multiples of the perfect fifth. For instance, starting from the note D and either ascending or descending by a perfect fifth, we can obtain the Pythagorean scale. When we ascend, we generate the sharped notes and when we descend, we generate flats:

A♭–E♭–B♭–F–C–G–**D**–A–E–B–F♯–C♯–G♯

Pythagorean intonation is associated with certain musical performance practices. It is sometimes ascribed to the phenomenon of raising sharped notes and lowering flat notes as well as raising leading tones when they precede tonics. Pythagorean intonation is especially associated with Pablo Casals who dubbed this phenomenon “expressive intonation” - the idea that one must alter the tuning of notes especially in relation to the musical and harmonic context.

1.2.3 Equal Temperament

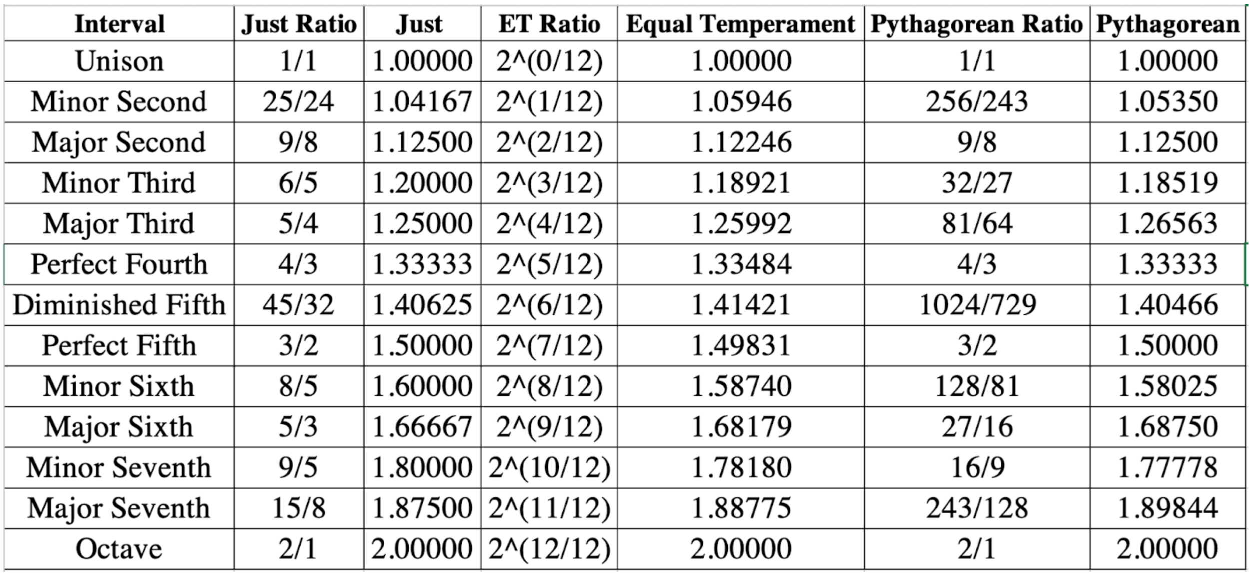
The final major tuning system used regularly in string playing is equal temperament which is the predominant standard in the tuning of pianos today. Equal temperament provides a solution to the issue that exists with just intonation: instead of constructing whole integer ratios based on the harmonic series, it generates a 12-tone scale by solving the following equation:

$$r^{12} \overset{?}{=} 2$$

The solution to this equation, $r = \sqrt[12]{2}$, ensures each semitone contains the same logarithmic frequency ratio. As such, only the interval of the octave is preserved compared to the harmonic series; every other interval is slightly more or less (tempered) compared to its just counterpart. The system is the most flexible as it allows for seamless playing in all keys; a piece in G major in just intonation would sound quite harsh and unpleasant in any other key while equal temperament would preserve the overall perception of pitch relationships.[3]

1.2.4 Comparison of Tuning Systems

A comparison of the frequency ratios of all three tuning systems is shown below. The octave is the only interval that is preserved among the three:



Interval	Just Ratio	Just	ET Ratio	Equal Temperament	Pythagorean Ratio	Pythagorean
Unison	1/1	1.00000	2^(0/12)	1.00000	1/1	1.00000
Minor Second	25/24	1.04167	2^(1/12)	1.05946	256/243	1.05350
Major Second	9/8	1.12500	2^(2/12)	1.12246	9/8	1.12500
Minor Third	6/5	1.20000	2^(3/12)	1.18921	32/27	1.18519
Major Third	5/4	1.25000	2^(4/12)	1.25992	81/64	1.26563
Perfect Fourth	4/3	1.33333	2^(5/12)	1.33484	4/3	1.33333
Diminished Fifth	45/32	1.40625	2^(6/12)	1.41421	1024/729	1.40466
Perfect Fifth	3/2	1.50000	2^(7/12)	1.49831	3/2	1.50000
Minor Sixth	8/5	1.60000	2^(8/12)	1.58740	128/81	1.58025
Major Sixth	5/3	1.66667	2^(9/12)	1.68179	27/16	1.68750
Minor Seventh	9/5	1.80000	2^(10/12)	1.78180	16/9	1.77778
Major Seventh	15/8	1.87500	2^(11/12)	1.88775	243/128	1.89844
Octave	2/1	2.00000	2^(12/12)	2.00000	2/1	2.00000

Figure 1.2: A comparison of the three intonation systems. The octave is the only musical interval preserved among the three.

1.3 Problem Formulation

Musicians, whether consciously or not, use a combination of tuning systems during performance. The placement of notes depends largely on a musician's individual preference for intonation. Because of tempo considerations and cognitive overload, in performance,

[3] 12-TET was calculated by Simon Stevin around 1600, but was actually widely rejected by musicians because it mistuned (almost) all the intervals, except for the unison and octave. It was first calculated in China, by Zhu Zaiyu in 1584.

musicians are unable to place every single note according to a predetermined intonation system. Musicians are not always aware of what intonation decisions they make. Understanding what intonation decisions musicians make would provide insight into whether there are fundamental mathematical and musical principles that guide these decisions. A clearer understanding of this could improve musical instruction and performance. Exploring this topic would start with a method to determine what intonation systems are used in musical performances. My thesis asks the following: given a raw audio recording of a movement from a Bach Cello Suite, how can we detect and identify different tuning systems? Particularly, how can we find instances of just, Pythagorean, and equal temperament from a signal processing and mathematical perspective?

1.3.1 Data

The Bach Cello Suites are considered to be the pinnacle of the cello repertoire. They are musical and harmonic masterpieces and thus give rise to many interesting questions concerning what particular music theoretical or harmonic features may be relevant in determining the usage of a particular intonation system. For instance, does the presence of a particular kind of musical structure such as a scale or arpeggio lend itself more often to Pythagorean intonation and does more vertical type music involving a reoccurring pedal tone to just? Additionally, debates surrounding intonation are arguably most potent in the Bach Cello Suites. In addition to how historically 'accurate' or stylistic they should be played, cellists frequently argue as to how the intonation should sound. For the purposes of the thesis, the audio data was limited to various recordings of the Third Cello Suite Prelude. Twelve cellists' recordings of the Prelude were used, including Anner Bylsma, Colin Carr, Pablo Casals, Ethan Cobb[4], Ralph Kirshbaum, Yo Yo Ma, Mischa Maisky, Jean-Guihen Queyras, Mstislav Rostropovich, Heinrich Schiff, Jan Vogler, Pieter Wispelwey.

For the purposes of fundamental frequency detection, which is the foundation of intona-

[4] The inclusion of myself in this list of legendary cellists is for my own sake and in no way a claim of equivalence.

tion analysis, the data was all converted to mono, 44.1 kHz sampled audio and edited to exclude instances of chords. It goes without saying that these twelve cellists do not represent the entirety of cello playing. They do however exhibit a fascinating set of results as we will see.

1.4 Overview of Pipeline

To begin to answer the question of how to identify a certain intonation system, it is evident that we must be able to extract accurate fundamental frequency information and timestamps of individual notes. This is the purpose of the note detection component of the data processing pipeline. Once we have this foundation, we can then run a probabilistic intonation detection algorithm to determine the final output of the pipeline: a set of timestamps and their associated intonation system labels for a particular raw audio recording. This is the intonation analysis component of the pipeline. A schematic of the pipeline is provided below and each step will be discussed in further detail in later chapters:

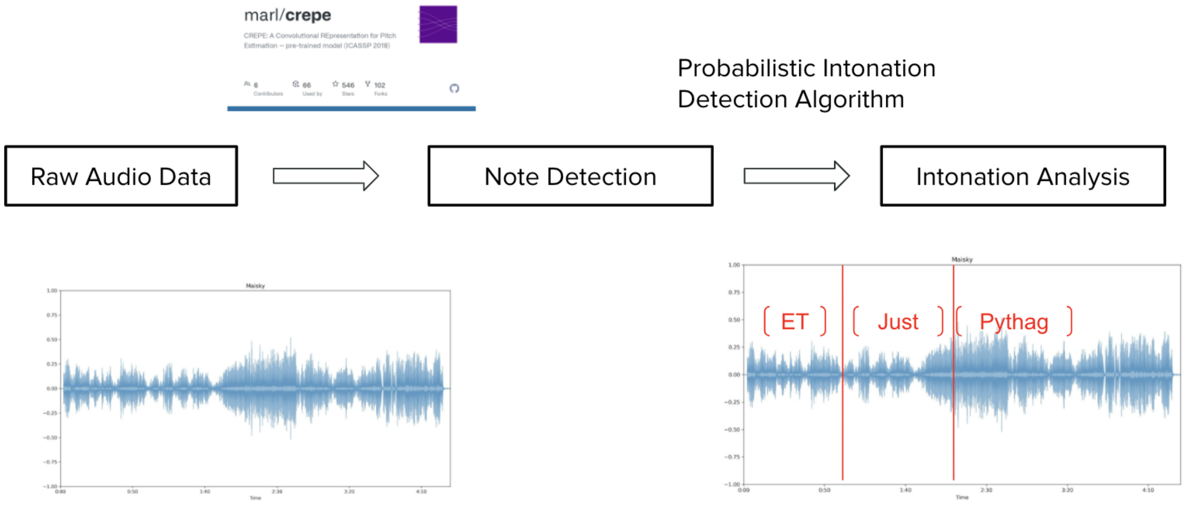

marl/crepe
CREPE: A Convolutional REpresentation for Pitch Estimation -- pre-trained model (ICASSP 2018)
6 Contributors
66 Used by
546 Stars
102 Forks
Probabilistic Intonation Detection Algorithm
Raw Audio Data
Note Detection
Intonation Analysis
Maisky
Time
ET
Just
Pythag

Figure 1.3: An overview of the overall data processing pipeline.

1.5 Summary of Contributions

I have developed the following:

- A method to extract individual notes and their associated timestamps and fundamental frequencies with high accuracy.
- An algorithm and associated heuristics to calculate the probability of a group of notes falling under a certain intonation system.
- An algorithm that accepts time series data including fundamental frequency information and identifies intonation systems over time.

Musicians, including myself, frequently argue about countless issues of intonation–whether or not someone is out of tune, whether a certain note should be raised or lowered according to the harmonic context, etc., and there is a dearth of computational and numerically-based software that can begin to shed light on some of these questions. Intonation occupies a majority of individual practice time and musicians often wonder whether or not they are playing "in tune." I hope to emphasize that there is no right answer to this question; you can only ask whether one is playing in tune in relation to a certain intonation system. I believe that my system will be a useful source of objective analysis for musicians to use and guide their playing.

Chapter 2

Literature Review

2.1 Temperament Estimation

The most relevant kind of research in relation to my project has been in the subject of temperament estimation in audio recordings. Simon Dixon, Matthias Mauch, and Dan Tidhar tried to address the problem of how to estimate the inharmonicity and temperament of a harpsichord given only an audio recording [6] . These quantities can be captured pretty easily looking only at individual notes but to do this with just an audio recording is nontrivial.

They produce an initial transcription to generate a list of note candidates and then employ high-precision frequency estimation techniques and statistics to estimate the inharmonicity and fundamental frequency f_0 of each note. They then match these estimates to a set of known keyboard temperaments and allow for variation in the reference tuning frequency to obtain the temperament.

The authors capture many of the challenges that remain the same for cello recordings. The classification of temperament requires a very high frequency resolution as the differences between different note frequency estimates among different temperaments can be on the order of a few cents or hundredths of a semitone. To capture this kind of frequency resolution, one needs a large FFT (Fast Fourier Transform) window length which of course diminishes the time resolution to the order of a few seconds. This is problematic because notes aren't stationary and don't usually last this long and because low order harmonics require even smaller frequency resolution to distinguish. Notes also rarely occur in isolation; there might be bias in frequency estimates - most favored intervals in western music are those in which many partials coincide - this is the basis of just intonation for instance. It's difficult to tell whether a peak in the Fourier Transform is a fundamental frequency or a partial of another note but this is obviously critical for the task of the authors.

To achieve frequency resolution on the order of cents, they use the FFT with quadratic interpolation and correction of the bias due to the window function. They also produce their own test dataset with both real and synthesized harpsichord music where the synthesized music has precise temperament but slightly different timboral/recording conditions.

2.1.1 Inharmonicity

The study of inharmonicity refers to investigation of the physical and acoustical properties of vibrating strings and the analysis of the perceptibility of inharmonicity from an psychoacoustic point of view. An example of inharmonicity is that vibrating strings have partials at frequencies slightly greater than integer multiples of the fundamental frequency; this phenomenon can be attributed to both the stiffness of the string and the amplitude of vibration [7].

The frequency of the kth partial is given by the following equation, where B is a inharmonicity constant:

$$f_k = kf_0\sqrt{1 + Bk^2}$$

2.1.2 Fundamental Frequency Estimation

There are notable limitations with existing fundamental frequency estimation methods. There are a number of often-made assumptions about the signal that do not always hold for musical signals. Among them are: monophonicity - the input signal at any point in time consists of a single pitched tone (although for my problem, this is the case); stationarity - that the signal's properties are stable over time; and that the properties of the input signal are known or match a small set of instruments. The authors also note that methods often ignore the effects of inharmonicity and human perception on the pitch estimation and that papers rarely deal with frequency resolution on the order of a few cents and the limitations that come with that (which is necessary for temperament and inharmonicity).

They show that the FFT with quadratic interpolation and correction of the bias due to the window function outperforms instantaneous frequency estimation using phase information and is suitable for estimating temperament and inharmonicity. As I also have thought of, they note that the ideal solution for f_0 estimation would involve identifying the existence and timing of each note in a recording but that no known transcription algorithm accomplishes this. I have achieved a version of this task. They employ a 2-stage approach to estimate f_0 and inharmonicity of unknown notes in the presence of multiple simultaneous tones. They first do a conservative transcription with high precision (keeping a high fraction of correctly-transcribed notes) at the cost of low recall (they might miss some notes) to identify notes, and then do the frequency estimation [8].

The conservative transcription consists of: frame-wise amplitude spectra with a STFT; sinusoid detection through peak-picking which yields initial frequency estimates; and deleting sinusoids with low confidence either because they're below an amplitude or duration threshold, or if they're overtones of a different sinusoid. For the partial detection, they identify peaks in the amplitude spectrum $|X(n,i)|$. They calculate the moving weighted mean $\mu(n,i)$ and moving weighted standard deviation $\sigma(n,i)$. A locally salient bin is identified when a spectral bin exceeds the moving weighted average plus half a moving standard deviation:

$$|X(n,i)| > \mu(n,i) + .5 \cdot \sigma(n,i)$$

Globally salient peaks correspond to bins that have an amplitude greater than 25dB less than the global maximum bin amplitude:

$$|X(n,i)| > 10^{-2.5} \max_{u,v}\{|X(u,v)|\}$$

They consider peaks that fulfill both of these inequalities and then estimate the frequency with quadratic interpolation of log magnitude of the peak bin and its two surrounding bins. They they go through all sorts of further deleting and refining of the potential fundamental frequencies. Frequency estimates are sorted into semitone bins from MIDI note 36 to 80.

After this conservative transcription, they then employ partial frequency estimation using the first equation with a frequency-dependent B (B is initially a constant). With two partial frequencies f_j, f_k, B can be estimated by the following equation, provided there is no interference between partials from other notes:

$$B_{j,k} = \frac{j^2 f_k^2 - k^2 f_j^2}{k^4 f_j^2 - j^4 f_k^2}$$

Dixon, Mauch, and Tidhar are quite successful in this regard however their task has crucial structural differences compared to my task. Most notably, the harpsichord is a *fixed-pitch* instrument while the cello is a *continuous-pitch* instrument. Because of this, there is one right answer to the question of what kind of temperament system is used in a harpsichord recording and there is a limit to the range of possible frequencies for each note. On the cello, the frequency of a note depends on one's finger placement on a fret-less fingerboard - an infinite number of possibilities. Moreover, in the case of a cello recording, there cannot be one classification of a certain temperament; these systems change over time depending on musical and harmonic context.

2.2 Intonation Studies

There is also a variety of literature concerned with identifying intonation tendencies and pitch-drift in various kinds of music and audio recordings. Johanna Devaney and Dan Ellis studied intonation tendencies in polyphonic vocal tracks, that is tracks with multiple voices [9]. Their central assertion was that the overall tuning of a vocal ensemble cannot be determined by a singular reference point; rather, horizontal and vertical factors inform the intonation. This is also the case for string players in general. To begin their analysis, they consider various music theoretical notions concerning consonance and voice leading. Particularly, they consider Ernst Terhardt's theory of consonance which places a lot of weight in the fundamental or lower overtones of a note. He describes consonance as when the real bass note and virtual fundamental note align, an idea which can be applied to the idea of tuning preferences in vertical sonorities.

In addressing vertical aspects of voice leading, Lerdahl's tonal pitch space theory deals with the tendency of a dissonant pitch to resolve to a consonant neighbor and follows a rule analogous to the inverse square law in Newtonian gravitation. In this model, the attraction of one pitch to another is the anchoring strength of the goal pitch s_2 divided by the anchoring strength of the source pitch s_1 times the inverse of the square of the number of semitones between the two pitches n:

$$\frac{s_2}{s_1} \cdot \frac{1}{n^2}$$

Larson also deals with melodic attraction in his work on melodic forces, correlating gravity, magnetism, and inertia in a single equation [10]. The total force acting on a note in a given context or pattern is calculated by summing the results of individual calculations for each force:

$$F = w_G G + w_M M + w_I I$$

Gravity, G, is the tendency of a musical line to go down and is a binary variable - it's either a 1 if a pattern descends towards a more stable pitch, and 0 otherwise. Magnetism, M, is the tendency of unstable notes to move to stable ones. The formula for magnetism is as follows:

$$M = \frac{1}{d_{to}^2} - \frac{1}{d_{from}^2}$$

d_{from} is the distance in semitones from the initial note to the closest stable pitch while d_{to} is the distance in semitones from the initial note to the goal note in the current musical context. Inertia, I, refers to the tendency of a musical line to continue rather than vary. Inertia is 1 when the musical pattern has inertial potential and fulfills it, 0 if it has no inertial potential, and -1 if it stays on the same pitch and has an I value of 0. w_G, w_M, and w_I represent the weightings on each of these variables and are found using multiple regression.

Lerdahl's model is useful because it is internally consistent and generates a full complement of attractional relations within a musical system; Larson's model requires some modification because it cannot accomodate a change in the governing tonic part-way through a musical sequence, calculate attractions from a stable pitch, or generate negative values which makes comparisons difficult.

The trained models will suggest horizontal intonation tendencies. The reconciliation of the vertical and horizontal depends on a number of factors, including the duration and metrical position of a given vertical sonority, its function within the musical context, and significance of the the horizontal lines moving through a vertical sonority (relation of pitch material to current harmonic context).

For their computational analysis, 12-tone equal temperament is used as a reference because it is invariant under changes to tuning references. Data collection is a 2-step process - temporally aligning a MIDI score of the work to the audio recording and developing a method to accurately extract pitch data from polyphonic vocal recordings. The MIDI version guides the pitch estimation but the alignment is tricky since note onsets can often be difficult to determine, and that the timbre of all the vocal parts are very similar and also the amount of reverb. They use a dynamic programming approach which relies on accuracy in pitch but is less sensitive to exact onset times.

The aligned score now contains the approximate time and frequency of each performed note, where the frequency is determined by an instantaneous-frequency spectral analysis which calculates a phase derivative within each time-frequency cell of a STFT. The accuracy of the estimate is limited by the amount of energy from noise, harmonics, etc that may be present in the bin. The IF spectrogram recovers the estimated energy and frequency of sinusoids at every time-frequency cell and single pitch values are estimated using an energy-weighted average of the instantaneous frequencies aligned to each note. As is the case for the cello, vibrato makes this task more difficult.

Mauch and Dixon have also explored intonation and intonation drift in vocal singing and proposed a model of reference pitch memory where the reference pitch is treated as a changing latent variable [11].

They treat intonation as the signed pitch difference (measured in semitones on an equal-tempered scale) relative to the reference pitch. They also use equal temperament as their reference tuning system but claim that this doesn't substantially affect their results. They convert frequencies to MIDI pitches by the usual conversion, with A440 as the reference pitch:

$$p = 69 + 12\log_2 \frac{f_0}{440}$$

Then, for a particular frequency estimation, the difference in that estimated MIDI pitch's value to the closest integer pitch is the deviation in cents from that particular pitch. They use the term nominal to refer to intervals or pitches with respect to ET. A perfect fifth corresponds to 7 equal semitones for instance, while in general a nominal interval can differ from an observed interval.

They underwent a semiautomatic pitch-tracking process - they annotated the note onsets and offsets by using Sonic Visualizer and identifying the stable part of the estimated pitch track. The annotations were fed into pitch tracking software based on YIN and the note tracks were analyzed with R. They took median pitch estimates to approximate the pitch value of each note.

They measure the interval of the ith note as the signed difference in semitones:

$$\Delta p_i = p_i - p_{i-1}$$

The interval error then of the observed interval is

$$e_i^{int} = \Delta p_i - \Delta p_i^0$$

where Δp_i^0 is the nominal interval in semitones in ET.

In trying to define pitch error, the authors claim that since the tuning changes over the course of the song with singers, there is no reference pitch to base intonation off of. Thus, to obtain a reference, they use a linear fit to a local tonic estimate. For the measured pitch of the ith note, p_i, they find the estimate:

$$t_i = p_i - s_i$$

where s_i is the nominal pitch relative to the estimated tonic.

2.3 Music Information Retrieval

Within the larger area of music information retrieval, there has been work in a number of different areas in data-driven pitch and music theory-related tasks [12]. Pitch histograms have been used extensively in music information retrieval studies. Some work has also been done in non-western music, particularly Turkish music [13]. A *makam* in Turkish music is a modal entity; every musical piece is identified and recognized by a particular type of makam. Many of the problems inherent in trying to estimate the makam of a given piece are similar to the issues inherent in the task of identifying intonation systems in cello recordings–pitch-class histogram-based methods often have a number of assumptions used to reduce the dimension of the pitch histogram space which don't always apply and much of MIR relies on western tonality and equal-tempered tuning (A4=440 Hz). Because many of these concepts do not follow in Turkish music, the analysis has to be done with little assumptions - particularly, they don't take any specific tuning system for granted. This is exactly analogous to identifying intonation systems in cello recordings. They also extract frequency data from monophonic audio recordings to construct pitch histograms.

