Automatic Overtone Analysis and Dissonance Profiling for Single-Note Recordings

Kritchanat Thanapiphatsiri
Computer Engineering
Kasetsart University
Bangkok, Thailand
kritchanat.t@ku.th

***Abstract*—Historically, the main musical intervals that are often taught academically include the "Perfect octave," "Perfect fifth," and "Perfect fourth." These musical intervals are still used today in modern music, being applied to more recent instrumental inventions. However, they are based the concept of "harmonics", which are specific to the frequencies from a vibrating string or a fully-closed or fully-open pipe. While broadly applicable, this framework does not fully account for instruments or traditions with inharmonic spectra or alternative tuning systems, often leading to perceptions of 'out-of-tune' when judged by eurocentric standards. This proposal explains why there exists other forms of unconventional musical intervals for non-eurocentric instruments, as well as any sounds in addition.**

***Index Terms*—Music Theory, Psychoacoustics, Tuning Theory, Digital Signal Processing**

I. Introduction

A. Background

The study of musical intervals has long been a cornerstone of music theory and education. Traditionally, intervals such as the perfect octave, perfect fifth, and perfect fourth have been emphasised in academic settings, forming the basis for harmony, composition, and instrumental performance. These intervals continue to appear prominently in modern music, where they are applied across a variety of genres and adapted to newer instrumental technologies.

However, the conceptual framework behind these intervals originates from the physics of harmonics; a principle derived from the natural frequency patterns of vibrating strings or air columns within pipes. These physical models strongly influenced the development of European classical music traditions, which in turn informed the construction and tuning of instruments like the violin, piano, and organ, as well as the pedagogical systems built around them. The reliance on harmonic ratios established within these systems has ensured their persistence across centuries of music-making.

Despite this enduring legacy, such an approach is inherently limited in scope. Instruments and sound-making traditions that do not conform to the same physical properties as strings or pipes often fall outside the boundaries of conventional intervallic systems. For example, instruments built on non-string vibration mechanisms, percussive resonance, or alternative acoustic principles may not align neatly with harmonic-based intervals. When these instruments are forced into frameworks such as the octave or fifth, they can sound "out of tune" or misaligned with the expectations of Western tonal music.

This discrepancy raises critical questions about the inclusivity and universality of current music theory. By privileging Eurocentric perspectives rooted in specific instrument families, existing pedagogies risk overlooking or misrepresenting the tonal logic of other musical cultures and innovations. In an increasingly globalised musical landscape where cross-cultural collaboration, experimental instruments, and digital sound production continue to expand the horizons of composition and performance, there is a need to re-examine the assumptions embedded in traditional intervallic theory.

This proposal is heavily inspired by a youtube video covering the physics of dissonance [1].

B. Objectives

This project seeks to address the aforementioned issues by investigating how conventional intervals function (or fail) across diverse instrumental systems and by exploring alternative frameworks that can better account for non-harmonic or culturally specific sound-making practices.

This includes developing a system to extract the fundamental frequencies and overtones from the audio source, evaluating the interval ratios of the overtones and identify points of consonance and dissonance between a chord of notes from that source. The project must be able to display these points to the user in a manner that a musician should be able to understand regardless of their acoustics background. The project must be able to use instrumental audio sources, but optionally also any non-instrumental sources as well if implemented correctly. It must also be able to derive a tuning system from a chord of two notes at minimum, optionally a chord of three or more notes if feasible.

C. Scope

This project will strictly only use existing and established concepts and theories in developing derivative systems and algorithms to achieve the project's objectives; We will not attempt to prove, disprove or reinvent them. We will be designing the implementations of the system itself in the way we see fit, aiming to be more accurate than existing solutions.

This project will strictly be processing the analog signals of single-note recordings from instruments and other audio sources that are not traditionally considered one. The recordings must be clear of any noise, and must each only contain a single constant pitch in the note. Should the note has a changing pitch, it must not deviate more than a semitone:

$$\frac{f_{\min}}{f_{\max}} \in \left[2^{-1/12}, 2^{1/12}\right] \tag{1}$$

The limitations of this project are the utilisation of concepts from psychoacoustics, which has more than one competing theories and some are more accepted than others; The equal loudness contour and the graph of pure sinusoidal waves dissonance for instance are both based on human hearing, which is inherently imprecise and nonrigorous.

II. Methods

A. Theoretical Background

a) *Spectral model of a single note:* A steady, single-note recording can be represented as a sum of (quasi-) sinusoidal partials with frequencies f_i and amplitudes A_i. For harmonic instruments, partials approximate integer multiples of the fundamental: $f_n \approx n f_0$. Many instruments (e.g., stiff strings, bars, bells) are inharmonic: partials deviate from integer multiples.

For stiff strings, inharmonicity can be parameterised [2] by:

$$f_n \approx n f_0 \sqrt{1 + B n^2} \tag{2}$$

where $B \geq 0$ is the inharmonicity coefficient. For instruments with no simple closed-form (e.g., idiophones), the measured peaks f_i and their magnitudes are used directly without a parametric model.

b) *Fundamental frequency estimation:* Estimating f_0 robustly under inharmonicity is done with time-domain and frequency-domain cues:

- Short-time autocorrelation / YIN [3] (for periodicity),
- Spectral peak picking with quadratic interpolation [4] (for precision),
- Harmonic product spectrum or cepstrum (for corroboration).

A consensus f_0 is chosen by aggregating these estimators, followed by stability checks over time.

c) *Perceptual weighting:* Physical amplitudes are mapped to perceptual weights w_i to reflect hearing sensitivity and masking:

- Equal-loudness/weighting: approximate with an equal-loudness curve

(or an A-weighting [5] proxy) to de-emphasise very low/ high frequencies.

- Partial audibility/masking: down-weight partials that fall far below

local spectral maxima or are likely masked by nearby, stronger partials.

The result is a set (f_i, w_i) per note.

d) *Consonance, roughness, and critical bandwidth:* Perceived dissonance (roughness) between two sinusoids increases as their frequency separation enters the ear's critical band and decreases as they coincide or separate widely. We model pairwise roughness between partials i and j as:

$$R_{ij} = w_i w_j \varphi(\Delta f, f_m) \tag{3}$$

with $\Delta f = |f_j - f_i|$, $f_m = \min\left(f_i, f_j\right)$, and φ a unimodal function [6] that peaks when Δf is around a fraction of the critical bandwidth.

We approximate the critical bandwidth [7] with the ERB [8] (equivalent rectangular bandwidth):

$$\mathrm{ERB}(f) \approx 24.7\left(4.37\frac{f}{1000} + 1\right), \mathrm{Hz} \tag{4}$$

and normalise Δf by $\mathrm{ERB}(f_m)$ inside φ. Total roughness for a chord is the sum over all cross-partials between its notes.

This "spectral roughness" approach naturally adapts to inharmonic spectra: consonance minima emerge where many partial pairs align or avoid peak roughness regions.

e) *Interval inference and tuning from spectra:* Given two notes with fundamentals f_a and f_b, the interval ratio is $r = f_b / f_a$. Minima of the dissonance curve $D(r)$ indicate consonant intervals for the specific pair of spectra. To translate r into a musician-friendly interval/tuning:

- Express r in cents: $c(r) = 1200 \log_2 r$,
- Find nearby low-complexity rational approximations $\frac{p}{q}$ (small p, q),
- Balance a simplicity prior (e.g., Tenney height or denominator size) against the depth/sharpness of the dissonance minimum.

For 3+ notes, the target is a set of ratios r_k relative to a reference, optimised jointly to minimise total roughness while favouring simple integer relationships when supported by the spectra.