Machine learning approaches have also been explored in music information retrieval. Harasim et al. attempted to infer the number and characteristics of modes in different historical periods dating from the Renaissance to the 19th century [14]. They used an unsupervised learning approach to determine the number of nodes with a geometric model, capturing modes as clusters of musical pieces in a non-Euclidean space, then using a Bayesian model to characterize the modes.

Chapter 3

Note Detection

At the foundation of any analysis of intonation or temperament is an accurate calculation of fundamental frequency on a note-by-note basis. The subject of fundamental frequency analysis has been thoroughly studied but while many existing pieces of software (Melodyne, Logicpro, etc.,) are able to extract individual notes with a high degree of accuracy, there are no existing modules within Python or other languages that perform the complete task of extracting fundamental frequency estimates and timestamps of every note in an audio recording. This task was thus one of the first steps needed in the overall pipeline.

Extremely high frequency resolution is needed for the task of estimating intonation systems as frequency estimates may differ on the order of a few cents and further work in this area is noted in section 6[8]. As a first approximation, and for its ease of use and high accuracy compared to other state-of-the-art monophonic fundamental frequency estimators, CREPE was used.

3.1 CREPE

CREPE is currently the highest-performing fundamental frequency estimator outperforming other leading programs like pYIN and SWIPE. CREPE is a pretrained deep convolutional neural network which operates on the time-domain waveform [15]. By default, CREPE operates every 10 milliseconds and outputs three readings at each time-step: the time (in seconds), the frequency estimate (in Hz), and a confidence estimate (confidence in the presence of a pitch). A sample output for an equal tempered C4 is shown below:

time	**frequency**	**confidence**
0.00	263.82	0.68
0.01	263.00	0.84
0.02	262.29	0.92
0.03	261.91	0.94
0.04	261.52	0.94
0.05	261.82	0.94
0.06	261.67	0.94
0.07	261.79	0.94
0.08	261.63	0.94
0.09	261.73	0.94
0.10	261.79	0.94
0.11	261.72	0.94
0.12	261.66	0.94
0.13	261.73	0.94
0.14	261.71	0.94
0.15	261.81	0.94
0.16	261.75	0.94
0.17	261.53	0.94
0.18	261.79	0.94
0.19	261.66	0.94
0.20	261.76	0.94

Table 3.1: Sample CREPE output

3.2 Note Detection Algorithm

The goal of the note detection algorithm is two-fold:

1. Calculate timestamps for each note in the audio
2. Calculate a fundamental frequency estimate for each note in the audio

We wish to retain temporal information for each note so to be able to ultimately perform analysis concerning the usages of intonation systems over time. This information may eventually combined with musical contextual information such as harmonic and melodic content to provide more insight as to how intonation is dependent on these features. High resolution frequency estimates are obviously essential for the task of assessing intonation. Each of these requirements may be solved individually with different kinds of approaches but the goal of the algorithm is to be able to capture both of these pieces of essential information in one pass.

The algorithm takes advantage of CREPE's high accuracy when it comes to time-domain frequency detection. Extensive data analysis showed that CREPE's confidence values tended to be quite high when the frequency estimate was within the ballpark of the actual note frequency estimate. It was possible to ascertain this correlation by testing with synthesized audio or an excerpt of the Prelude, for which we know the notes played and thus the theoretical frequency estimates. The algorithm leverages this correlation by assigning a confidence threshold to the frequency array to weed out noisy estimates. The remaining estimates then are segmented according to where consecutive frequency estimates exceed a presupplied frequency ratio threshold. For the purposes of the thesis, this threshold was set to $2^{\frac{1}{26}}$, a bit less than the equal tempered semitone ratio value of $2^{\frac{1}{12}}$.

The algorithm can be described as follows, letting:

- h denote the step size in seconds.
- ϵ denote a small deviation value in seconds, (on the order of 10^{-5})

Algorithm 1 Note Detection Algorithm

```
$N \leftarrow length(F)$
$F_{current} \leftarrow F[0]$
for $i \leftarrow 1$ to $N$ do
    if $t[i] - t[i-1] > 4 \cdot (.001h) + \epsilon$ then
        $f_r \leftarrow \frac{max(F[i],F[i-1])}{min(F[i],F[i-1])}$
        if $f_r < f_t$ then
            continue
        end if
        $t_f \leftarrow t[i-1] + .001 * h - \epsilon$
        $dict[(t_i, t_f)] \leftarrow \frac{\sum F_{current}}{length(F_{current})}$
        $t_i \leftarrow t[i]$
        $F_{current} \leftarrow [f[i]]$
    else
        $F_{current}.append(f[i])$
    end if
end for
```

- F denote the frequency array containing the valid frequency estimates (after confidence-thresholding the CREPE output)
- $F_{current}$ denote the current array of frequency estimates for the current pitch under consideration.
- f_r denote the frequency ratio calculated between consecutive frequency estimates.
- t_i denote the initial time in seconds of the current pitch under consideration.
- t_f denote the final time in seconds of the current pitch under consideration.

3.3 Algorithm Performance

To test the accuracy of the note detection algorithm, we test it on a mix of synthesized and normal audio. For simplicity, we tested the algorithm on a synthesized equal-tempered scale and calculated the cent difference for each detected note:

Note	C4	D4	E4	F4	G4	A4	B4	C5
Freq. Pred.	261.737	294.057	329.661	348.925	392.293	440.348	494.153	524.522
ET Value	261.63	293.67	329.633	349.234	392.002	440.007	493.892	523.26
Cent Dev.	0.709	2.28	0.147	-1.532	1.286	1.339	0.917	4.171

Table 3.2: Results of freq prediction on c major equal temperament scale

Note:	C4	D4	E4	F4	G4	A4	B4	C5
Time Prediction:	0.02	.515	1.015	1.515	2.025	2.52	3.015	3.51
Time:	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
Time Difference:	0.02	0.015	0.015	0.015	0.025	0.02	0.015	0.01]

Table 3.3: Results of time predictions on c major equal temperament scale

As we can observe, the algorithm performs quite well- the frequency deviation stays under 5 cents, the JND or "just-noticeable difference." Between equal-tempered B4 and C5, the just noticeable difference would correspond to:

$$\text{JND (B4, C5)} \ = \frac{f_{\Delta}}{20} = \frac{523.26 - 493.892}{20} = 1.4684 \text{ Hz}$$

The time prediction also performs quite well; the deviation never exceeds .025 seconds or 25 milliseconds.

Chapter 4

Probabilistic Intonation Detection

Now that we have a system that can parse through an audio file and obtain accurate time and frequency estimates of each note, we are able to ask the original question of the thesis: how can we detect intonation systems over time?

There are many possible approaches to answering this question. I begin by describing just a few of these possible approaches. Naturally, the process of detecting intonation systems depends entirely on frequency relationships; musical intervals and their corresponding ratios in just, equal-temperament, and Pythagorean intonation are only defined with two notes. As such, one could possibly only compute frequency information for every note in an audio file and group this information into pitch class data. Each pitch class would have a value equal to the mean fundamental frequency of that pitch class normalized to some octave. With this data, it would be possible to compute average musical interval information and classify the entire audio file as the intonation system for which the musical interval ratios are closest. This approach has potential for interesting data analysis but lacks temporal information and thus fails to completely answer the question of the thesis. It also represents a one-answer simplification to the question of what intonation system is a given player using which in reality is not the case. Performers change their intonation on a variety of musical features and musical contextual information [1]. Among these, the particular musical intervals between notes are highly influential in the choice of intonation system. For instance, it is rare and virtually unacceptable to play a perfect interval according to just intonation. This is of course due to the fact that the beating between two of these frequencies is extremely audible and quite harsh-sounding. It must be acknowledged, however, that this kind of musical interval preference is indeed subjective.

To obtain temporal intonation information, we first estimate the probability of a note falling under a certain intonation system. We approximate a given note to be normally distributed with mean described by the reference frequency under a certain intonation system and standard deviation approximated to be a fixed constant value. The reference frequency corresponds to the closest possible frequency in a certain intonation system stencil as judged by the minimum absolute value of the difference between a given frequency. We describe the process for calculating the stencils and corresponding intonation system probabilities for each of the three intonation systems:

4.1 Tuning System Stencil Creation

4.1.1 Equal Temperament

Equal temperament is the simplest of the three systems to approximate. The stencil is generated by a certain reference frequency as judged by the frequency value for A4. This frequency can be approximated for each of the cellists with exploratory data analysis using the note detection algorithm. Once a reference frequency is given, every note in the stencil is calculated by multiplying (or dividing) by $2^{\frac{1}{12}}$ and scaling by a power of 2 to stretch the range of four octaves on the cello corresponding to the approximate frequency range 65 Hz - 500 Hz.

4.1.2 Pythagorean

To generate the Pythagorean stencil, we again use a frequency approximation for A4 and a particular key-invariant definition derived from the just perfect fifth frequency ratio. Based on this reference frequency, we traverse by perfect fifths up and down and scale by the necessary powers of two to end up in the same cello frequency range.

4.1.3 Just

The just stencil is the most complex of the three intonation systems given its key dependence. The process for generating a just stencil in a particular key again starts with a frequency approximation for A4. To calculate the just stencil in C major for instance, we calculate the equivalent frequency for open C on the cello (C2) by descending three perfect fifths from the frequency approximation for A4 and scaling by an octave:

$$\mathrm{C}_f = 441 \times \left(\frac{2}{3}\right)^3 \times \frac{1}{2}$$

Each note of the stencil is generated by multiplying by the necessary just intonation interval ratio and scaling. For instance, the note E is calculated by multiplying the just intonation major third ratio ($\frac{5}{4}$) by the fundamental frequency of C. Once we've calculated E, we can calculate its sharp and flat equivalents by multiplying and dividing by the just minor second ratio ($\frac{25}{24}$ respectively). The other keys are derived in exactly the same fashion; they only differ by the base reference frequency for which all of the other notes in the scale are derived.

C	65.33
C#	$C \cdot \frac{25}{24}$ *
Db	$D \cdot \frac{24}{25}$ *
D	$C \cdot \frac{9}{8}$ *
D#	$D \cdot \frac{25}{24}$ *
Eb	$E \cdot \frac{24}{25}$ or $C \cdot \frac{6}{5}$ *
E	$C \cdot \frac{5}{4}$
E#	$E \cdot \frac{25}{24}$ *
Fb	$F \cdot \frac{24}{25}$
F	$C \cdot \frac{4}{3}$
F#	$F \cdot \frac{25}{24}$ *
Gb	$G \cdot \frac{24}{25}$ *
G	$C \cdot \frac{3}{2}$
G#	$G \cdot \frac{25}{24}$ *
Ab	$A \cdot \frac{24}{25}$ or $C \cdot \frac{8}{5}$ *
A	$C \cdot \frac{5}{3}$ *
A#	$A \cdot \frac{25}{24}$
Bb	$B \cdot \frac{24}{25}$ *
B	$C \cdot \frac{15}{8}$ *
B#	$B \cdot \frac{25}{24}$ *

Table 4.1: C major just intonation stencil. The starred notes are shown just for completeness but do not appear in the C major scale from a music theoretical standpoint.

4.2 Calculating Tuning System Probability

Each note has an associated probability vector

$$\vec{p}(f) = [p_j(f), p_e(f), p_p(f)] \tag{4.1}$$

where $p_j(t)$ is the probability of a justly-tuned pitch, $p_e(t)$ is the probability of an equal-tempered pitch and $p_p(t)$ is the probability of a Pythagorean-tuned pitch. $p_i(t) \sim \mathcal{N}(f_r, \sigma^2)$ where f_r is the reference frequency in that particular intonation system and σ^2 is a fixed constant value. The probability of a note falling under a certain intonation system is given by:

$$p_i(f) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(f - f_r)^2}{2\sigma^2} \tag{4.2}$$

4.3 Tuning System Change Detection Algorithm

The algorithm takes as input the time-frequency data generated by the previous step in the pipeline. Using these fundamental frequency estimates, $\vec{p}(t)$ is calculated for each note. Using a certain window size of notes, an individual tuning system label is generated for each note by the tuning system that maximizes the average probability over the window size, effectively applying a moving average filter. Equal temperament and Pythagorean only have one possible reference frequency for each frequency under consideration, but just has 5 possible keys (C, G, D, F, A). Because of this, a certain frequency may have the same probability estimate for several just intonation stencils. For this reason, a certain kind of classification criterion needs to be employed for just intonation - either all the stencils are grouped under the larger “just” umbrella, or each key is handled separately. We explore both implementations in chapter 5.

Each side of the frequency array is padded accordingly so to end up with the same number of labels as notes. We describe the classification algorithm below:

Let f_i be the fundamental frequency of the ith note, W the window size (# of notes). The tuning system label corresponds to the system with the highest probability as calculated by equation 4.2.

$$\begin{aligned} p_i &= \max[p_j(f_i), p_e(f_i)), p_p(f_i)] \\ &= \max\left[\frac{1}{W}\vec{p}(f_{i-W/2}) + \ldots + \vec{p}(f_i) + \vec{p}(f_{i+1}) + \ldots + \vec{p}(f_{i+W/2})\right] \end{aligned} \tag{4.3}$$

If the answer to equation 4.3 is p_j, then the label is just, and so on. $p_e(f)$ and $p_p(f)$ are straightforward to calculate because we merely consult the lookup table for each of their stencils and find the answer that corresponds to the minimum absolute value with the frequency in question. $p_p(f)$ is more complex and corresponds to the maximum probability among the six possible key estimates:

$$p_p(f) = \max[p_C(f), p_G(f), p_D(f), p_A(f), p_E(f), p_F(f)]$$

For each frequency and its associated neighbors moving window, many just stencils may be possible and many of them may overlap as well. For this reason, the associated key label is calculated by the mode of the total number of keys represented (which still may not be unique, but is fine for the purposes of obtaining a just classification).

Once this has been applied to the entire data, we end up with an associated tuning system label for each fundamental frequency f_i for $i = 1, \ldots, N$ where N is the total number of detected notes. Once we have this per-note data, we need to extrapolate information about more global intonation patterns rather than this localized information. There are many possible approaches to this task - one could simply count the total number of instances of each label to achieve some sort of average global estimate of the usage of each intonation system but this would importantly once again sacrifice temporal information, which we need for analysis. Instead we apply the following algorithm to identify sequences of intonation systems in the tuning label time series:

We iterate through the array of tuning system predictions y_i until we find a pair of consecutive unequal labels. This defines the index location of a split. We store the time information of the last note in the current intonation system sequence as well as the total number of notes in the sequence. In the case that all labels are the same, the final output will be the entire sequence. Once we've gone through the entire array, we do one final sweep to to remove any sequences with a number of notes smaller than a given threshold (for the results in the following chapter, the minimum length of a sequence was set to two, meaning any sequences of one note were removed from the final output). We also check any instances of adjacent sequences in which the tuning system labels are the same and if $t_{\text{start},i} - t_{\text{final},i-1} < t_{\text{thresh}}$, then the sequences are grouped together into one sequence. In the end, we end up with a set of sequences with four components:

- Start (s) - the starting time of the sequence, as defined by $t_{initial}$ of the first note
- End (s) - the ending time of the sequence, as defined by t_{final} of the last note in the sequence
- Tuning System - the estimated tuning system of the sequence
- Total (# of notes) - the total number of notes in the sequence, including any labels that may have been "sandwiched" into the sequence

Chapter 5

Empirical Results

The purpose of this chapter is to showcase several sets of results for different kinds of audio that illustrate both the frequency detection capabilities of CREPE and the effectiveness of the tuning system detection algorithms but more importantly, what both of these components are not able to capture and why. We start with analyzing the algorithm's results on synthesized scales of the three intonation systems and then in more detail at the results for just the opening phrase of the Third Suite Prelude for each of the twelve cellists. Finally, we analyze the full results of the entire Third Suite Prelude for all cellists and examine one particular output in detail to shed light on what sorts of interpretive and musical factors may influence the results of the model.

A figure of the output is shown for each example case corresponding to the tuning system classification versus time. Two associated tables are also shown - one details which tuning system(s) were estimated, the total number of notes detected, and the fraction of the total number of notes represented by each tuning system in the audio, and the other outputs the total percent of just intonation and the most likely intonation system of the audio in question as judged by the system which corresponds to the largest percentage.

5.1 Scale Recognition

To begin testing the results of the algorithm, we start with synthesized diatonic A major scales consisting of eight notes total of the three different systems. We use a reference A4 frequency of 440.0 Hz and start on the note A3 = 220.0 Hz. Each note of the scale is determined by the particular frequency ratio specific to that system relevant to the tonic.

5.1.1 Equal Temperament

As we can see, the algorithm correctly classifies the scale and even detects and correctly labels all eight notes.

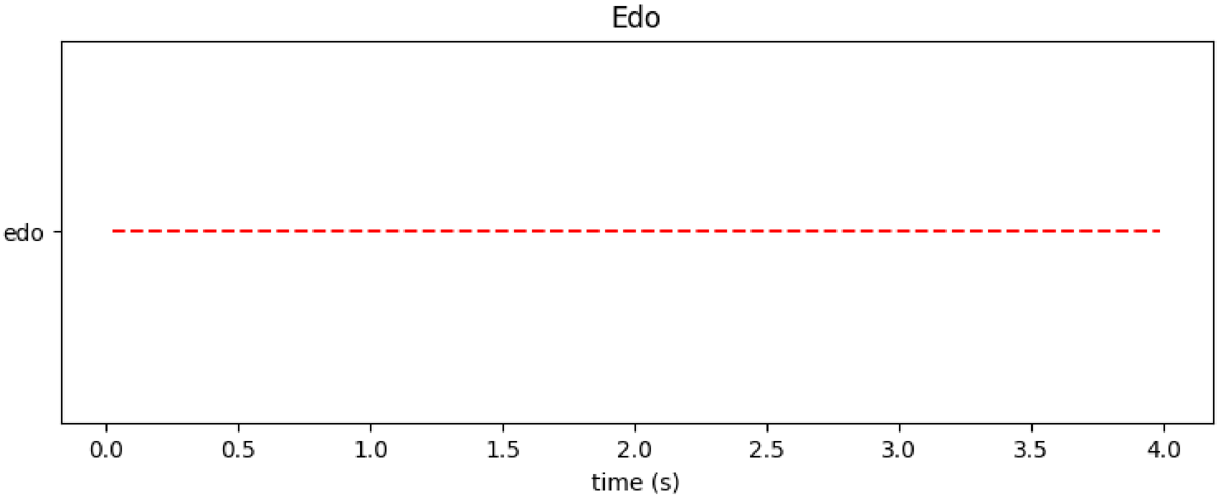


Figure 5.1: Output of equal tempered scale.

Table 5.1: Equal Temperament Scale

Tuning System	Total	Percentage
edo	8	100.0

Percent Just:	0.0
Most Probable Intonation System:	edo

5.1.2 Just

Similarly, the algorithm correctly classified the scale as the correct just intonation stencil and nearly detected every note with an accuracy of $\frac{7}{8} = 87.5\%$.

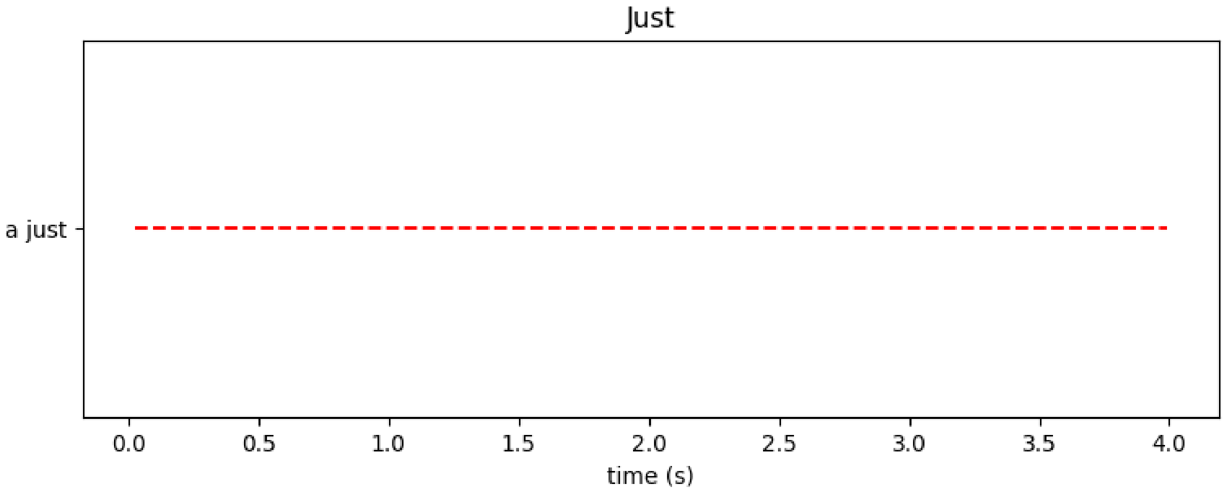


Figure 5.2: Output of just scale.

Table 5.2: Just Scale

Tuning System	**Total**	**Percentage**
a just	7	100.0

Percent Just:	100.0
Most Probable Intonation System:	a just

5.1.3 Pythagorean

The Pythagorean stencil is more tricky to capture in a scale because naturally the frequency differences between equal temperament except for two intervals - the major third and major seventh differ on the order of a few cents. This is with high probability beyond the frequency detection capabilities of CREPE. CREPE classifies the major second, major third, and perfect fourth as equal temperament but classifies the perfect fifth, major sixth, and major seventh correctly as Pythagorean.

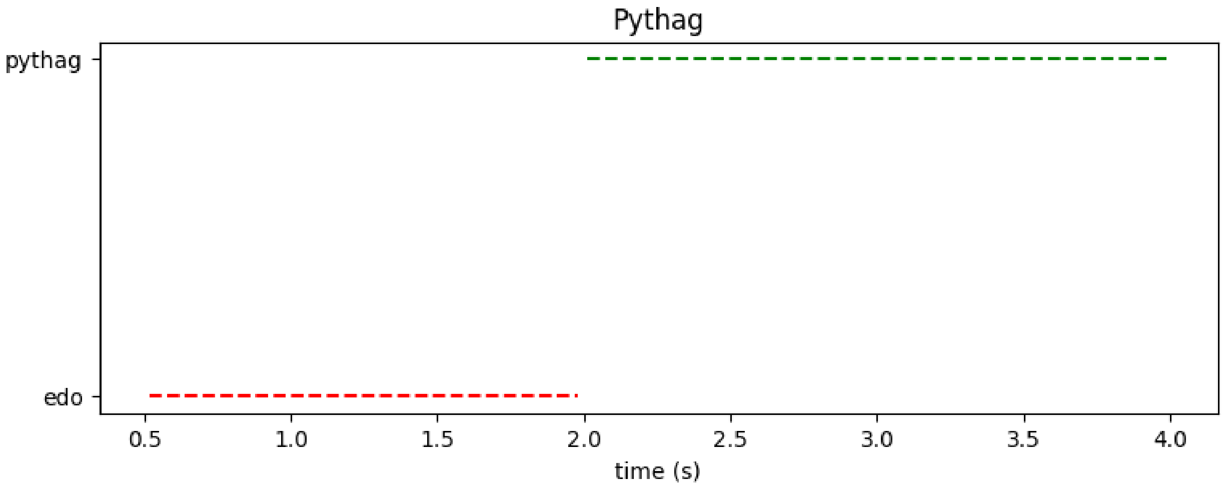


Figure 5.3: Output of Pythagorean scale

Table 5.3: Pythagorean Scale

Tuning System	Total	Percentage
edo	3	42.86
pythag	4	57.14

Percent Just:	0.0
Most Probable Intonation System:	pythag

5.2 Cello Suite No. 3 Prelude Opening Phrase

Now we begin our exploration of the main part of the project: the Bach Cello Suites. We start with just the opening phrase of the Third Suite Prelude, which starts on the note C4, follows a downward scale to an octave below on the note C3 and then descends by an arpeggio to the open C string on the cello (C2). This opening phrase contains only twelve notes within the C major scale and two kinds of musical structures - a scale and arpeggio.

As such, we would most likely expect cellists to use a combination of just intonation and Pythagorean intonation because of the combination of harmonic and melodic material. The sheet music to the opening phrase is shown below in figure 5.4.

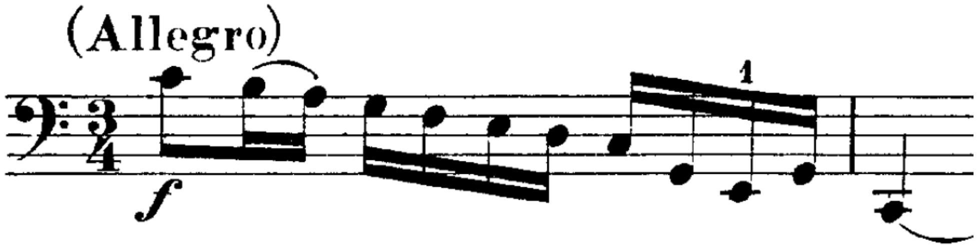


Figure 5.4: Opening Phrase of Suite 3 Prelude

The model has a small number of hyperparameters: a frequency estimate for A4 (to calibrate the intonation system stencils accordingly), a confidence threshold (for note detection, see chapter 3), a time threshold (to determine when adjacent equivalently-labeled sequences can be merged, see chapter 4), a window length (# of notes for use in the individual note-by-note label process, see section 4.3), and a step size (in milliseconds, for note detection, the default value for CREPE is 10 milliseconds). For all cellists, the time threshold, window length, and step size were set to .55, 3, and 5 respectively. The other parameters were estimated by a small grid search on a clip of audio.

cellist	**A4 freq**	**confidence**	**retention**
bylsma	407.15	0.8	54.06
carr	441.99	0.85	55.04
casals	432.67	0.88	52.84
kirshbaum	447.31	0.88	41.77
ma	439.96	0.85	47.39
maisky	440.13	0.9	47.93
queyras	444.38	0.8	32.71
rostropovich	440.54	0.7	67.31
schiff	443.42	0.8	34.06
vogler	442.22	0.8	56.70
wispelwey	394.54	0.8	54.39
cobb	441.47	0.8	49.72

Table 5.4: Hyperparameters used for each cellist

The overall output statistics of the opening phrase using both only the C Just stencil and all possible stencils are shown below. We will delve more into each individual cellist in the following subsections. By looking at a smaller example of the piece, we can gain insight into the behavior of the algorithm. As is certainly also the case and even more pronounced in the following section and appendix, much of the variation in the the overall most probable tuning system is replaced in favor of just when all stencils are used. When examining only the C stencil data, the algorithm confirms our expectation of Pythagorean intonation for Casals but interestingly classifies many examples of equal temperament that we wouldn't necessarily expect. We'd expect Bylsma, being a baroque cellist, to exhibit more bias towards just intonation but this isn't the case. Queyras is even entirely classified as equal temperament, which either could be due to numerical error or a suggestion that his frequent playing with equal-tempered piano as part of his solo and chamber music career has steered his intonation in the equal-tempered direction.

Using C Just Stencil		
	Percent Just	**Most Probable Intonation System**
Bylsma	37.4	edo
Carr	42.86	just
Casals	33.33	pythag
Cobb	40.0	edo/just
Kirshbaum	0.0	pythag
Ma	22.22	pythag
Maisky	22.22	edo
Queyras	0.0	edo
Rostropovich	25.0	edo
Schiff	25.0	edo/pythag
Vogler	66.67	just
Wispelwey	37.5	pythag/just

Table 5.5: Fraction of notes played with just intonation and overall most probable intonation system for all 12 cellists only using the C just stencil for opening phrase

With all stencils, detection is naturally pushed in favor of just intonation. The only impressive exception is Kirshbaum, which may support that his choice of Pythagorean intonation is indeed intentional.

Using All Stencils		
	Percent Just	**Most Probable Intonation System**
Bylsma	57.14	just
Carr	66.67	just
Casals	71.43	just
Cobb	60.0	just
Kirshbaum	28.57	pythag
Ma	100.0	just
Maisky	37.5	edo
Queyras	50.0	just
Rostropovich	100.0	just
Schiff	66.67	just
Vogler	100.0	just
Wispelwey	42.86	just

Table 5.6: Fraction of notes played with just intonation and overall most probable intonation system for all 12 cellists using all just stencils on opening phrase

5.2.1 Anner Bylsma

Forcing the c major stencil, Bylsma indeed uses c major just intonation but the algorithm detects an equal tempered majority. With the fraction of notes remaining that were missed by CREPE (most likely due to noise in the data and corresponding low confidence) it is possible that a few more detected notes could have tilted Bylsma in the c major just intonation direction.

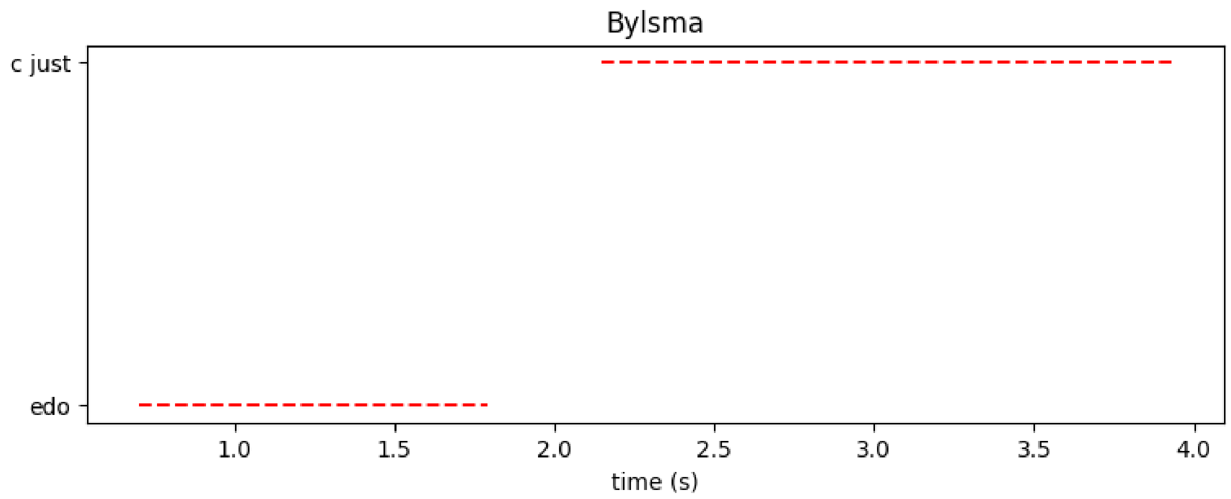


Figure 5.5: Output of Bylsma opening phrase with only c major stencil..

Table 5.7: Bylsma Opening Phrase with only c major stencil..

Tuning System	Total	Percentage
edo	5	62.5
c just	3	37.5

When we allow all just intonation stencils to be classified, we interestingly find that both the C and E stencils are detected. This makes sense as there is much overlap between the two stencils.

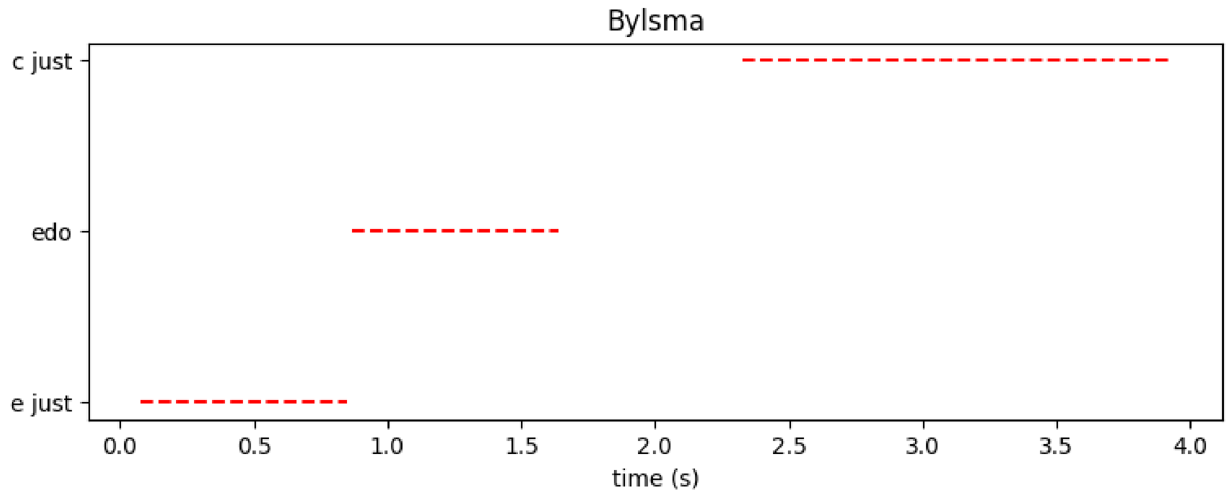


Figure 5.6: Output of Bylsma opening phrase with all just stencils..

Table 5.8: Bylsma Opening Phrase with all just stencils..

Tuning System	Total	Percentage
e just	2	28.57
edo	3	42.86
c just	2	28.57

5.2.2 Colin Carr

Carr plays the majority of notes with c major just intonation. He uses all three systems.

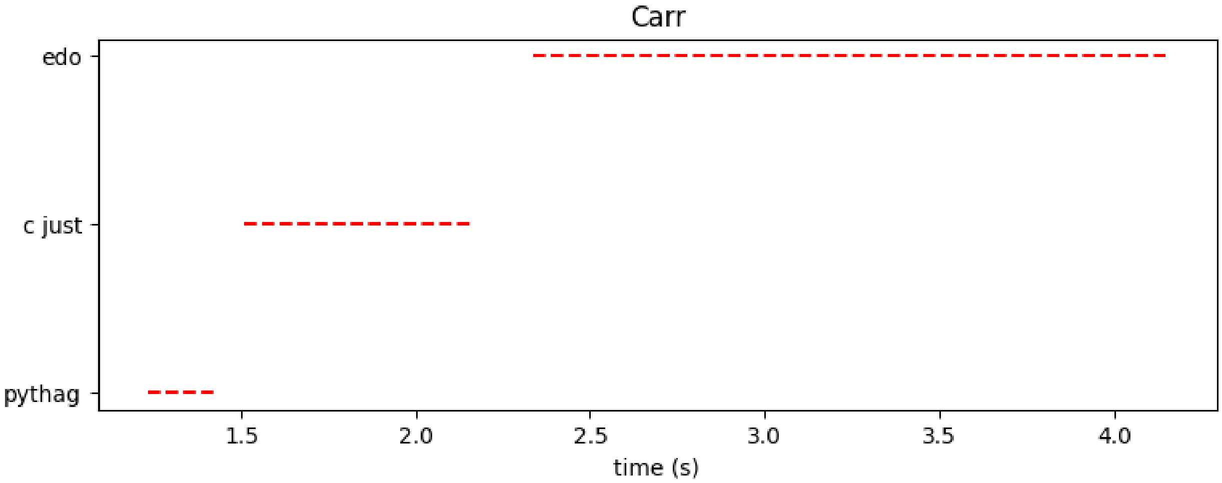


Figure 5.7: Output of Carr opening phrase with only c major stencil..

Table 5.9: Carr Opening Phrase with only c major stencil..

Tuning System	Total	Percentage
pythag	2	28.57
c just	3	42.86
edo	2	28.57

With all stencils, Carr still plays a majority just intonation. The D just intonation stencil is detected most likely again because of the significant overlap between the C and D stencils.

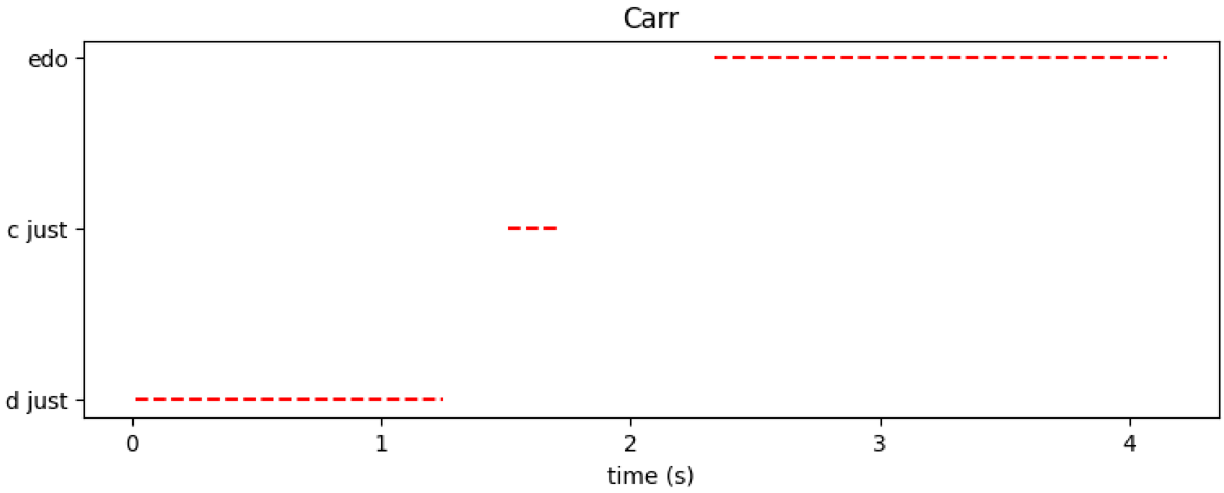


Figure 5.8: Output of Carr opening phrase with all just stencils..

Table 5.10: Carr Opening Phrase with all just stencils..

Tuning System	Total	Percentage
d just	2	33.33
c just	2	33.33
edo	2	33.33

5.2.3 Pablo Casals

The algorithm's detection of Pythagorean intonation for Casals is fitting - Casals is known for his use of 'expressive intonation.' The beginning may be classified as Pythagorean because of the small minor second interval between the first note and second note, which has a tendency to be squeezed. The c just detection in the middle of the audio is interestingly centered around the entirely melodic downward scale. The last Pythagorean detection may be exaggerated because the Casals hits the C string quite loudly which can raise the pitch of the open string by a small, albeit significant amount.

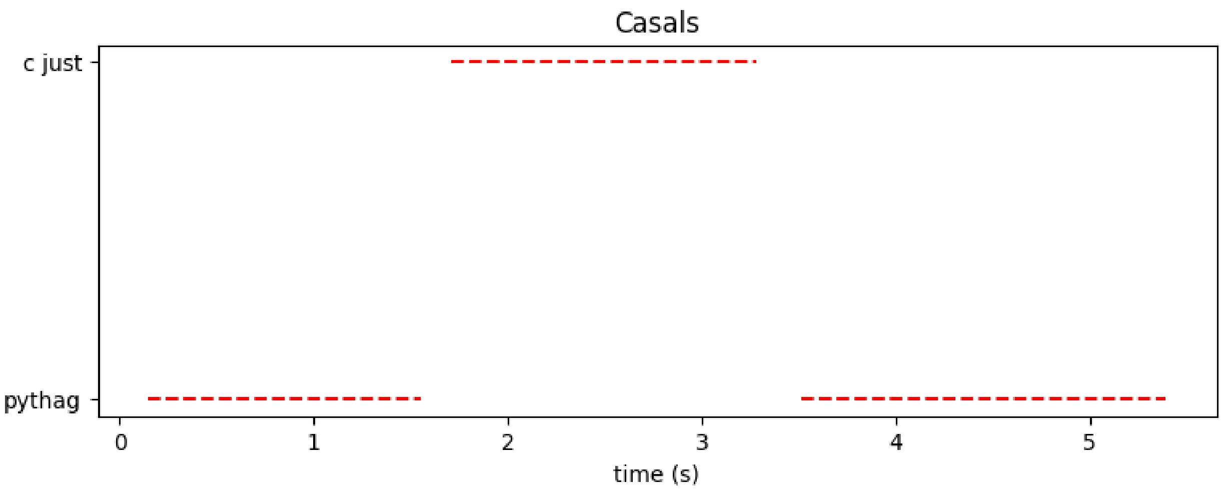


Figure 5.9: Output of Casals opening phrase with only c major stencil..

Table 5.11: Casals Opening Phrase with only c major stencil..

Tuning System	**Total**	**Percentage**
pythag	6	66.67
c just	3	33.33

With all stencils, the first Pythagorean detection gets replaced in favor of the E just stencil. This does not make much intuitive sense other than to ascribe the switch due to small frequency differences and the sensitivity of the detection algorithm once again.

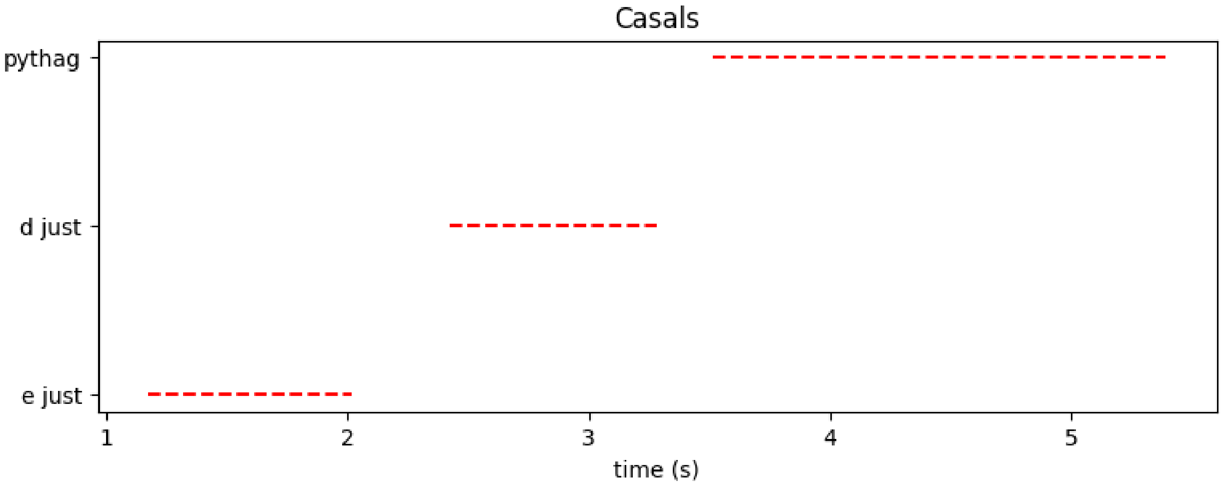


Figure 5.10: Output of Casals opening phrase with all just stencils..

Table 5.12: Casals Opening Phrase with all just stencils..

Tuning System	Total	Percentage
e just	3	42.86
d just	2	28.57
pythag	2	28.57

5.2.4 Ethan Cobb

The relatively equal division between equal temperament and just intonation may confirm the author's frequent practice with an equal tempered tuner as well as justly tuned double stops. Or, the algorithm may suggest that the author's playing is somewhat inconsistent. The first equal temperament detection is between the first and second notes, a departure from Casals as we saw before.

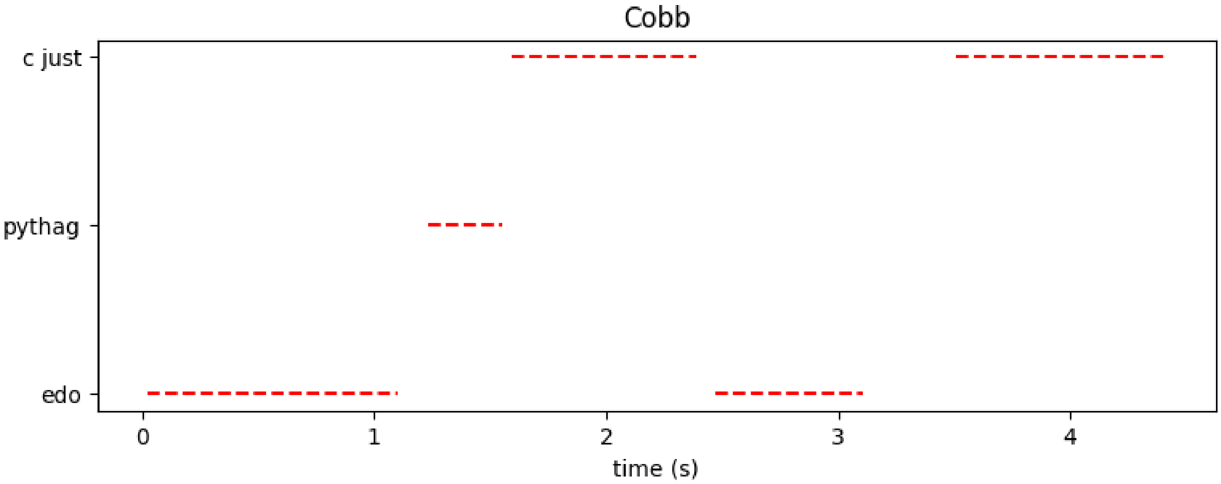


Figure 5.11: Output of Cobb opening phrase with only c major stencil..

Table 5.13: Cobb Opening Phrase with only c major stencil..

Tuning System	Total	Percentage
edo	4	40.0
pythag	2	20.0
c just	4	40.0

With all stencils, Pythagorean is replaced with the G just stencil. The final G just classification would be equivalent to a C just classification since the algorithm is looking at the last two notes of the phrase - G2 followed by C2, a perfect fifth. This interval would be equivalent in both stencils.