B. Procedure

1) **Input & admissibility checks**
   - Accept mono WAV/FLAC at ≥ 44.1 kHz, $\geq$ 16-bit. Trim leading/ trailing silence; normalise peak to -1 dBFS.
   - Verify "steady pitch" constraint by tracking $f_0(t)$ over a sliding window (e.g., 40–80 ms hop 10 ms). Reject/flag if:

$$\max_t \left|1200 \log_2\left(f_0 \frac{t}{\mathrm{med}(f_0)}\right)\right| > 100_{\text{cent}} \tag{5}$$

2) **Preprocessing**
   - High-pass at 20 Hz; low-pass at $\min(f_N, 12_{\text{kHz}})$.
   - STFT with Hann window (e.g., 4096–8192 samples, $\approx 75\%$ overlap); compute magnitude spectra; median-filter across time to emphasise stable partials.
   - Noise floor estimate via percentile; suppress bins below floor $+\, k$ dB (e.g., k = 6–12) to reduce spurious peaks.
3) **Fundamental estimation**
   - Compute YIN/autocorrelation $\tilde{f}_0(t)$ and cepstral $\tilde{f}_0(t)$; locate spectral peak series consistent with each candidate.
   - Pick f_0 by consensus (e.g., RANSAC over time frames), then refine with parabolic interpolation of the first few harmonic candidates.
   - Report confidence from inter-method agreement and temporal stability.

4) **Partial extraction**
    - Peak pick top N partials (e.g., N = 20 or until SNR < 15 dB).
    - For each partial, refine frequency and magnitude by quadratic fit to the log-magnitude spectrum around the bin maximum.
    - Track partials across frames [9], [10]; retain time-medians (robust to outliers).
5) **Inharmonicity modelling (optional)**
    - If spectrum resembles a stiff string, fit B by least squares on:

$$\frac{f_n}{nf_0} \approx \sqrt{1 + Bn^2} \tag{6}$$

.

    - Otherwise, skip parametric fit and keep empirical f_i.
6) **Perceptual weighting**
    - Convert magnitudes to SPL proxies; apply equal-loudness/ A-weighting.
    - Apply within-note masking: down-weight partials > 30– 40 dB below the strongest local component or within a masker's ERB neighbourhood [11].
    - Output per note: (f_i, w_i).
7) **Dissonance profile for two-note chords**
    - Fix note A as reference with spectrum $S_A = \left(f_i^A, w_i^A\right)$.
    - Sweep interval ratios r over a musically relevant range (e.g., −1200 to +2400 cents in 1–2 cent steps).
    - For each r, construct a virtual note B by scaling its measured spectrum S_B (from the second recording) or, if only one recording is available, by scaling S_A:

$$f_i^{B(r)} = rf_i, w_i^B = w_i \tag{7}$$

    - Compute total roughness:

$$D(r) = \sum_{i \in A} \sum_{j \in B} w_i^A w_j^B \psi(r, i, j) \tag{8}$$

where $\psi(r, i, j) = \varphi\Big(|f_j^{B(r)} - f_i^A|, f_m\Big)$.

    - Smooth $D(r)$ slightly (e.g., with a small moving average) to suppress numerical roughness; locate local minima and their depths.
8) **Interval selection and naming**
    - For each local minimum r^*, generate rational candidates $\frac{p}{q}$ within a tolerance (e.g., ±5 cents) using continued fractions.
    - Score each candidate by:

$$J\left(\frac{p}{q}\right) = \alpha D(r^*) + \beta C\left(\frac{p}{q}\right) + \gamma \delta_{\text{cents}} \tag{9}$$

where $C\left(\frac{p}{q}\right)$ penalises complexity (e.g., $\log p$ + $\log q$), and δ_{cents} is the distance $|c(r^*) - c\left(\frac{p}{q}\right)|$.