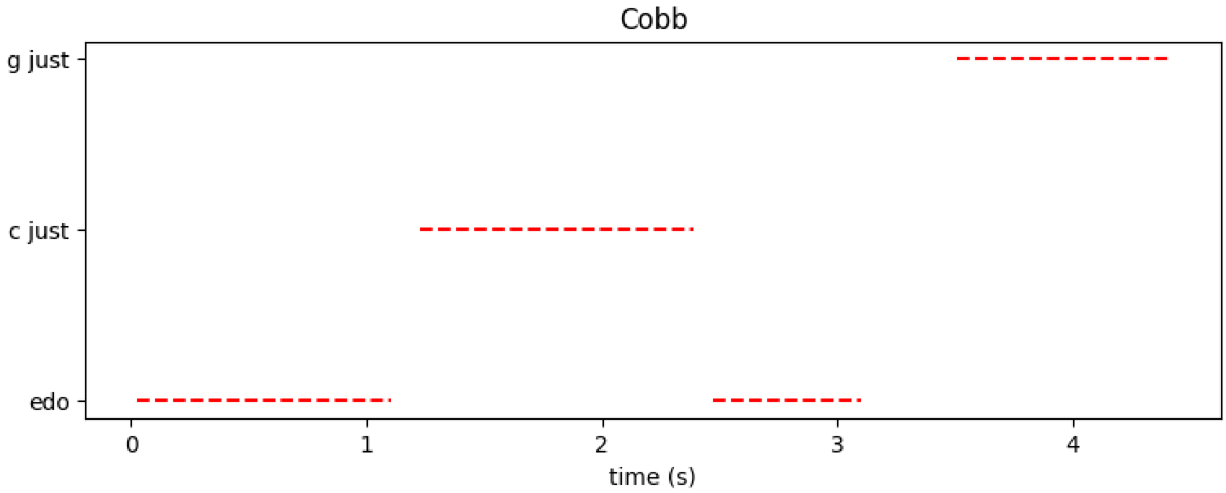


Figure 5.12: Output of Cobb opening phrase with all just stencils..

Table 5.14: Cobb Opening Phrase with all just stencils..

Tuning System	Total	Percentage
edo	4	40.0
c just	4	40.0

Continued on next page

Table 5.14 – continued from previous page

Tuning System	Total	Percentage
g just	2	20.0

5.2.5 Ralph Kirshbaum

Kirshbaum is the first instance of an entire classification of one system - in this case Pythagorean. Matching to the audio output, the only notes detected are part of the downward scale suggesting that scalewise motion like this may lend itself more often to Pythagorean intonation.

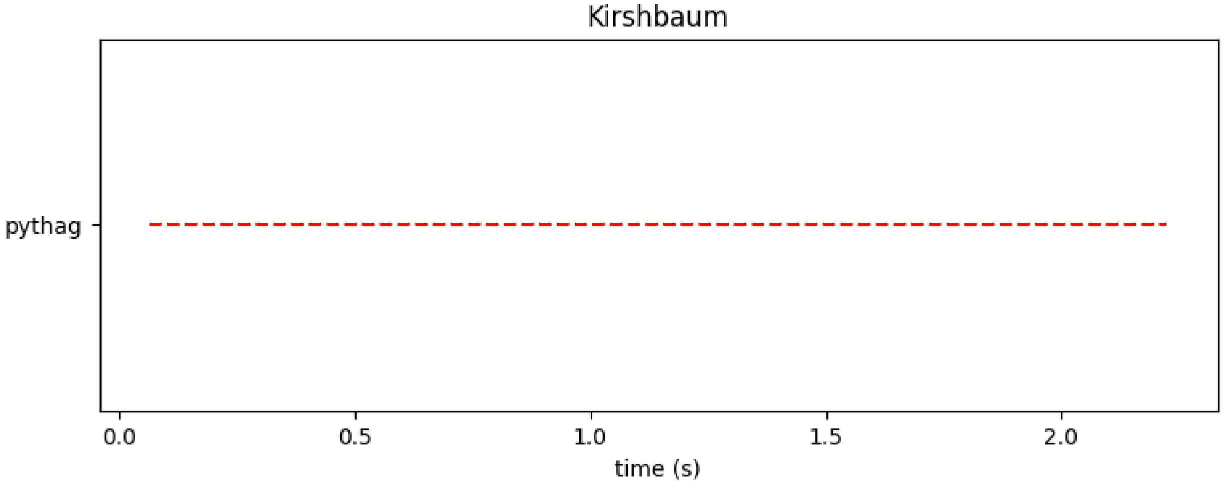


Figure 5.13: Output of Kirshbaum opening phrase with only c major stencil.

Table 5.15: Kirshbaum Opening Phrase with only c major stencil.

Tuning System	Total	Percentage
pythag	7	100.0

Allowing all stencils does not change the output dramatically; Kirshbaum still uses Pythagorean the majority of the opening phrase.

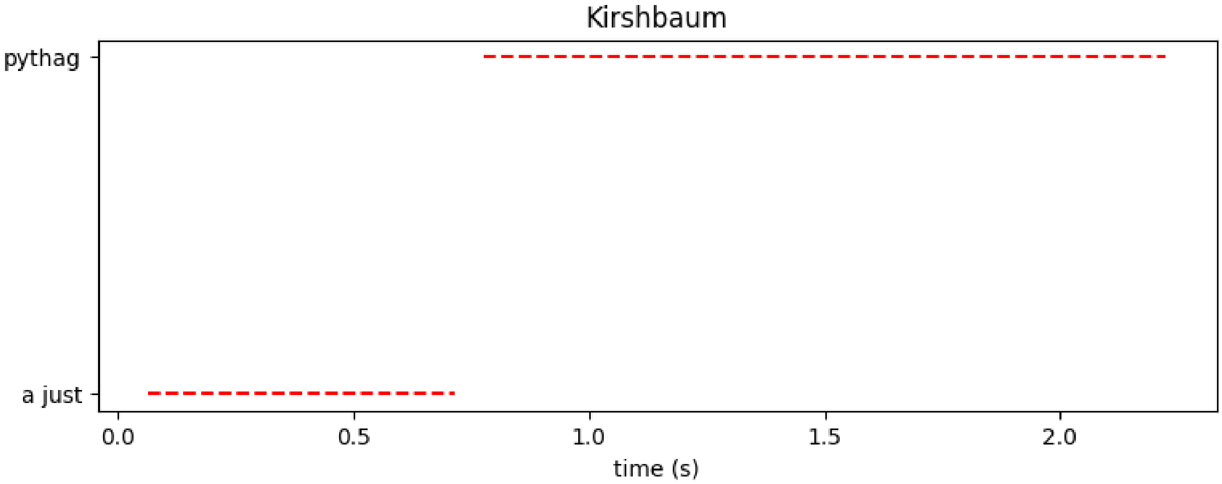


Figure 5.14: Output of Kirshbaum opening phrase with all just stencils.

Table 5.16: Kirshbaum Opening Phrase with all just stencils.

Tuning System	Total	Percentage
a just	2	28.57
pythag	5	71.43

5.2.6 Yo Yo Ma

Ma plays with Pythagorean intonation the majority of the opening phrase and also uses C just. This is entirely what we would predict.

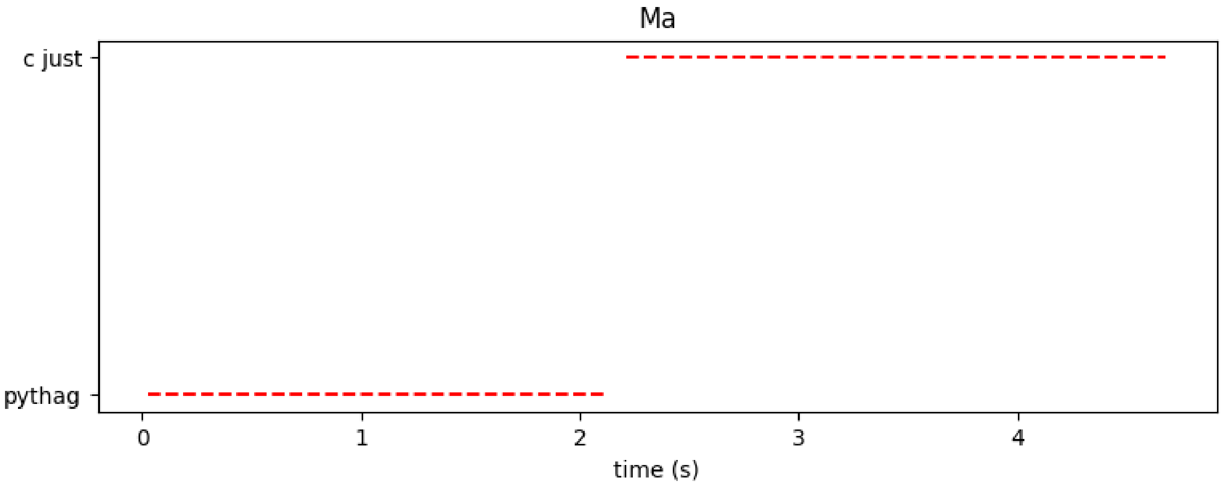


Figure 5.15: Output of Ma opening phrase with only c major stencil.

Table 5.17: Ma Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
pythag	7	77.78
c just	2	22.22

Interestingly, when all stencils are allowed, practically all of the Pythagorean classification is abandoned in favor of the E and D just stencils.

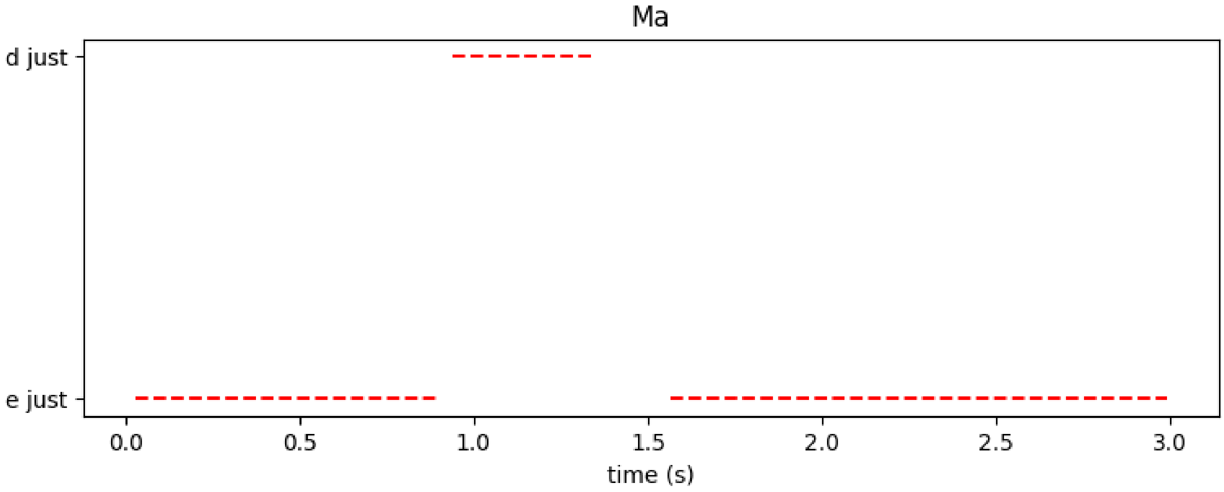


Figure 5.16: Output of Ma opening phrase with all just stencils.

Table 5.18: Ma Opening Phrase with all just stencils.

Tuning System	**Total**	**Percentage**
e just	5	71.43
d just	2	28.57

5.2.7 Misha Maisky

Maisky by a majority plays with mostly equal temperament but still uses the other two systems. Like the author, the first two notes are played with equal temperament. The

Pythagorean classification is likely not entirely correct because Maisky happens to place the G in beat two of the first measure quite high. This is most definitely because his finger happened to land there on that day the audio was recorded. With a high degree of certainty, if Maisky were asked to prepare and play only that same note without context, it would be tuned with his open G string to accord with G major just intonation. This points to an important caveat to the entire set of results - there is a degree of uncertainty in all of the output because cellists cannot perfectly perform a piece multiple times hitting the same frequency for every note every time; it is physically impossible.[1]

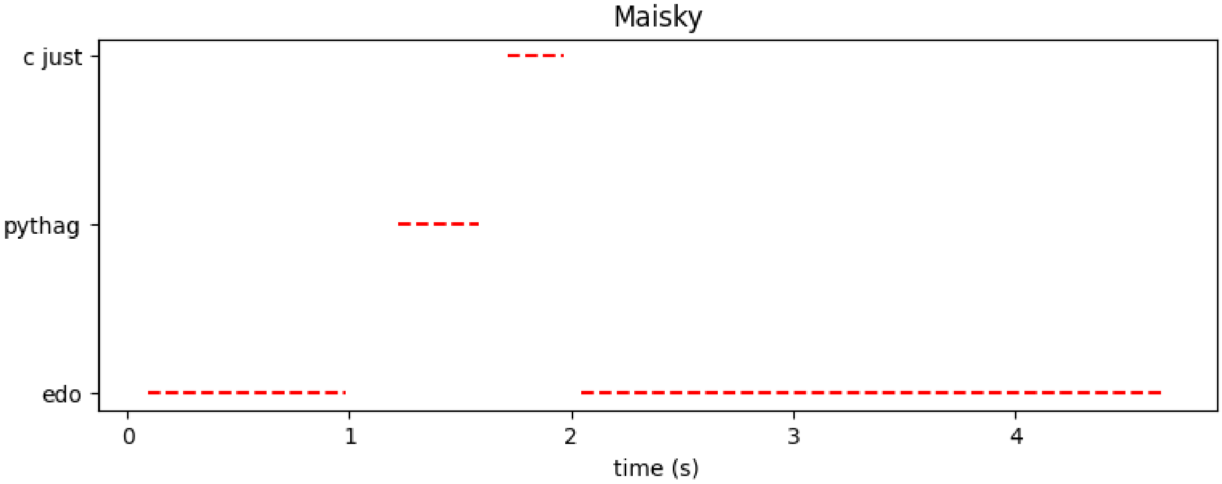


Figure 5.17: Output of Maisky opening phrase with only c major stencil.

Table 5.19: Maisky Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
edo	5	55.56
pythag	2	22.22
c just	2	22.22

[1] Perhaps with only a small number of notes - two notes, for instance, would this be possible. But even still it would be impossible to confirm with 100% confidence because it goes without saying that no quantity can be measured with infinite precision

With all stencils, Maisky still is classified as equal temperament, forming a stronger case for the labeling.

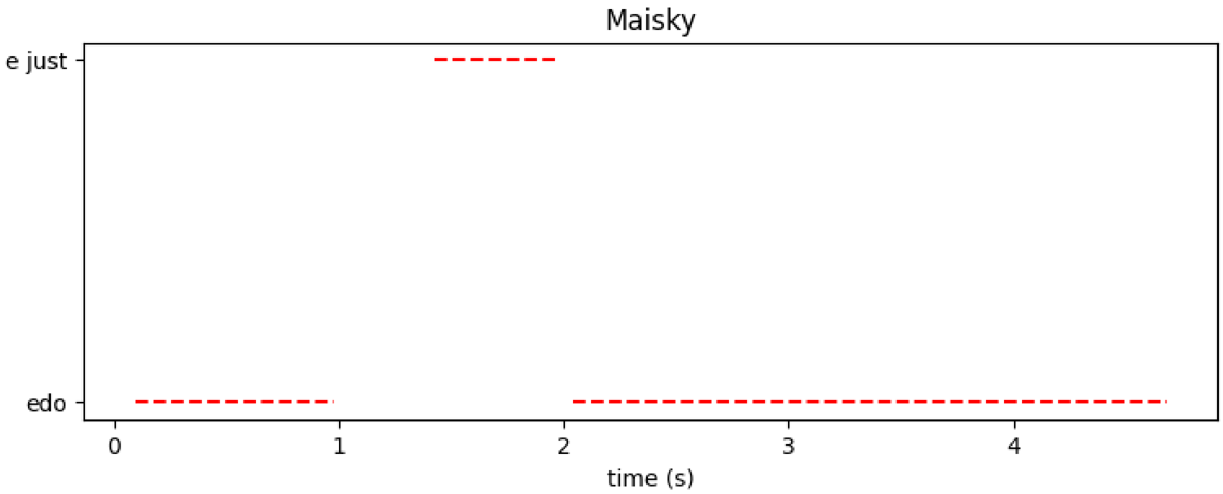


Figure 5.18: Output of Maisky opening phrase with all just stencils.

Table 5.20: Maisky Opening Phrase with all just stencils.

Tuning System	Total	Percentage
edo	5	62.5
e just	3	37.5

5.2.8 Jean-Guihen Queyras

Queyras uses equal temperament a majority of the time in addition to Pythagorean. This is somewhat of a surprise since Queyras is known for his somewhat baroque interpretations of the cello suites in addition to other repertoire. The recording of his used for these results importantly was not one in which his strings were tuned down as is usually the case when cellists attempt to play Bach in a more “historically-informed” manner.

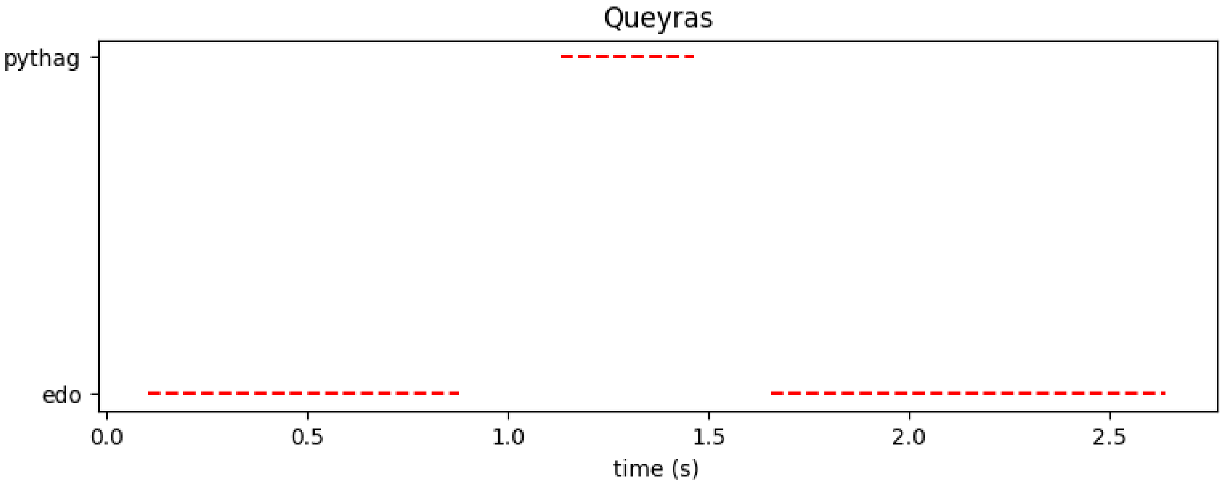


Figure 5.19: Output of Queyras opening phrase with only c major stencil.

Table 5.21: Queyras Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
edo	6	75.0
pythag	2	25.0

With all stencils, equal temperament is abandoned in favor of just intonation. The two notes detected as Pythagorean correspond to the notes F and E in between beats two and 3. For this reason, it could be that Queyras is intentionally making this interval smaller so to accord to a Pythagorean semitone ratio. This is also reminiscent of the music theoretical concept that the third degree of the scale ('submediant') has a tendency to resolve to the fourth degree of the scale (or 'subdominant').

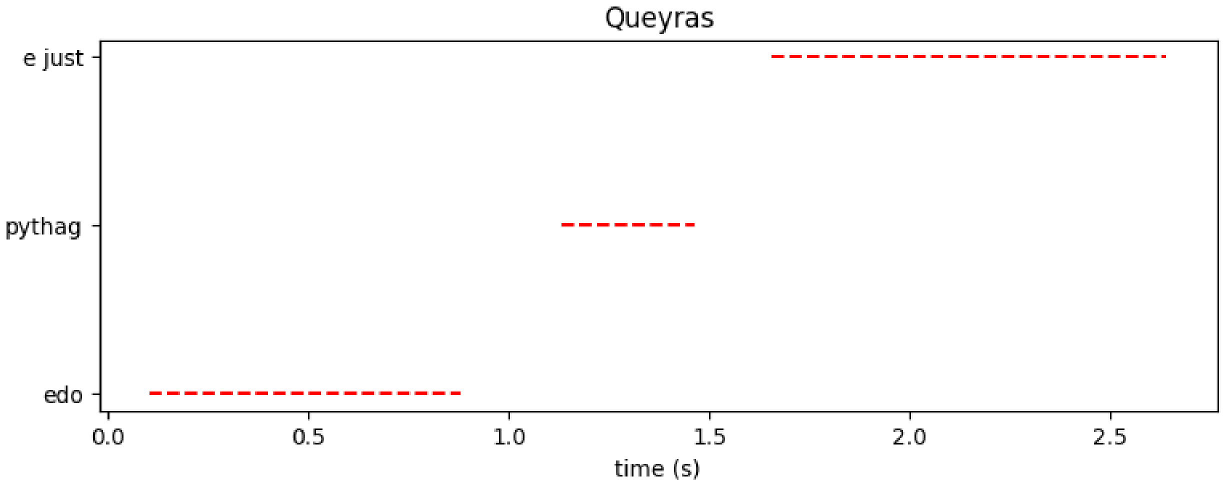


Figure 5.20: Output of Queyras opening phrase with all just stencils.

Table 5.22: Queyras Opening Phrase with all just stencils.

Tuning System	**Total**	**Percentage**
edo	2	25.0
pythag	2	25.0
e just	4	50.0

5.2.9 Mstislav Rostropovich

With only the C just stencil, somewhat surprisingly, Rostropovich mostly uses equal temperament.

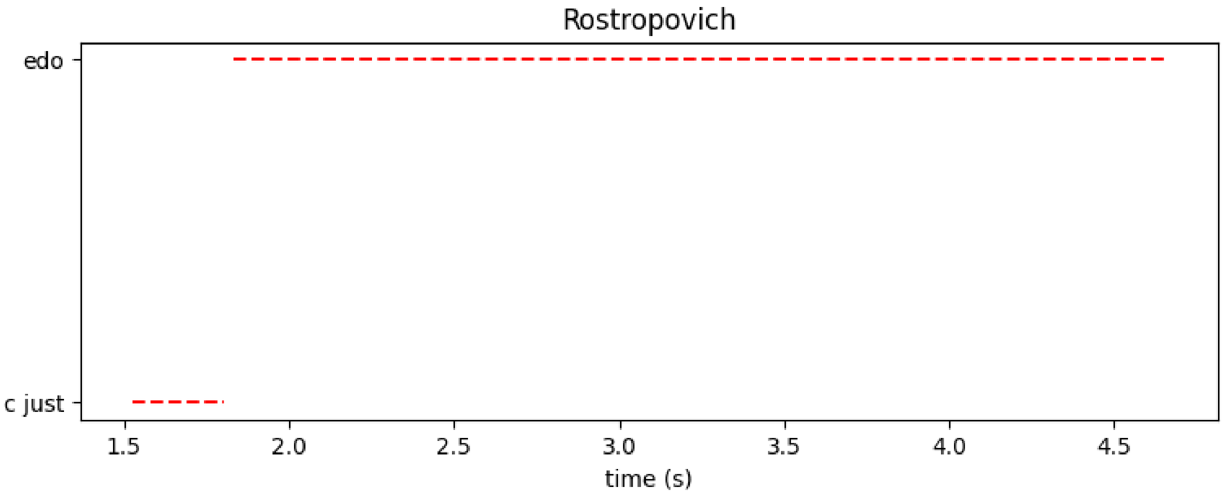


Figure 5.21: Output of Rostropovich opening phrase with only c major stencil.

Table 5.23: Rostropovich Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
c just	2	25.0
edo	6	75.0

With all stencils, the previous classification is replaced with an overwhelming just classification, particularly the A just stencil. Rather than for a music theoretical reason, this switch is more likely because the frequencies happened to fall closer on average to the six stencils than the equal tempered values.

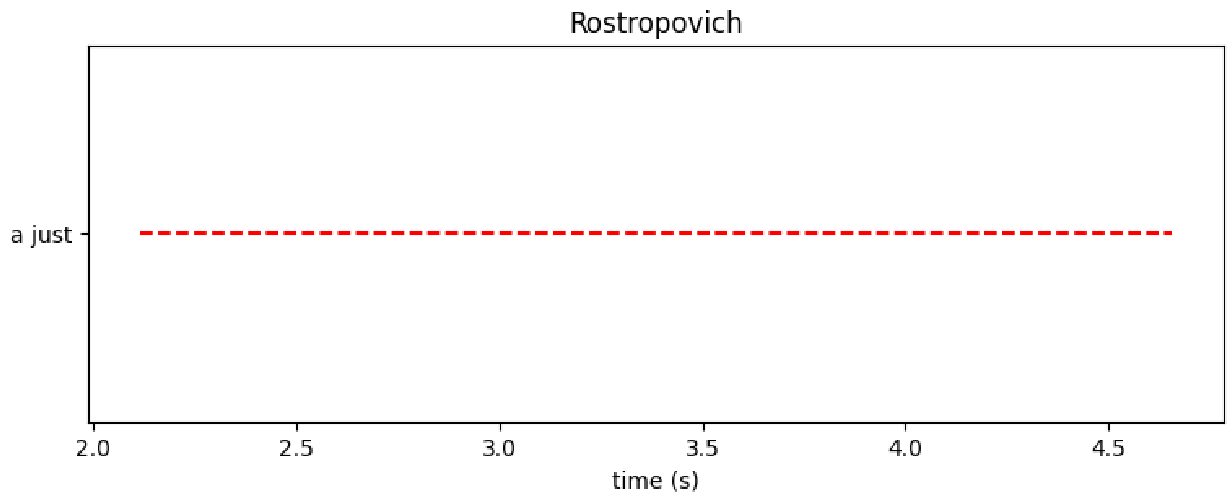


Figure 5.22: Output of Rostropovich opening phrase with all just stencils.

Table 5.24: Rostropovich Opening Phrase with all just stencils.

Tuning System	**Total**	**Percentage**
a just	5	100.0

5.2.10 Heinrich Schiff

With only the C just stencil, Schiff is equally divided between equal temperament and Pythagorean intonation. The Pythagorean classification is for the last five notes of the first measure. This could be due to the minor third interval between E2 and G2. Thirds in general are where Pythagorean has the most uniqueness compared to the other intonation systems.

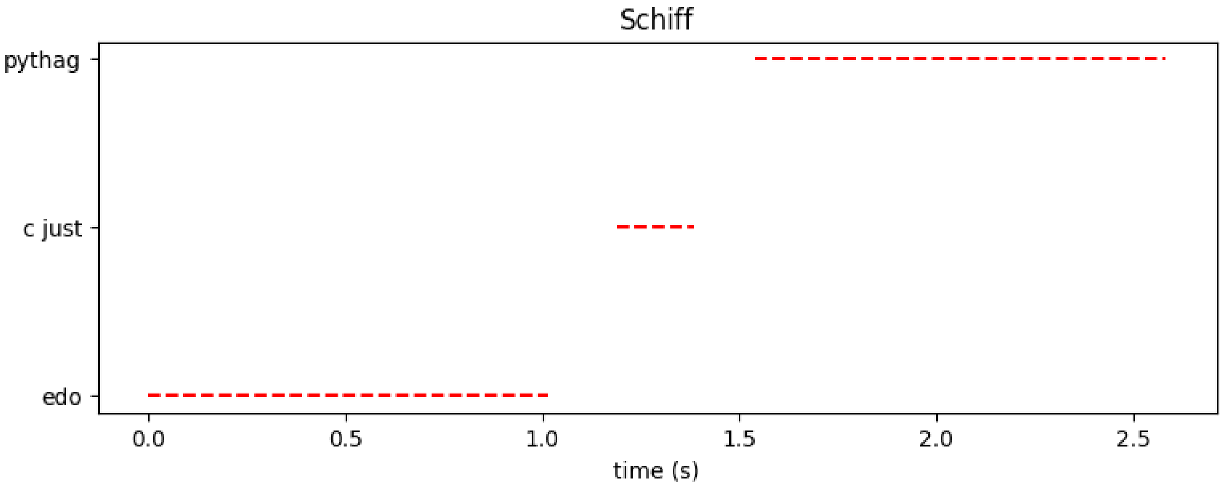


Figure 5.23: Output of Schiff opening phrase with only c major stencil.

Table 5.25: Schiff Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
edo	3	37.5
c just	2	25.0
pythag	3	37.5

With all just stencils, Schiff appears to play a majority with just intonation, particularly the E stencil which we have seen multiple times already.

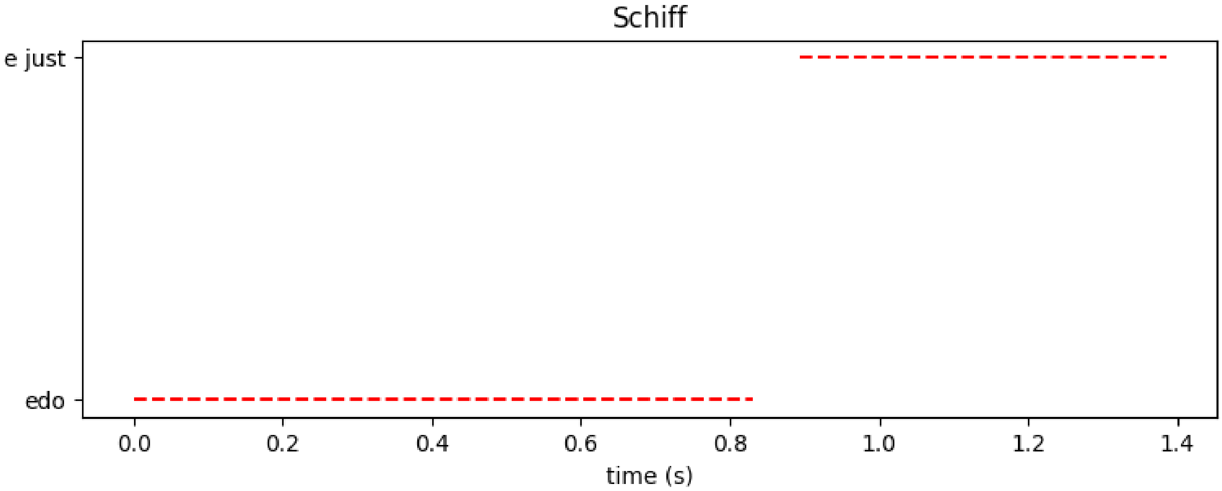


Figure 5.24: Output of Schiff opening phrase with all just stencils.

Table 5.26: Schiff Opening Phrase with all just stencils.

Tuning System	**Total**	**Percentage**
edo	2	33.33
e just	4	66.67

5.2.11 Jan Vogler

Vogler uses just intonation for the majority of the phrase although due to the noise and reverberation in the audio, the algorithm is only able to detect six total notes. This gives less certainty to the classification.

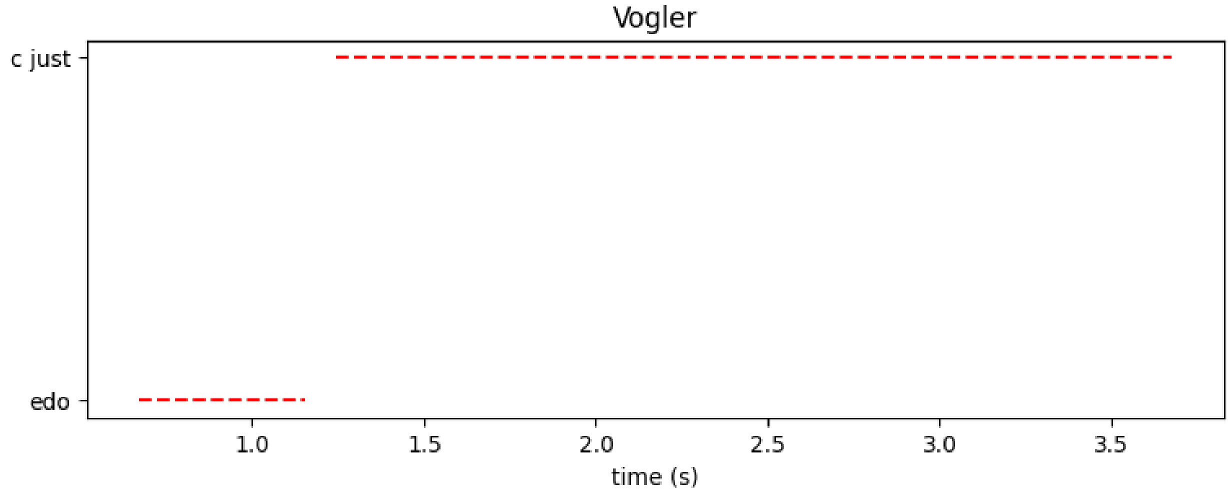


Figure 5.25: Output of Vogler opening phrase with only c major stencil.

Table 5.27: Vogler Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
edo	2	33.33
c just	4	66.67

With all stencils, Vogler is classified entirely as C just, what we would predict.

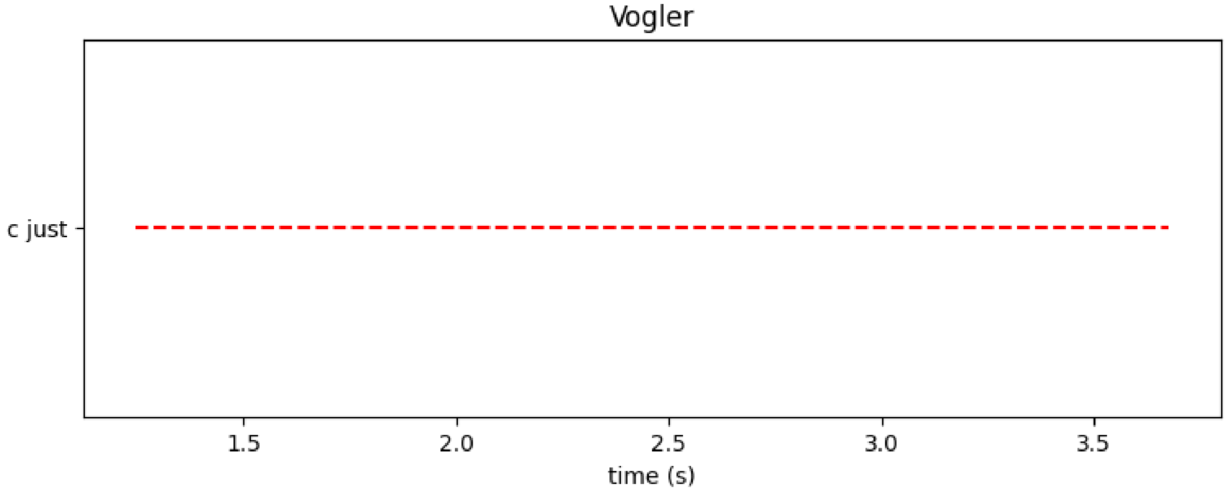


Figure 5.26: Output of Vogler opening phrase with all just stencils.

Table 5.28: Vogler Opening Phrase with all just stencils.

Tuning System	**Total**	**Percentage**
c just	4	100.0

5.2.12 Pieter Wispelwey

Wispelwey uses all three systems but interestingly uses Pythagorean in the start of the opening phrase. This may reflect the first note-second note succession (C4 - B3) which is an instance of tonic followed by leading tone, albeit downward descending. Even though the direction is flipped compared to normal instances of raised leading tones followed by tonics, this could be Wispelwey's intention. Wispelwey though uses the most extreme version of baroque tuning of any of the cellists analyzed with an A4 estimate of 394 Hz! One may argue it is no longer meaningful to think in terms of the original key of the phrase (C major) for the classification; naturally, there is a lot more variation in terms of the cent difference between notes in the phrase for the various intonation systems and is thus difficult to compare with the other cellists who especially don't attempt to tune down (Carr, Casals, Cobb, Kirshbaum, Ma, Maisky, Queyras, Rostropovich, Schiff, Vogler).

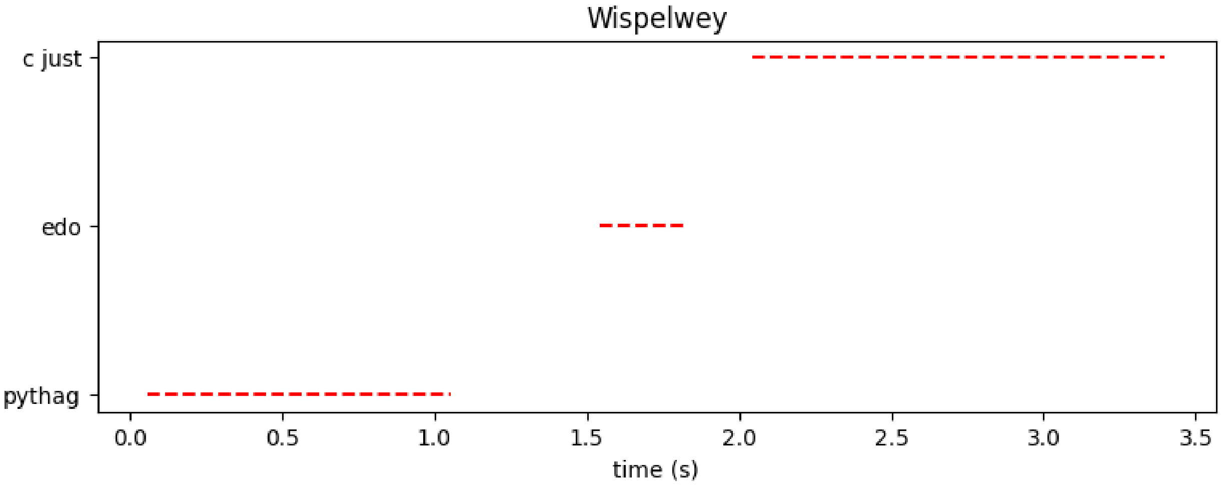


Figure 5.27: Output of Wispelwey opening phrase with only c major stencil.

Table 5.29: Wispelwey Opening Phrase with only c major stencil.

Tuning System	**Total**	**Percentage**
pythag	3	37.5
edo	2	25.0
c just	3	37.5

While the distribution of systems is still fairly equal, the algorithm outputs just as the most probable system, again what we would expect from a historically-informed cellist.

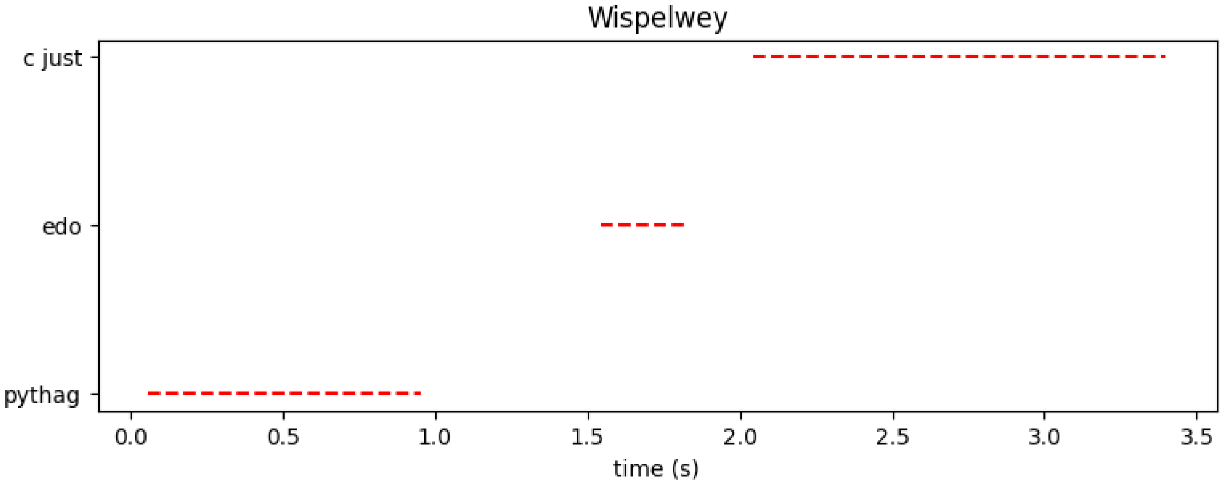


Figure 5.28: Output of Wispelwey opening phrase with all just stencils.

Table 5.30: Wispelwey Opening Phrase with all just stencils.

Tuning System	Total	Percentage
pythag	2	28.57
edo	2	28.57
c just	3	42.86

5.3 Full Cello Suite No. 3 Prelude

We begin our discussion of the results on the complete prelude with some remarks on overall trends in the analysis. While communicating results about the overall most likely intonation system for a given piece is not as meaningful as a more local time-based analysis (which we will do after in section 5.3.2), it presents clear trends obeyed by the data that are worth exploring.

5.3.1 Overall Statistics

When forcing the algorithm to only use the C just stencil, i.e., to do a three-class classification task, we get an extremely interesting result: the overwhelming majority of cellists play with Pythagorean intonation. Most would perhaps expect the algorithm to answer just intonation (which is indeed the case when we allow for all possible just stencils to be used) but this could suggest that the number of modulations within the Third Suite Prelude provide enough variation so to prevent the algorithm from getting 'stuck' on the C just stencil. Since the frequency differences in each of the three stencils are more distinct when dealing with only a three-class classification problem, this may support the notion that Pythagorean is indeed the most probable intonation system used among the twelve cellists. Casals also has the third largest share of Pythagorean intonation (which may in fact be an underestimate due to the quality of the recording) which we also hoped to be reflected in the output.

When all stencils are pooled together into one just estimate, as we would expect, the majority of cellists play with just intonation. Rostropovich leads the crowd with an overwhelming 90.24% share of notes falling under a just classification. Even more impressive is that he is still classified as just when only using the c major stencil. He may very well be aligning much of his intonation according to C major just intonation. Casals's bias towards Pythagorean is confirmed again on the entire prelude albeit not the strongest share which is somewhat surprising. Carr appears to lead the Pythagorean crowd in this regard. The overall just intonation majority makes sense for two large reasons:

1. For baroque music, we would expect an overwhelming amount of just intonation as this was mostly the tradition during this era and has been taught as a "best tuning practice" for the Bach Cello Suites [16].

2. When pooling all six key stencils into one just stencil, the probability of a just classification is naturally higher. In fact, on average, we'd expect about $\frac{6}{8} = .75$ percent of the piece to be classified as just intonation in this definition and indeed the numbers

are close to this amount. The average just share among the majority just cellists is actually 68.19 %.

Using C Just Stencil		
	Percent Just	**Most Probable Intonation System**
Bylsma	29.2	pythag
Carr	39.07	pythag
Casals	28.9	pythag
Cobb	19.46	pythag
Kirshbaum	15.1	pythag
Ma	29.17	pythag
Maisky	24.12	pythag
Queyras	16.35	pythag
Rostropovich	48.41	just
Schiff	21.21	pythag
Vogler	28.7	pythag
Wispelwey	29.5	pythag

Table 5.31: Fraction of notes played with just intonation and overall most probable intonation system for all 12 cellists only using the C just stencil

Using All Stencils		
	Percent Just	**Most Probable Intonation System**
Bylsma	63.28	just
Carr	81.7	just
Casals	83.19	just
Cobb	67.25	just
Kirshbaum	57.14	just
Ma	75.82	just
Maisky	76.21	just
Queyras	62.33	just
Rostropovich	90.24	just
Schiff	69.68	just
Vogler	74.93	just
Wispelwey	80.4	just

Table 5.32: Fraction of notes played with just intonation and overall most probable intonation system for all 12 cellists using all just stencils

5.3.2 Analysis of Yo Yo Ma Suite 3 Prelude

Perhaps the most famous of all cellists considered is Yo Yo Ma who is especially known for his brilliant performances of the Cello Suites. We present a more localized analysis of the results of the algorithm and speculate on the choices of intonation system. Plots of the complete output using both only the C just stencil and all key stencils are shown below as well as tables detailing the results of the tuning system detection. The rest of the data for the other 11 cellists is included in the appendix A.

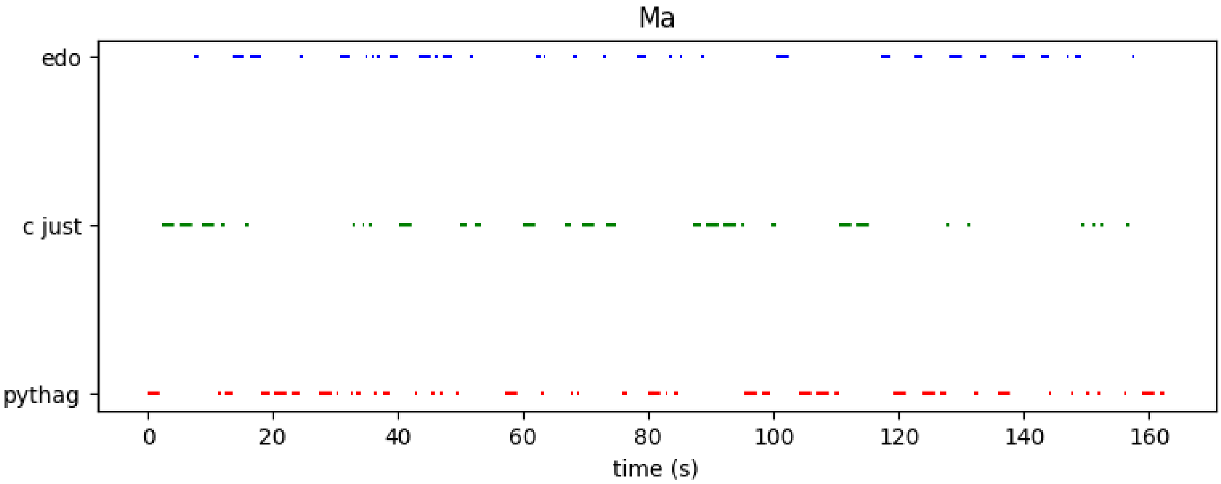


Figure 5.29: Output of Ma Full Suite 3 Prelude with only c major stencil.