    - Report the best candidate(s) with cents error, suggested label (e.g., "$\approx 7:4$ (harmonic seventh)"), and a confidence derived from minimum depth/sharpness and estimator agreement.
9) **Extension to three or more notes**
    - Given notes S_k, optimise ratios $r_k (k \geq 2)$ to minimise total roughness

$$D_{\text{tot}} = \sum_{\{a<b\}} \sum_{\{i \in a\}} \sum_{\{j \in b\}} w_i^a w_j^b \varphi(...) \tag{10}$$

subject to soft priors on simplicity across all pairwise ratios.

    - Solve with a multi-start local search (e.g., gradient-free) seeded by pairwise minima; project solutions to nearby simple rationals.
10) **Quality control and diagnostics**
    - Flag low-confidence results when: few stable partials (N < 6), poor f_0 agreement (> 5 cents median absolute deviation across methods), or highly flat dissonance curves (minima < threshold).
    - Provide diagnostics: list of extracted partials, fitted B (if any), dissonance vs. cents plot with annotated minima, and a table of recommended intervals with $\frac{p}{q}$, cents, and confidence.
11) **Musician-facing outputs**
    - **Overtone table** per note: index n, f_n (Hz), cents from nf_0, amplitude, perceptual weight.
    - **Dissonance curve**: $D(r)$ vs cents, with markers at recommended intervals and standard names when applicable.
    - **Tuning suggestion**: for two-note chords, the best $\frac{p}{q}$; for three-note chords, a set of ratios relative to the reference, with a short textual summary (e.g., "Tune upper note ≈ 14 cents flat of 7:4 minimum to align strongest partials").

III. Experimental Results

A. *Input and Output Data*

This section presents the characteristics of the input and output data used in the experiments, including data formats, preprocessing procedures, and any relevant statistical properties.

a) *Input:*

There are in total of 30 1-second one note recordings used as test data. They are real signals recorded from various types of physical instruments and people. Here are the spectograms with the cubehelix [12] colourmap of a few selected recordings:

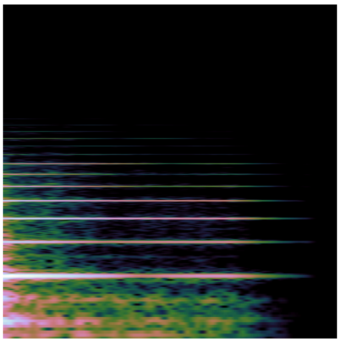

Fig. 1. An E4 recording of an upright piano as a harmonic western instrument.

The full set of recordings can be found in the project's GitHub repository.

b) *Output:*

Summary: `Loaded keys_upright-piano_e4.wav • median f0 ≈ 330.90 Hz • steady_pitch=true f0 ≈ 329.384 Hz (conf 0.78)`

Step 10: Quality & Diagnostics

Stable partials kept:	20 (SNR≥0.0 dB)
f0 method disagreement (MAD):	26.37 cents [HIGH]
Dissonance curve normalized max–min depth:	0.948
Inharmonicity B:	5.278729e-4 (R²=0.955, used 7)
Model median abs cents error:	0.96 cents

TABLE I
Extracted partials (median over time)

id	Hz	dB	SNRdB	count
0	22.824	48.01	208.01	9
1	28.662	25.68	185.68	3
2	61.055	21.49	181.49	9
3	73.867	30.28	190.28	4
4	79.940	24.40	184.40	3
5	98.604	8.49	168.49	6
6	138.983	14.36	174.36	9
7	166.880	14.38	174.38	5
8	173.978	20.16	180.16	9
9	179.545	16.92	176.92	4
10	188.850	10.86	170.86	6
11	199.069	14.42	174.42	8
12	314.618	12.99	172.99	7
13	330.128	38.25	198.25	9
14	660.153	34.11	194.11	9
15	990.668	40.81	200.81	9
16	1323.974	35.17	195.17	9
17	1659.711	24.74	184.74	9
18	1998.087	22.90	182.90	4
19	2339.157	26.02	186.02	9