Table 5.33: Ma Full Suite 3 Prelude with only c major stencil.

Tuning System	Total	Percentage
pythag	183	42.36
c just	126	29.17
edo	123	28.47

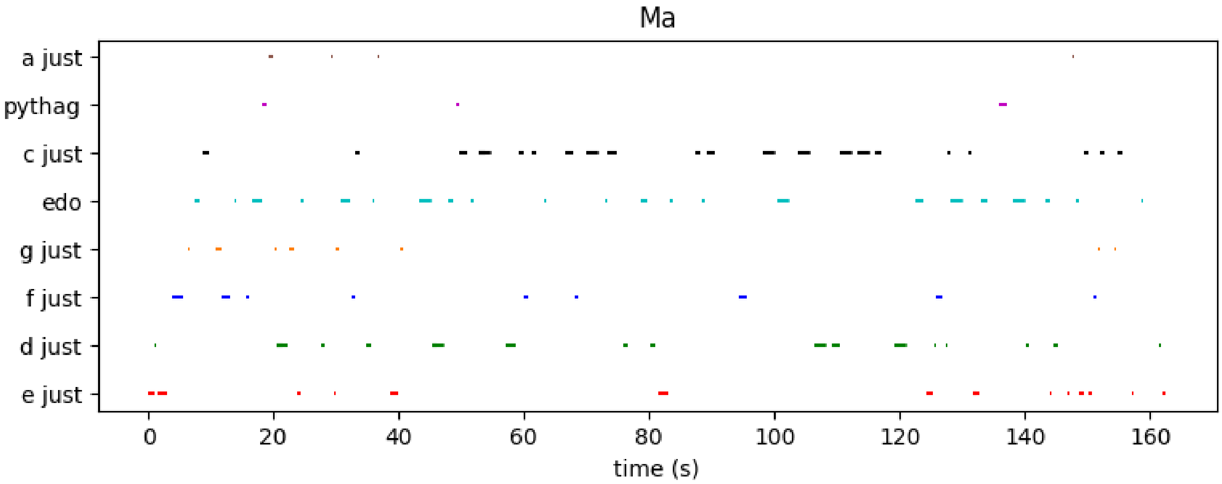


Figure 5.30: Output of Ma Full Suite 3 Prelude with all just stencils.

Table 5.34: Ma Full Suite 3 Prelude with all just stencils.

Tuning System	**Total**	**Percentage**
e just	48	14.33
d just	56	16.72
f just	29	8.66
g just	20	5.97
edo	73	21.79
c just	91	27.16
pythag	8	2.39
a just	10	2.99

As we saw in figure 5.18, Ma seems to express a tendency towards just intonation especially when involving melodic lines in the key of C major. This is indeed captured in the data for much of the beginning of the Prelude which contains exactly this.

The algorithm also detects just intonation, albeit in the D just stencil for the following passage. The determination of D as the stencil is once again likely equivalent with a number of other stencils for the given passage, such as A. In each measure, a note is repeated in

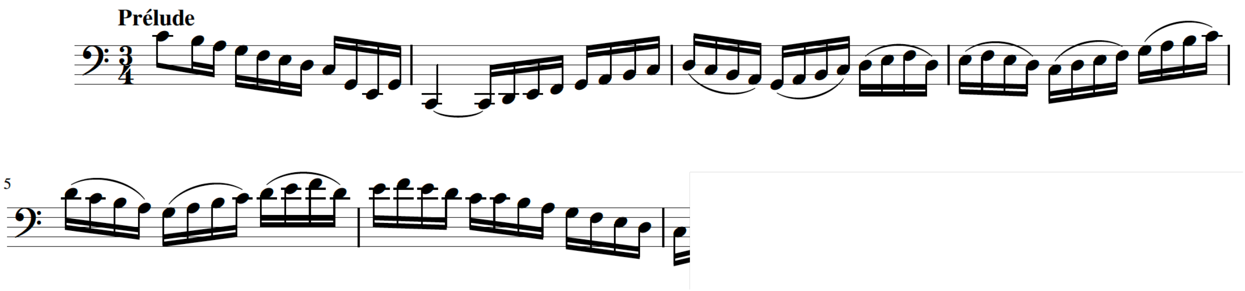


Figure 5.31: Instance of just intonation

the same rhythmic form - in measure 21, it is the note A; in measure 22, C, and in 23, D before descending to the next phrase. This repeated structure may warrant the use of just intonation in relation to the repeated notes as it may be argued that these notes serve as a kind of pedal point musical structure. If this is the case, Ma has the option of tuning the notes decorating the repeated note (for instance D# and B in mm 21) according to just interval ratios. The last three notes of the phrase get classified as Pythagorean - this could be intentional on Ma's part to serve as some sort of leading quality to the next phrase or simply due to numerical error. The intervals correspond to two major seconds which have the same interval ratio in both just and Pythagorean intonation.

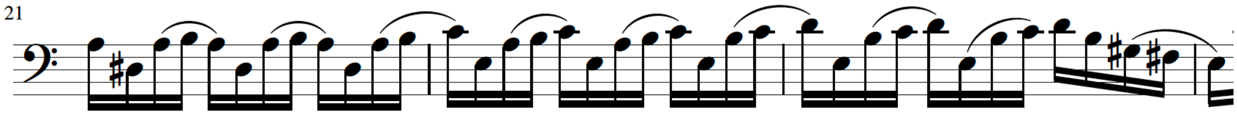


Figure 5.32: Instance of just intonation

The algorithm detects the following passage as just for roughly the first measure and then equal temperament for the second. While the passage corresponds to transition material in the latter half of the prelude, the Pythagorean detection may be due to the repeated C# note in the first measure, which has a leading tone quality in relation to the D minor resolution in the following measure. Ma may be intentionally raising this note to increase the sense of resolution when the leading tone is finally resolved.

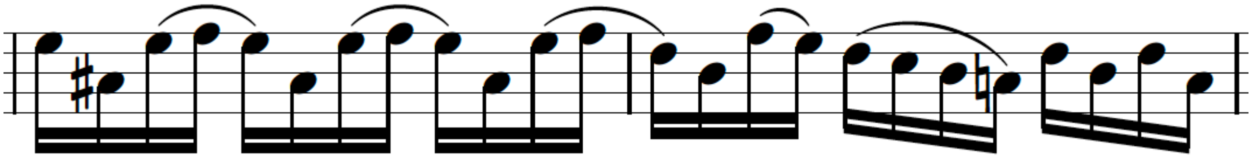


Figure 5.33: Instance of equal temperament and Pythagorean intonation

While there is some variation of the detection within this next passage (Pythagorean, equal temperament), the algorithm predicts a just intonation majority, which is what we would hope for and expect. Cellists often practice this especially difficult passage by playing the notes as double stops tuning to the cello g string. This is in effect tuning according to just intonation. The passage is a clear instance of pedal point which is usually a sign that just intonation will be involved.

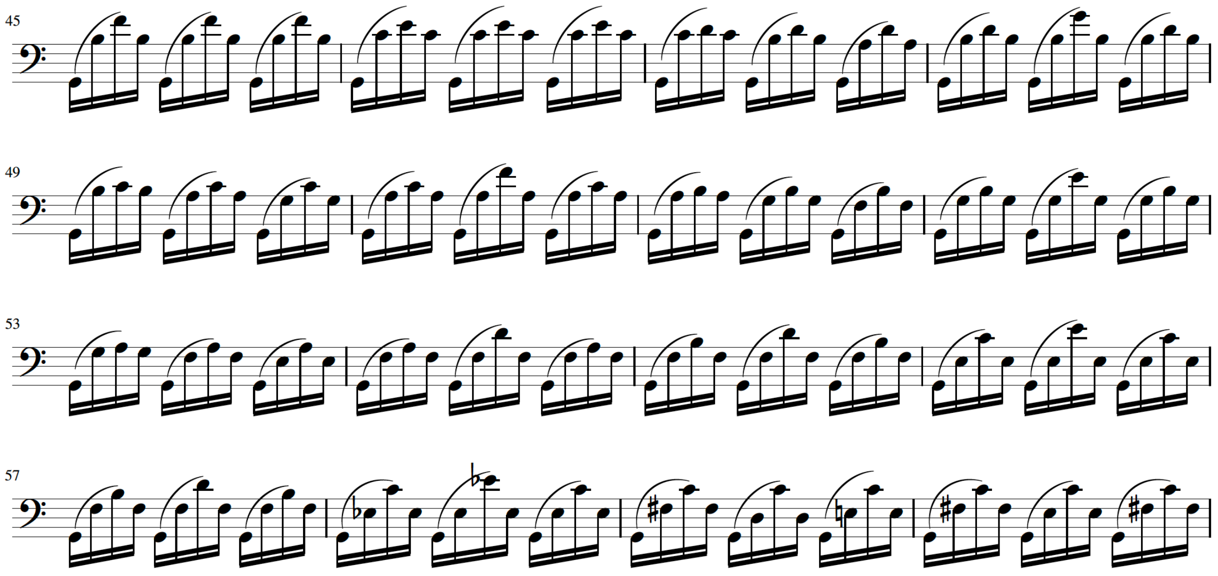


Figure 5.34: Instance of just intonation

The final descending scale and arpeggio in the penultimate measure of the Prelude is classified as Pythagorean according to the algorithm. This passage is an exact repeat of the opening measure of the Prelude which we observed before. When only the C just stencil was used, the algorithm predicted Pythagorean but otherwise predicted just when all stencils were used. To this end, either Ma intentionally is using Pythagorean to create a greater sense of fulfillment especially with a high Pythagorean major third corresponding to the low E2 or, his fingers did not land exactly how he intended, or the frequency differences between systems were simply not significant enough to produce a reliable classification.

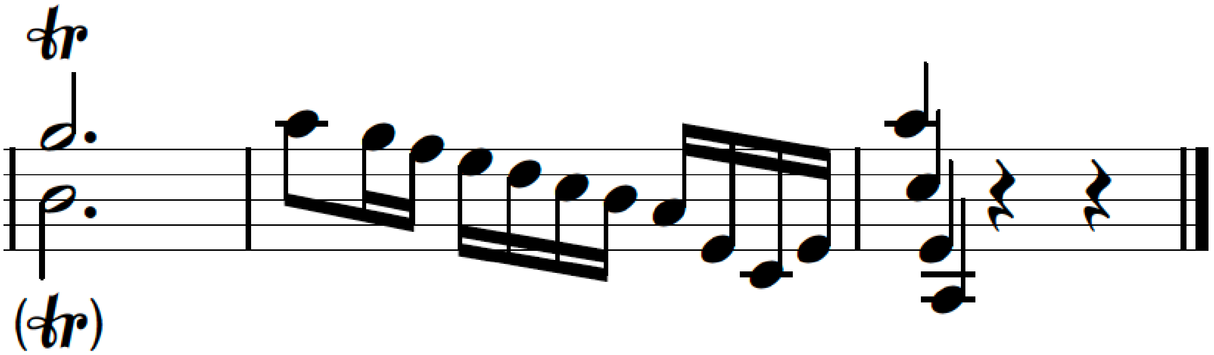


Figure 5.35: Instance of Pythagorean intonation

5.4 Discussion

As was previously shown, large overall classifications of intonation system depend largely on whether or not all just intonation key stencils are used as this naturally tends to skew the classifications in the just direction. The choice of using only one particular key stencil or all should indeed depend on the kind of piece one wants to analyze. If the piece is atonal and doesn't lend itself to a classical harmonic analysis, then applying a just intonation stencil perhaps is not as meaningful. This is why the Third Suite Prelude was chosen in particular - there are not too many frequent key modulations in the piece so we would hope to observe some kind of consistency in the classifications.

In the other direction, one may argue that just intonation does not always follow the kind of key-following template as we've described; many cellists tune notes according to the cello open strings or notes that directly precede or follow a given note with double stops. They are not always necessarily tuning justly by following the key-centric approach as the algorithm is employing. To this end, it may be more fitting to test for just intonation solely by examining adjacent frequency ratios and their closeness to just intervals. This however is an extremely local approach and may fail to provide larger intonation trends. In performance, cellists also may be informed by pitch memory meaning that for any given note, intonation may be determined by the intervals created by several notes played before; it quickly becomes extremely computationally costly [17].

In general there may exist bias towards just intonation because it naturally brings with it many physical phenomena humans find pleasing. Aligning two frequencies such that their partials align produces resonance, harmony, and purity, all physically realized by the response of the instrument. One need not know the mathematics of Pythagoras to be able to feel when the instrument rings when an octave is played perfectly in tune, to see when the A string sympathetically vibrates when a D tuned perfectly with the D string is played, or hear a higher partial when the same note is played. To many in the baroque era and earlier, this resonance and purity created by just intonation affirmed it to be the "God-Given" basis of music [18]. Thus as baroque cellists, Wispelwey and Bylsma would hope to play in just intonation the majority of the time. It should also be noted that Pythagorean is indeed a strict subset of just intonation; every interval is derived from the just intonation ratio of the perfect fifth and the major second, perfect fourth, and perfect fifth are exactly the same in both systems (see section 1.2.2. Thus even if a cellist is classified as Pythagorean for a section of the audio, the distinction is not entirely black and white.

Chapter 6

Final Remarks

6.1 Conclusion

In performance, musicians use a combination of tuning systems that depend on a number of factors including musical and harmonic context. While work has been done in estimating the temperament of a fixed-pitch instrument such as the harpsichord, no work has been done in detecting instances of tuning systems in audio recordings, especially cello recordings of the Bach Cello Suites where intonation is critical.

I proposed several algorithms and signal to detect and classify three possible intonation systems - just, equal-tempered, and Pythagorean - in recordings of the Third Bach Cello Suite Prelude, which I chose for its musical and harmonic complexity. First, timestamps and fundamental frequency estimates were obtained for every note based on a note detection algorithm which utilizes confidence estimates outputted by CREPE. Then, each note was assigned an estimated tuning system label by calculating a unique probability for each tuning system combined with a centered moving average process. Finally, sequences of intonation systems were identified by splitting the intonation system-labeled time series at instances where consecutive labels differ. One final sweep of the output was taken to remove any sequences with a number of notes less than a given threshold and combine any consecutive equivalently-labeled sequences that are within a given time threshold. The results of the algorithm illustrate that cellists most often use either Pythagorean or just intonation depending on whether only the C just stencil is used or not in the algorithms. Physical and musical theoretical reasons may explain the tendency towards these particular systems.

6.2 Future Directions

There are many aspects of the pipeline that can be improved or substituted with more sophisticated methods. Fundamental frequency estimation is the foundation of intonation analysis and while CREPE is very promising and simple to use, a less automated approach involving techniques such as Dixon, Mauch, and Tidhar's frequency detection algorithm may prove to be more successful; any increase in frequency resolution can only serve to improve the accuracy and reliability of the final output [6]. Further, the analysis of polyphonic audio data including chords and double stops is an obvious point of further research as these kinds of musical instances almost always are instances of just intonation. If equipped to handle polyphonic intonation system detection, the model would have tremendous applications in smaller musical group settings such as chamber music where conflicts over intonation occur quite frequently. Having an objective, numerical model of intonation would ease some of the hearing conflicts that stem from the subjectiveness of intonation.

This thesis has effectively created an algorithmic and mathematically-based way of generating a training set of intonation system labels from raw audio data. Fascinating insight may be gained if a machine learning model is fed this resulting data and asked what sorts of music theoretical or physical features of the audio are most influential in determining the choice of intonation system in a piece of music? The range of data can be expanded to include more cellists, the rest of the Bach Cello Suites, and much of the other solo repertoire so to provide enough training data for machine learning.

In addition to providing insight into how musicians use intonation, this research offers musicians a tool for understanding and assessing their intonation by providing an objective measure of intonation. It also offers a way to gain insight into the history of intonation and how, depending on the era of musical composition and other factors, intonation patterns change. When applied to the history of recorded music, it provides a new tool for analyzing the work of master performers and great performances. This research thus has the potential to inspire computationally-driven intonation analysis research within musicology and music history.

Appendix A

Full Results

We now provide complete data for all twelve cellists of the Third Suite Prelude. As a reminder, chordal content was trimmed from the audio so to allow for monophonic fundamental frequency detection. Below are two tables with the overall percentage of notes played with just intonation and the overall most probable intonation system for each cellist:

A.1 Anner Bylsma

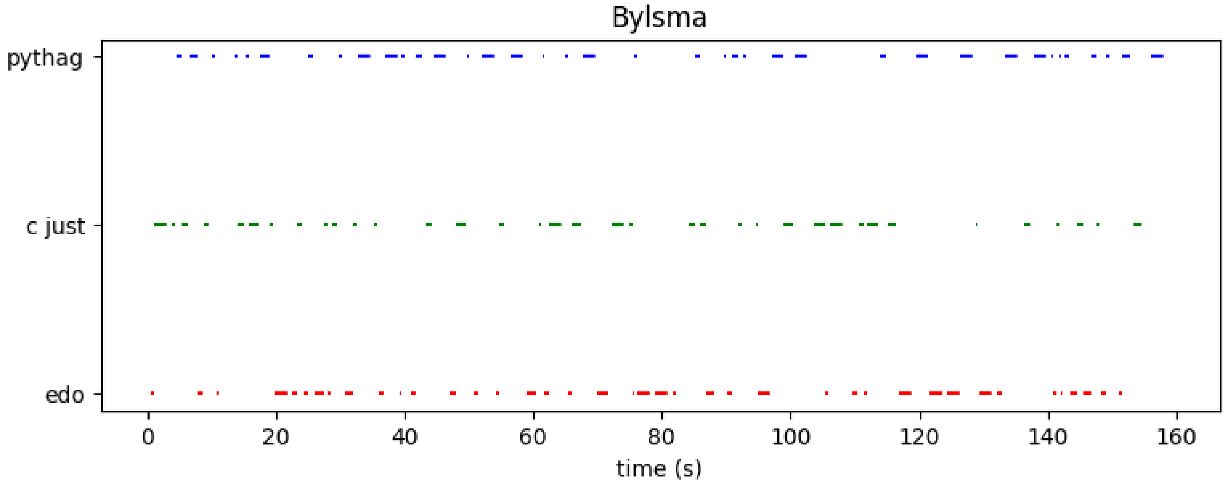


Figure A.1: Output of Byslma Full Suite 3 Prelude with only c major stencil

Table A.1: Byslma Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
edo	164	31.3
c just	153	29.2
pythag	207	39.5

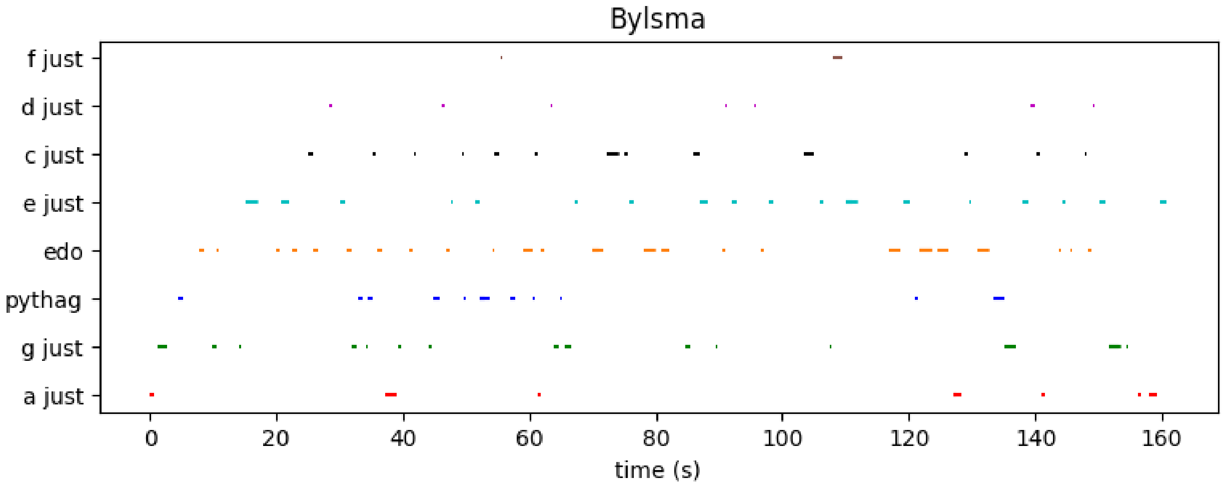


Figure A.2: Output of Bylsma Full Suite 3 Prelude with all just stencils

Table A.2: Bylsma Full Suite 3 Prelude with all just stencils

Tuning System	Total	Percentage
a just	24	7.16
g just	51	15.22
pythag	35	10.45
edo	88	26.27
e just	76	22.69
c just	41	12.24
d just	15	4.48

Continued on next page

Table A.2 – continued from previous page

Tuning System	Total	Percentage
f just	5	1.49

A.2 Colin Carr

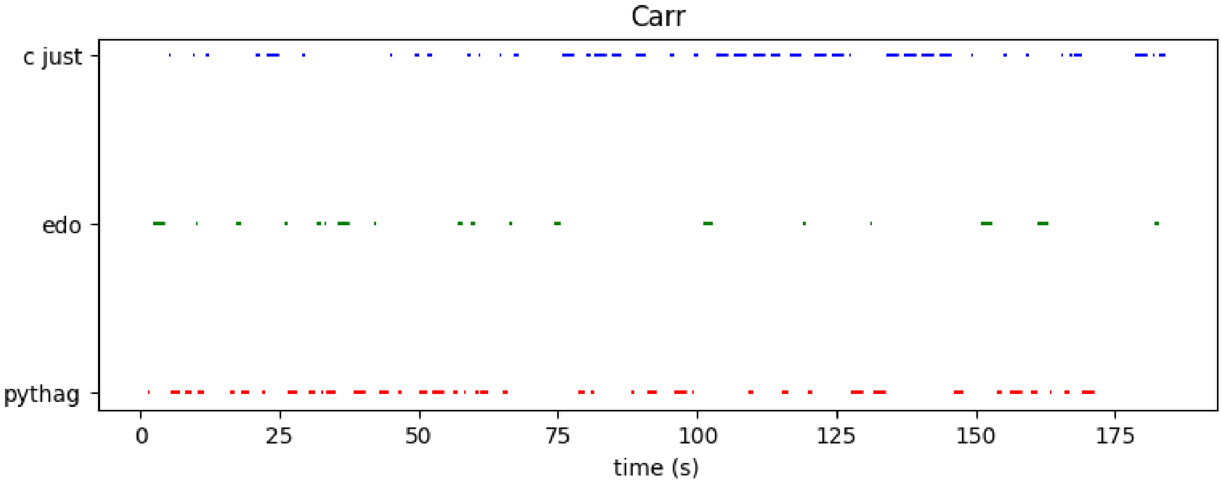


Figure A.3: Output of Carr Full Suite 3 Prelude with only c major stencil

Table A.3: Carr Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
pythag	177	45.5
edo	60	15.42
c just	152	39.07

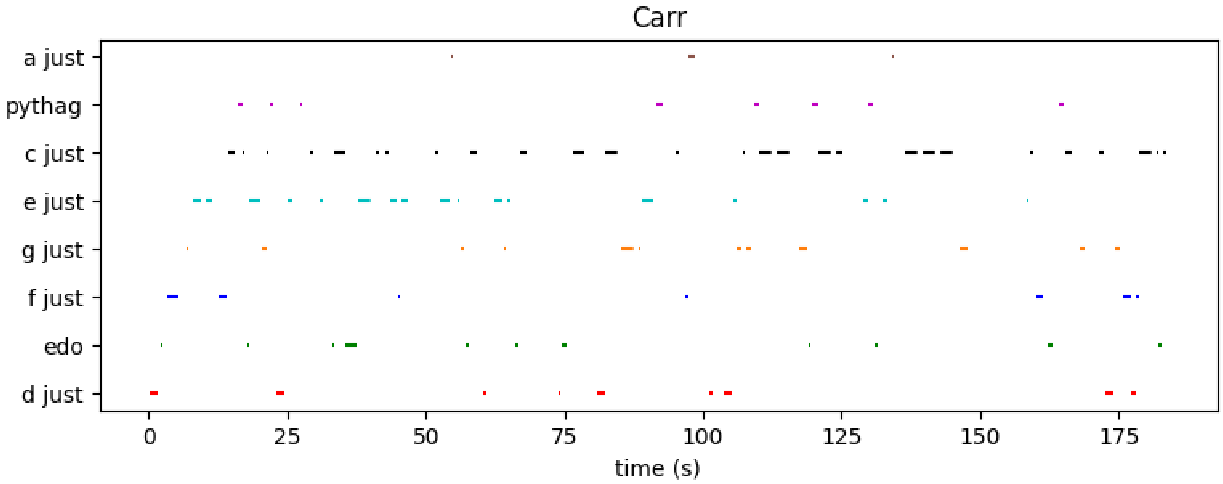


Figure A.4: Output of Carr Full Suite 3 Prelude with all just stencils

Table A.4: Carr Full Suite 3 Prelude with all just stencils

Tuning System	Total	Percentage
d just	34	11.11
edo	31	10.13
f just	19	6.21
g just	33	10.78
e just	68	22.22
c just	90	29.41
pythag	25	8.17
a just	6	1.96

A.3 Pablo Casals

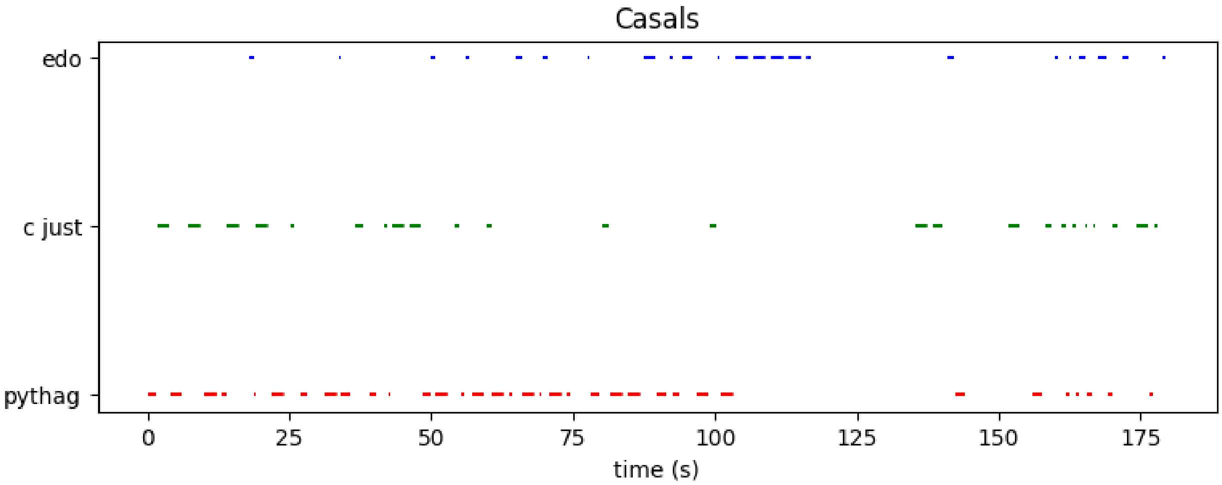


Figure A.5: Output of Casals Full Suite 3 Prelude with only c major stencil

Table A.5: Casals Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
pythag	155	51.5
c just	87	28.9
edo	59	19.6

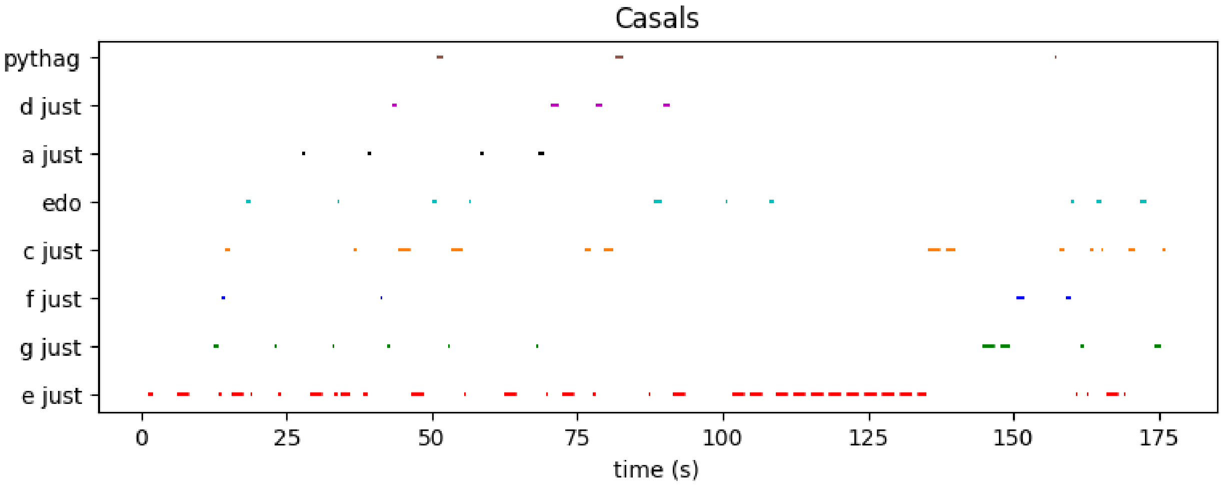


Figure A.6: Output of Casals Full Suite 3 Prelude with all just stencils

Table A.6: Casals Full Suite 3 Prelude with all just stencils

Tuning System	**Total**	**Percentage**
e just	102	42.86
g just	22	9.24
f just	10	4.2
c just	42	17.65
edo	27	11.34
a just	12	5.04
d just	10	4.2
pythag	13	5.46

A.4 Ethan Cobb

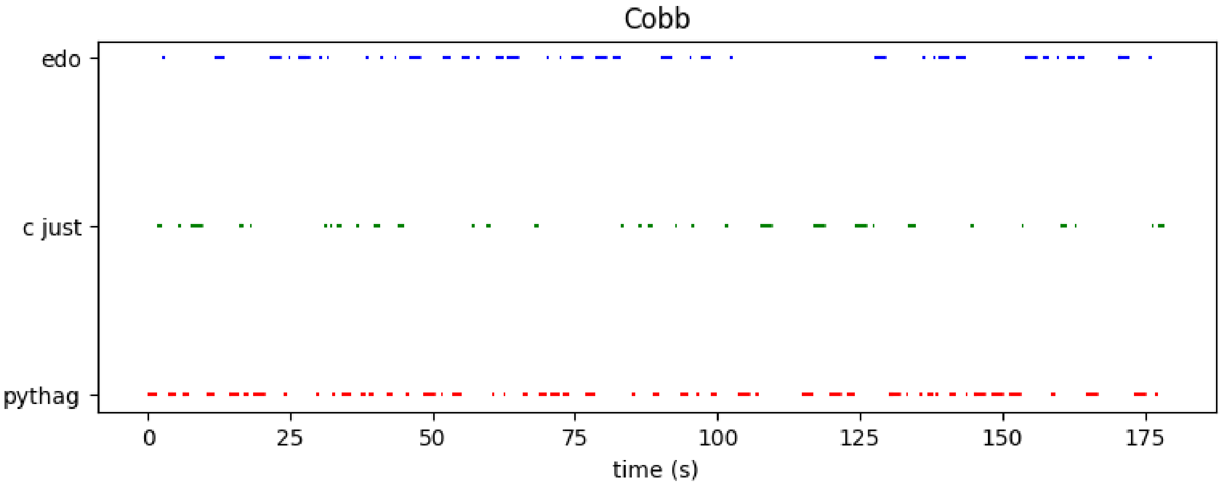


Figure A.7: Output of Cobb Full Suite 3 Prelude with only c major stencil

Table A.7: Cobb Full Suite 3 Prelude with only c major stencil

Tuning System	**Total**	**Percentage**
pythag	240	46.69
c just	100	19.46
edo	174	33.85

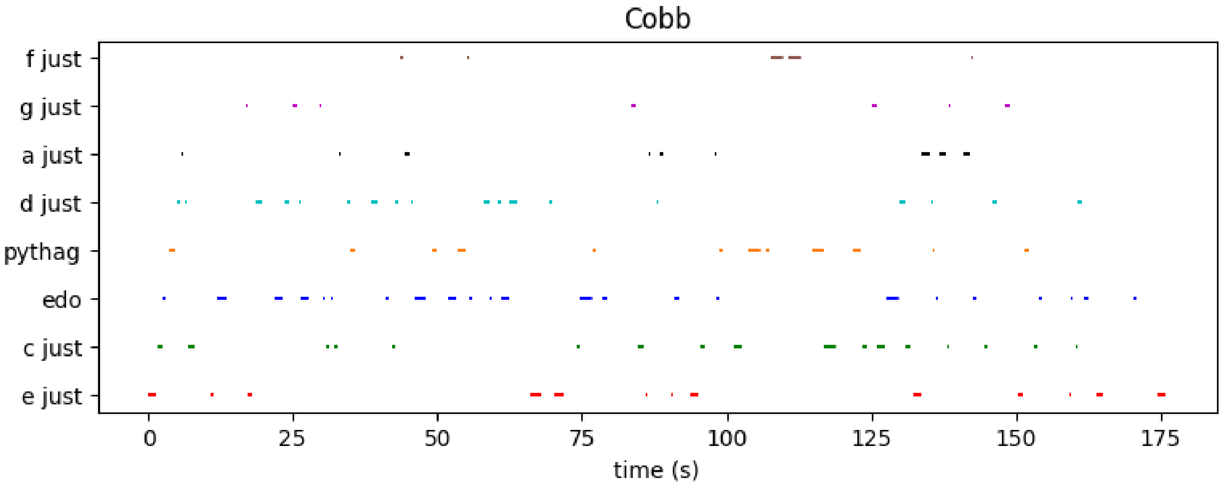


Figure A.8: Output of Cobb Full Suite 3 Prelude with all just stencils

Table A.8: Cobb Full Suite 3 Prelude with all just stencils

Tuning System	**Total**	**Percentage**
e just	51	14.78
c just	60	17.39
edo	78	22.61
pythag	35	10.14
d just	63	18.26
a just	27	7.83
g just	20	5.8
		Continued on next page

Table A.8 – continued from previous page

Tuning System	Total	Percentage
f just	11	3.19

A.5 Ralph Kirshbaum

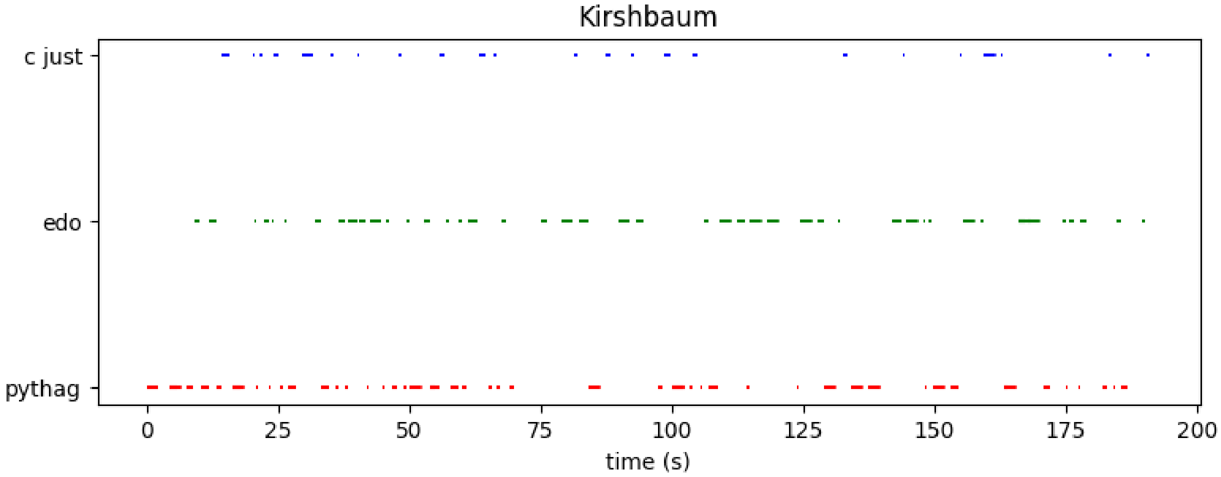


Figure A.9: Output of Kirshbaum Full Suite 3 Prelude with only c major stencil

Table A.9: Kirshbaum Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
pythag	212	46.39
edo	176	38.51
c just	69	15.1

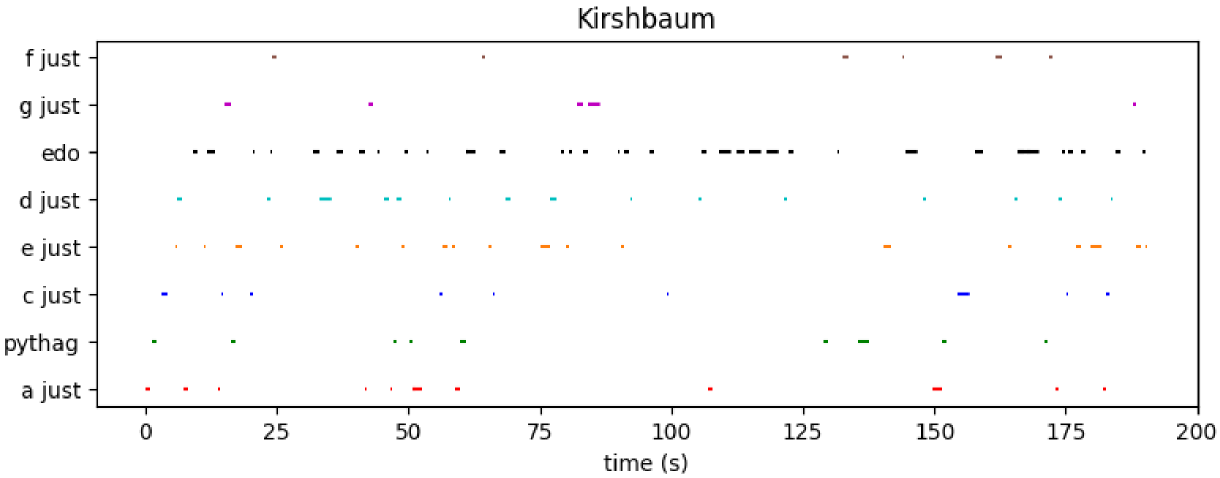


Figure A.10: Output of Kirshbaum Full Suite 3 Prelude with all just stencils

Table A.10: Kirshbaum Full Suite 3 Prelude with all just stencils

Tuning System	**Total**	**Percentage**
a just	35	10.42
pythag	30	8.93
c just	25	7.44
e just	46	13.69
d just	44	13.1
edo	128	38.1
g just	11	3.27
f just	17	5.06

A.6 Misha Maisky

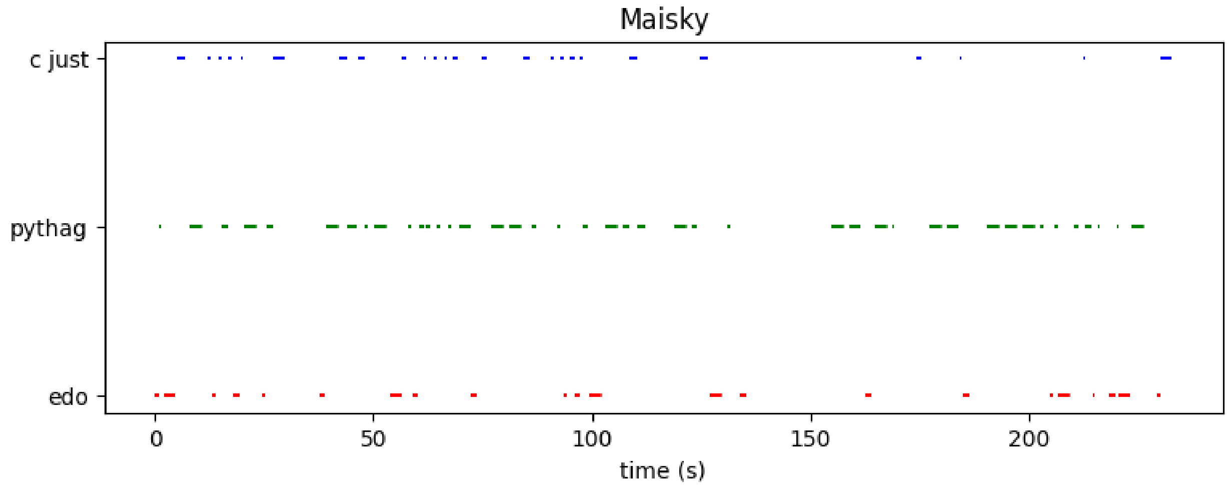


Figure A.11: Output of Maisky Full Suite 3 Prelude with only c major stencil

Table A.11: Maisky Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
edo	83	24.41
pythag	175	51.47
c just	82	24.12

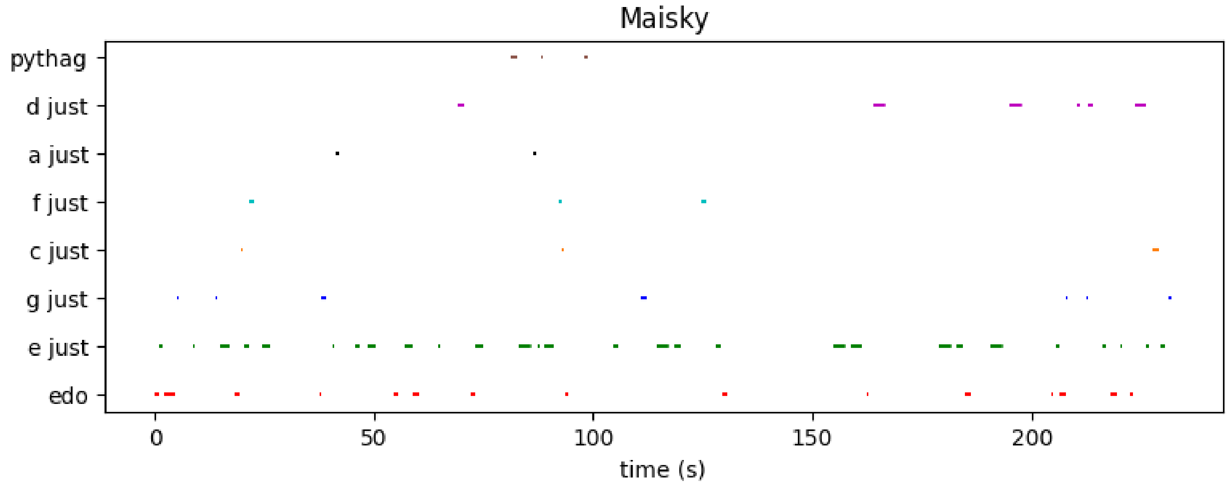


Figure A.12: Output of Maisky Full Suite 3 Prelude with all just stencils

Table A.12: Maisky Full Suite 3 Prelude with all just stencils

Tuning System	Total	Percentage
edo	42	20.39
e just	100	48.54
g just	14	6.8
c just	7	3.4
f just	9	4.37
a just	5	2.43
d just	22	10.68
pythag	7	3.4

A.7 Jean-Guihen Queyras

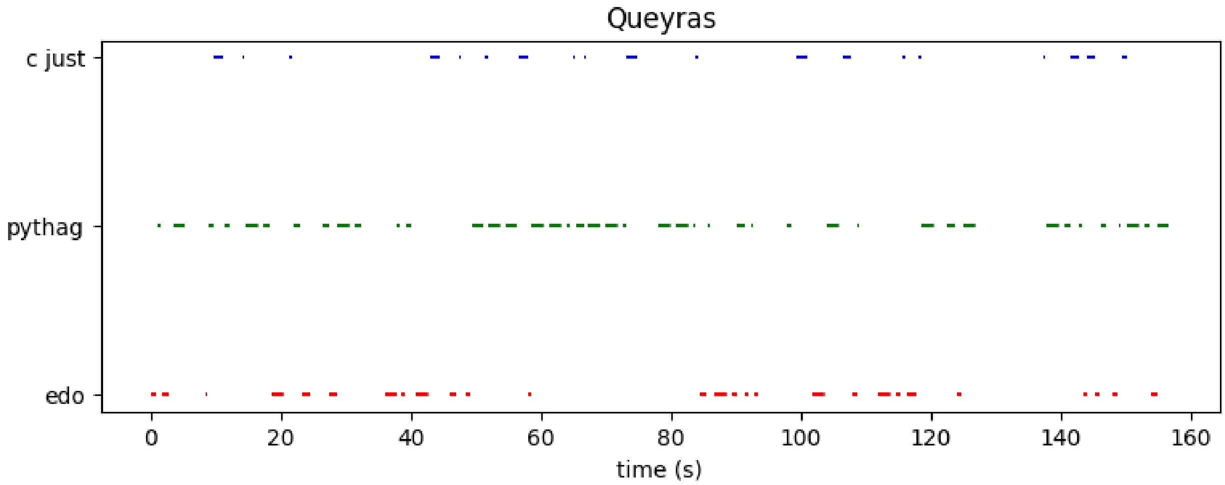


Figure A.13: Output of Queyras Full Suite 3 Prelude with only c major stencil

Table A.13: Queyras Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
edo	93	25.34
pythag	214	58.31
c just	60	16.35