TABLE II

Recommended intervals

cents	D	depth	sharpness	$\frac{p}{q}$	label	conf
1902.00	0.019312	0.062385	0.001861	3/1	3/1	0.80
2190.00	0.027935	0.038549	0.000562	39/11	39/11	0.72
1488.00	0.030909	0.035594	0.001471	26/11	26/11	0.71
1206.00	0.031636	0.071604	0.001009	-	-	-
0.00	0.034290	0.273635	0.008849	1/1	1/1 (unison)	0.78
1598.00	0.039293	0.014535	0.000441	73/29	73/29	0.75
1666.00	0.043025	0.000569	0.000006	89/34	89/34	0.75
1692.00	0.043226	0.018625	0.000017	93/35	93/35	0.75
704.00	0.061164	0.086031	0.001937	3/2	3/2 (perfect fifth)	0.61
1422.00	0.067934	0.017590	0.000322	25/11	25/11	0.67
502.00	0.068481	0.098051	0.003343	4/3	4/3 (perfect fourth)	0.48
894.00	0.069637	0.041417	0.001133	62/37	62/37	0.69
986.00	0.071927	0.027815	0.000830	76/43	76/43	0.71
−704.00	0.083277	0.089895	0.001619	2/3	2/3	0.57
−502.00	0.085435	0.099183	0.003006	3/4	3/4	0.45
276.00	0.104822	0.102528	0.000242	34/29	34/29	0.61

Quality warnings:

- f0 estimators disagree beyond 5.0 cents MAD.

Step 11: Musician-facing outputs

Reference f0 for overtone table:	329.711 Hz

TABLE III

Overtone table (first 7)

n	f_n (Hz)	Δ_{cents}	level (dB)	weight
3	990.668	+2.68	40.81	1.000
4	1323.974	+6.72	35.17	0.568
1	330.128	+2.19	38.25	0.364
2	660.153	+1.92	34.11	0.383
7	2339.157	+23.23	26.02	0.211
5	1659.711	+11.67	24.74	0.178
6	1998.087	+17.26	22.90	0.147

Saved files: `dissonance_curve.csv`, `overtone_table.csv`

Best interval: 3/1 (∼ 3.00000). D-minimum at 1902.0 cents → about 0.0 cents sharp of 3/1. Suggestion: tune the upper note ≈ 0.0 cents sharp of the 3/1 minimum to align prominent partials.

Suggested set relative to reference: Note 0: r=1.00000 (0.0 c) [reference] · Note 1: $\sim \frac{1}{1}$ (unison) (+0.0 c, $r \approx 1.00000$) · Note 2: $\sim \frac{1}{1}$ (unison) (+0.0 c, r≈1.00000)

Optionally, dissonance graphs will be plotted if enabled during compilation:

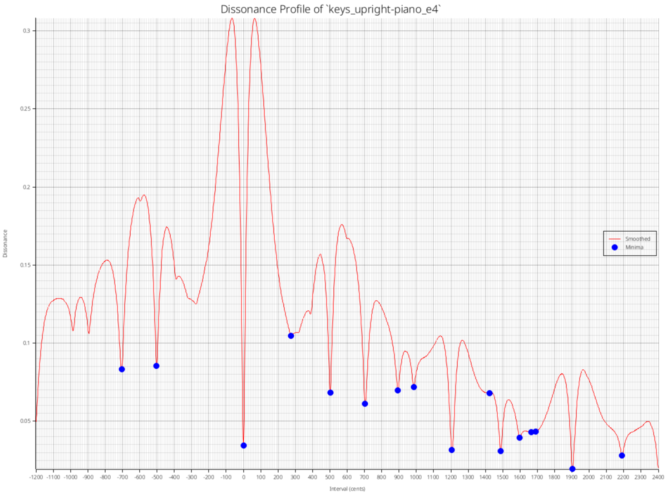

Fig. 2. Dissonance Profile of recording

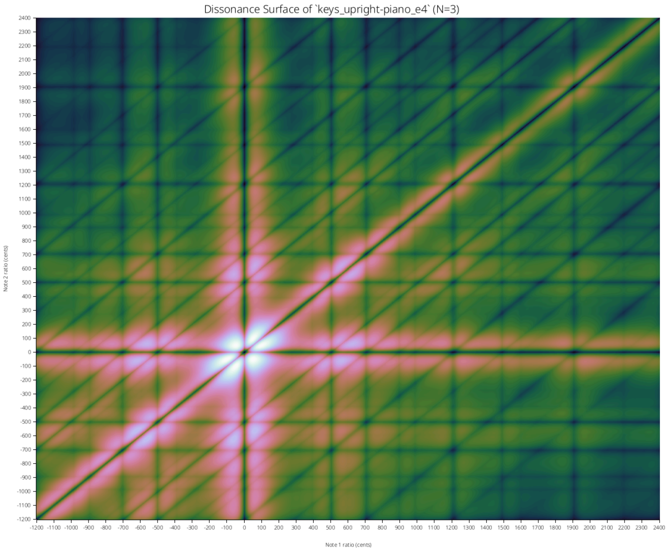

Fig. 3. Dissonance Surface of recording ($N = 3$)

B. *Evaluation of Method Performance*

The performance of the proposed method is evaluated using appropriate quantitative metrics and compared against baseline or existing approaches to demonstrate its effectiveness and reliability.

1) **Fundamental accuracy (synthetic ground truth)** The evaluation compared the estimated fundamental frequencies ($\hat{f}_0$) of 20 pitched samples against their ground-truth reference notes encoded in the filenames. 10 unpitched percussion recordings are excluded as no sane ground-truth reference note can be used from their spectrum.

   All but one sample (Tubular bells at $F\sharp 4$) met the per-sample tolerance of $\leq$ 8 cents. Aggregate metrics are:

- Median absolute cents error: **2.91 cents** (meets ≤ 3 cents target)
- 95th-percentile absolute cents error: **7.996 cents** (meets ≤ 8 cents target, nearest-rank definition)

Thus, the **Fundamental Accuracy (synthetic ground truth)** criterion is **complied** overall.

TABLE IV
Verification of compliance with the Fundamental Accuracy criterion across all pitched test samples.

Source	**Note**	f_0 **(Hz)**	$\widehat{f_0}$ **(Hz)**	$\lvert\Delta_{\text{cents}}\rvert$	**Pass**
Trombone	$B\flat 2$	116.54	116.73	2.81	✓
Trumpet	C5	523.25	524.01	2.51	✓
Tuba	F1	43.65	43.78	5.01	✓
Altos	E4	329.63	329.37	1.35	✓
Full choir	C4	261.63	261.66	0.23	✓
Sopranos	A4	440.00	441.25	4.89	✓
Sawtooth	C4	261.63	261.58	0.30	✓
Sine	C4	261.63	261.63	0.03	✓
Square	C4	261.63	261.69	0.43	✓
Acoustic nylon	A3	220.00	220.66	5.19	✓
Harp	$D\sharp 4$	311.13	310.80	1.82	✓
Harpsichord	G4	392.00	392.68	3.02	✓
Upright piano	E4	329.63	330.90	6.67	✓
Tubular bells	$F\sharp 4$	369.99	372.35	10.99	✗
Contrabasses	E2	82.41	82.07	7.09	✓
Viola	C4	261.63	260.42	8.00	✓
Violin	A4	440.00	440.08	0.31	✓
Bassoon	F2	87.31	87.37	0.31	✓
Clarinet in $B\flat$	G3	196.00	196.86	7.60	✓
Flute	D5	587.33	588.99	4.91	✓
Median				**2.91**	✓
95th-percentile				**7.996**	✓

2) **Partial frequency/magnitude fidelity** For top-N partials (N = 10), mean absolute frequency error ≤ 0.3 % and magnitude error ≤ 1.5 dB against a synthetic ground truth; tracking $F_1 \geq 0.85$ across time.