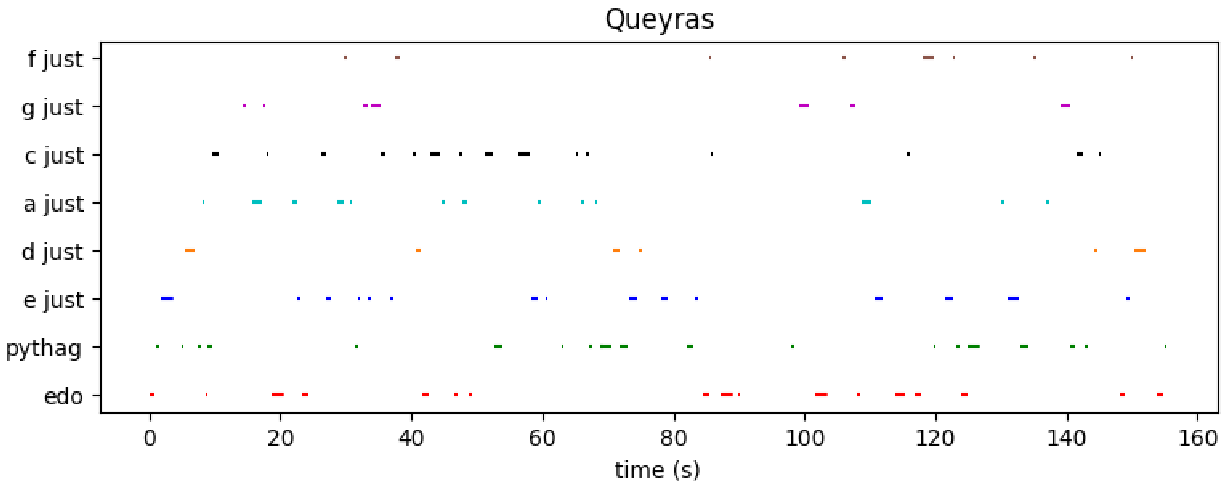


Figure A.14: Output of Queyras Full Suite 3 Prelude with all just stencils

Table A.14: Queyras Full Suite 3 Prelude with all just stencils

Tuning System	Total	Percentage
edo	51	17.47
pythag	59	20.21
e just	45	15.41
d just	17	5.82
a just	34	11.64
c just	44	15.07
g just	18	6.16

Continued on next page

Table A.14 – continued from previous page

Tuning System	Total	Percentage
f just	24	8.22

A.8 Mstislav Rostropovich

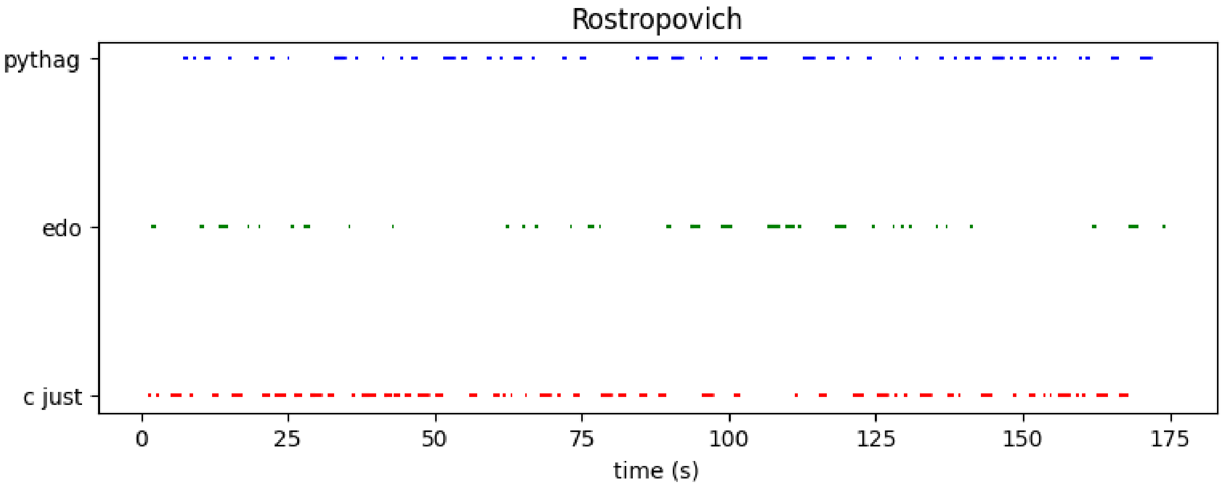


Figure A.15: Output of Rostropovich Full Suite 3 Prelude with only c major stencil

Table A.15: Rostropovich Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
c just	319	48.41
edo	113	17.15
pythag	227	34.45

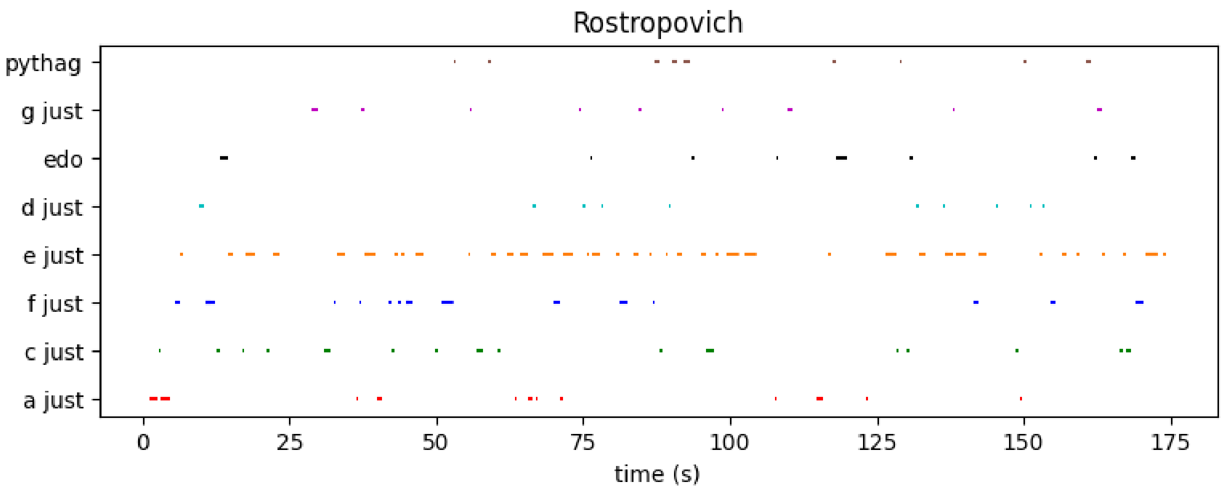


Figure A.16: Output of Rostropovich Full Suite 3 Prelude with all just stencils

Table A.16: Rostropovich Full Suite 3 Prelude with all just stencils

Tuning System	**Total**	**Percentage**
a just	29	6.9
c just	50	11.9
f just	58	13.81
e just	194	46.19
d just	24	5.71
edo	17	4.05
g just	24	5.71
pythag	24	5.71

A.9 Heinrich Schiff

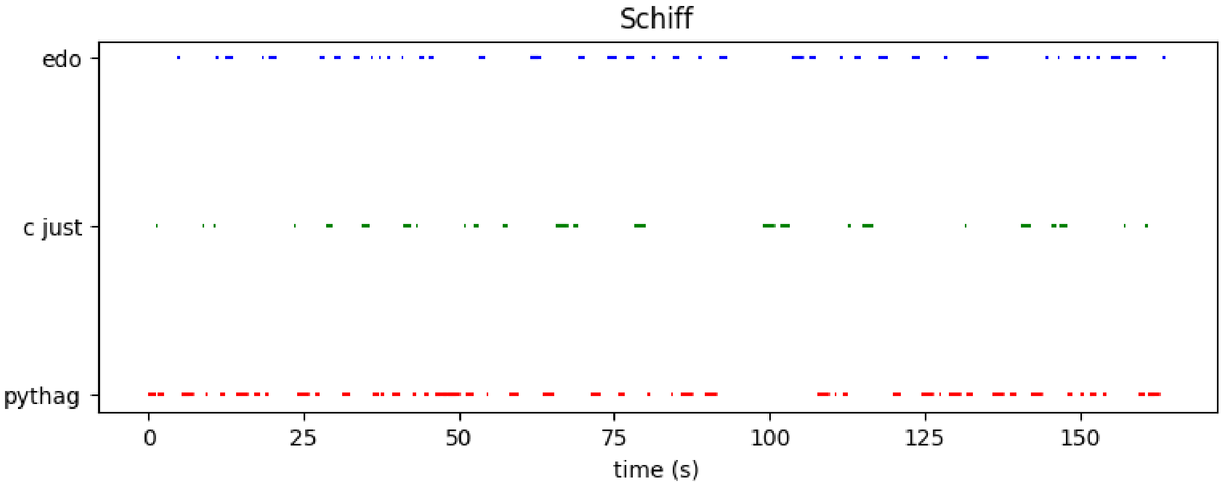


Figure A.17: Output of Schiff Full Suite 3 Prelude with only c major stencil

Table A.17: Schiff Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
pythag	203	47.32
c just	91	21.21
edo	135	31.47

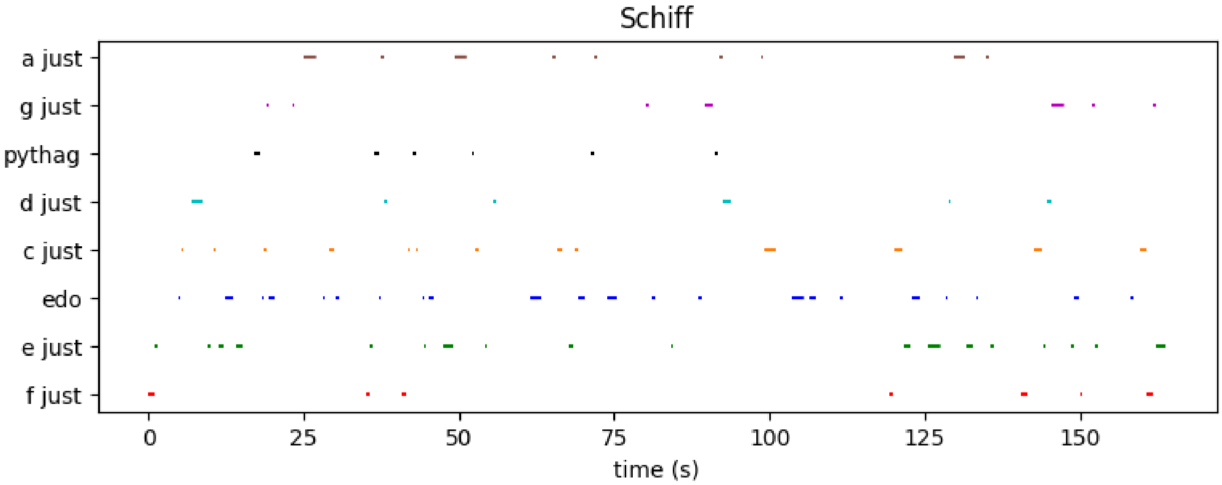


Figure A.18: Output of Schiff Full Suite 3 Prelude with all just stencils

Table A.18: Schiff Full Suite 3 Prelude with all just stencils

Tuning System	Total	Percentage
f just	24	8.66
e just	57	20.58
edo	68	24.55
c just	34	12.27
d just	18	6.5
pythag	16	5.78
g just	21	7.58
a just	39	14.08

A.10 Jan Vogler

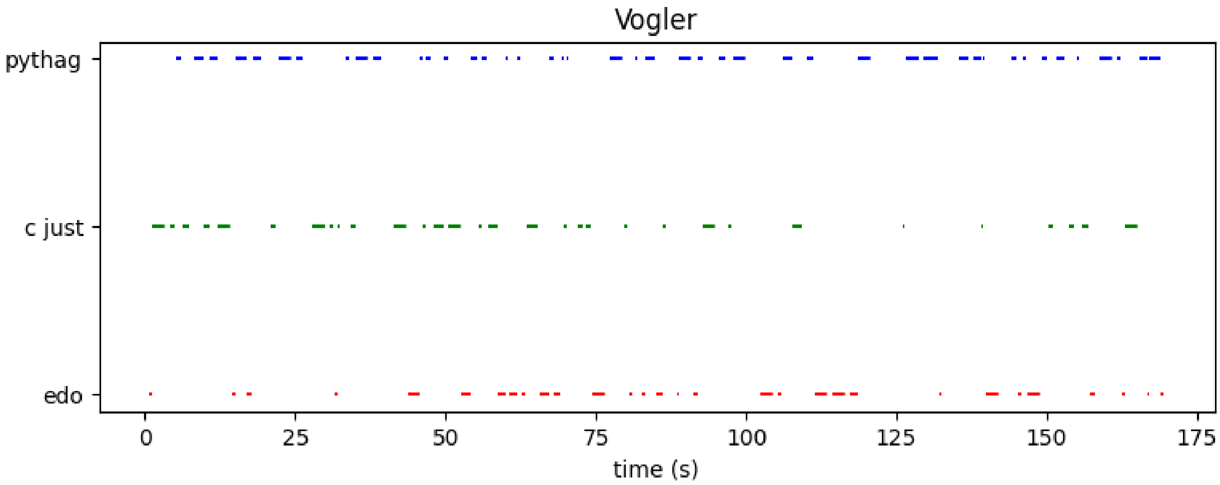


Figure A.19: Output of Vogler Full Suite 3 Prelude with only c major stencil

Table A.19: Vogler Full Suite 3 Prelude with only c major stencil

Tuning System	**Total**	**Percentage**
edo	104	24.07
c just	124	28.7
pythag	204	47.22

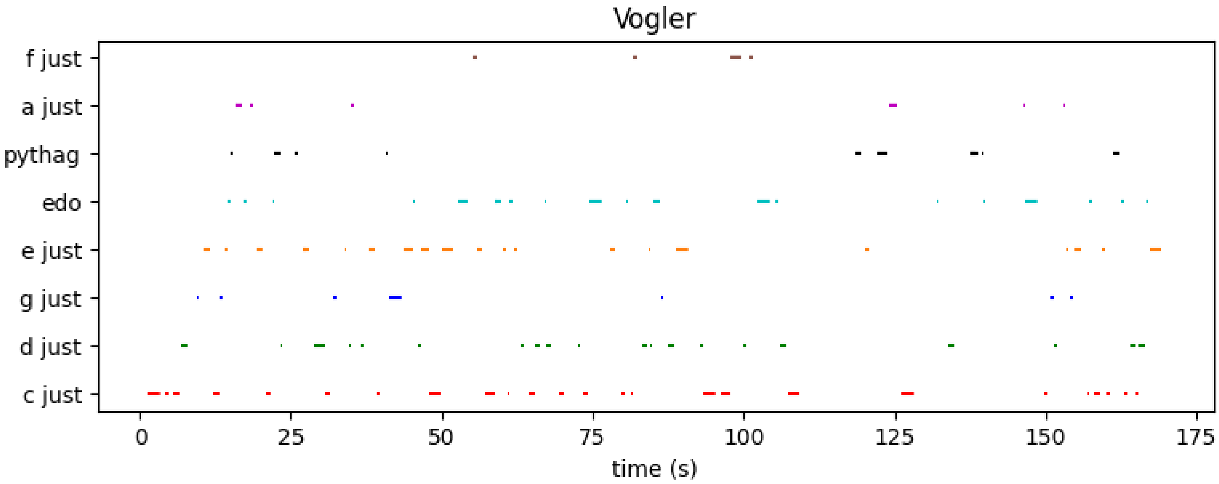


Figure A.20: Output of Vogler Full Suite 3 Prelude with all just stencils

Table A.20: Vogler Full Suite 3 Prelude with all just stencils

Tuning System	**Total**	**Percentage**
c just	79	23.58
d just	51	15.22
g just	19	5.67
e just	71	21.19
edo	55	16.42
pythag	29	8.66
a just	19	5.67

Continued on next page

Table A.20 – continued from previous page

Tuning System	Total	Percentage
f just	12	3.58

A.11 Pieter Wispelwey

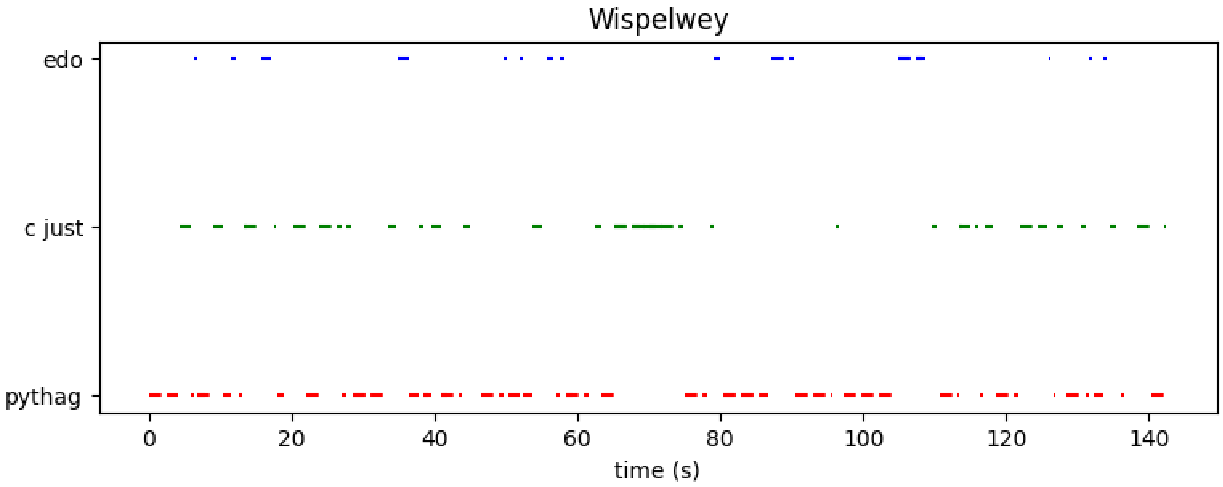


Figure A.21: Output of Wispelwey Full Suite 3 Prelude with only c major stencil

Table A.21: Wispelwey Full Suite 3 Prelude with only c major stencil

Tuning System	Total	Percentage
pythag	243	54.73
c just	131	29.5
edo	70	15.77

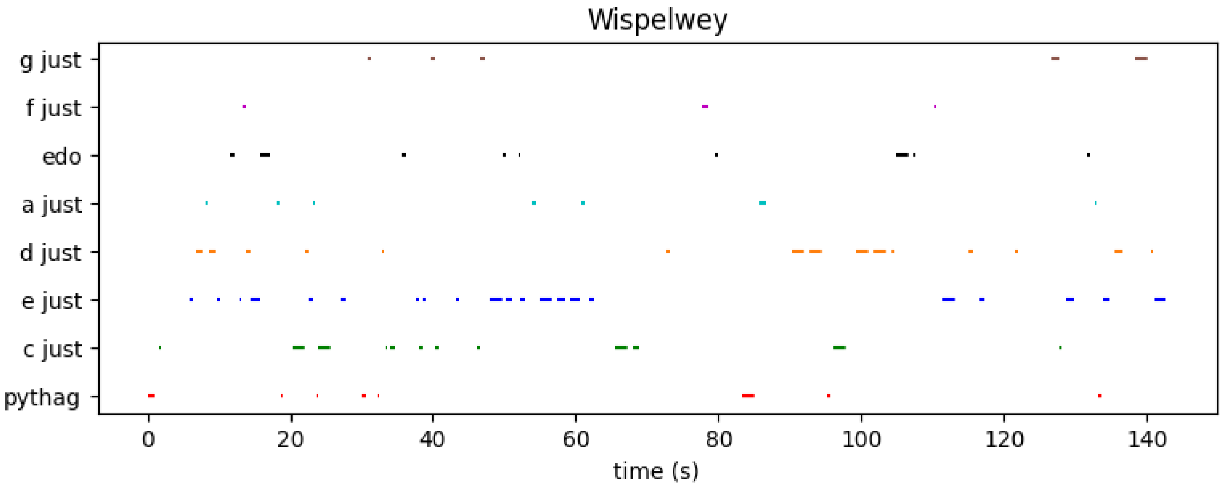


Figure A.22: Output of Wispelwey Full Suite 3 Prelude with all just stencils

Table A.22: Wispelwey Full Suite 3 Prelude with all just stencils

Tuning System	Total	Percentage
pythag	23	9.2
c just	32	12.8
e just	80	32.0
d just	44	17.6
a just	22	8.8
edo	26	10.4
f just	9	3.6
g just	14	5.6

Bibliography

[1] H. J. Jensen and M. R. Chung, *CelloMind: intonation and technique.* Ovation Press, Ltd., 2017.

[2] N. Cazden, "Pythagoras and aristoxenos reconciled," *Journal of the American Musicological Society*, vol. 11, no. 2/3, pp. 97–105, 1958.

[3] Aristoxenus and H. S. Macran, *Aristoxenu Harmonika stoicheia = the harmonics of aristoxenus.* Olms, 1990.

[4] R. P. Winnington-Ingram, "Aristoxenus and the intervals of greek music," *The Classical Quarterly*, vol. 26, no. 3-4, pp. 195–208, 1932.

[5] A. Daum, "The establishment of equal temperament," 2011.

[6] S. Dixon, M. Mauch, and D. Tidhar, "Estimation of harpsichord inharmonicity and temperament from musical recordings," *The Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 878–887, 2012.

[7] H. Fletcher, "Normal vibration frequencies of a stiff piano string," *The Journal of the Acoustical Society of America*, vol. 36, no. 1, pp. 203–209, 1964.

[8] D. Tidhar, S. Dixon, E. Benetos, and T. Weyde, "The temperament police," *Early Music*, vol. 42, no. 4, pp. 579–590, 2014.

[9] J. Devaney and D. P. Ellis, "An empirical approach to studying intonation tendencies in polyphonic vocal performances," 2008.

[10] S. McAdams, "Musical forces and melodic expectations: Comparing computer models and experimental results," *Music Perception*, vol. 21, no. 4, pp. 457–498, 2004.

[11] M. Mauch, K. Frieler, and S. Dixon, "Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory," *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. 401–411, 2014.

[12] D. P. Ellis, "Extracting information from music audio," *Communications of the ACM*, vol. 49, no. 8, pp. 32–37, 2006.

[13] A. C. Gedik and B. Bozkurt, "Pitch-frequency histogram-based music information retrieval for turkish music," *Signal Processing*, vol. 90, no. 4, pp. 1049–1063, 2010.

[14] D. Harasim, F. C. Moss, M. Ramirez, and M. Rohrmeier, "Exploring the foundations of tonality: statistical cognitive modeling of modes in the history of western classical music," *Humanities and Social Sciences Communications*, vol. 8, no. 1, pp. 1–11, 2021.

[15] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, "Crepe: A convolutional representation for pitch estimation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 161–165, IEEE, 2018.

[16] C. Denton, "The history of musical temperament and pitch before 1750," *History*, 1996.

[17] D. J. Levitin, "Absolute memory for musical pitch: Evidence from the production of learned melodies," *Perception & Psychophysics*, vol. 56, no. 4, pp. 414–423, 1994.

[18] P. R. Farnsworth, "Sacred cows in the psychology of music," *The Journal of Aesthetics and Art Criticism*, vol. 7, no. 1, pp. 48–51, 1948.