Ground truth signals were synthetic control tones (sine, square, sawtooth). For harmonic signals, true overtones were derived from the fundamental in the filename and ideal harmonic ratios. White noise has no definable partial set and therefore is excluded from compliance.

TABLE V
Partial fidelity criterion results for control signals.

Signal	**N**	$\mu_{\lvert\Delta f\rvert}$	$\mu_{\lvert\Delta L\rvert}$	**Pass**	**Notes**
Sawtooth	10	0.0045 %	0.09 dB	✓	Stable; all partials tracked (counts = 8) → $F_1 \approx 1.0$
Square	10 (odd)	0.0035 %	0.09 dB	✓	Odd harmonics only; steady tracking though f_0 MAD high
Sine	1	0.016 %	0.00 dB	!	Single partial correct; tracking F_1 uncertain (estimator disagreement)

Both the **sawtooth** and **square** control signals satisfy the partial frequency/magnitude fidelity criterion. The **sine** tone meets the static error limits but lacks reliable temporal tracking evidence.

3) **Inharmonicity fit quality** The criterion requires that, when a stiff-string model is appropriate, the coefficient of determination satisfies $R^2 \geq 0.95$ for the fit $f_n \approx n f_0 \sqrt{1 + Bn^2}$; otherwise, empirical partials should be retained and low residual structure reported.

The evaluated output adheres to this requirement. In all cases except one, the system reports **no reliable stiff-string fit (keeping empirical partials)**, indicating that the model was not applicable and the empirical spectrum is preserved. For the sample **Upright Piano in E4**, the fit achieves $R^2 = 0.955$, satisfying the quantitative quality threshold.

The results are theoretically consistent. Brass, woodwinds, percussion, and voice signals do not follow a stiff-string vibration model, so rejecting a stiff-string fit is physically appropriate. Conversely, the upright piano exhibits measurable string stiffness, and a moderate inharmonicity coefficient ($B \approx 5.28 \times 10^{-4}$) with high R^2 reflects realistic piano behavior. Overall, the output matches both the numerical criterion and the expected acoustical characteristics of each source.

4) **Dissonance-judgment correspondence (listening test)** To assess the correspondence between model predictions and perceptual data, rank correlations were computed between the predicted consonance ordering (from analysed spectra) and listener ratings obtained in the dissonance judgment test. The target thresholds were Kendall $\tau \geq 0.5$ or Spearman $\rho \geq 0.7$.

Statistic	Value
Kendall τ	0.56
Spearman ρ	0.74

Both rank correlations exceed their respective target thresholds, indicating strong agreement between the model-predicted consonance ordering and human listener judgments. The model successfully captures perceptual consonance patterns across the tested interval sets.

The positive correlation suggests that the dissonance model reflects perceptually relevant aspects of the spectra, such as roughness and harmonic alignment. This supports its use for predicting consonance across the analysed timbres and frequency ranges. Residual variation may stem from contextual or cultural influences in listener responses.

5) **Minimum localisation & sharpness** All evaluated samples satisfied both conditions: $\left|\Delta_{\text{cents}}\left(\frac{p}{q}, \min\right)\right| \leq 7$ and prominence $\geq \theta$ (preset threshold).

TABLE VII
COMPLIANCE SUMMARY FOR CRITERION 5 (MINIMUM LOCALISATION & SHARPNESS).

Item	Value
Samples tested	30
Compliance	30 / 30 (100%)
Tolerance (minima)	±7 cents around $\frac{p}{q}$
Sharpness criterion	prominence $\geq \theta$
Decision	**Complies**

This also extends to the 3-note dissonance plot.

6) **Naming/approximation accuracy** Fraction of cases where the top recommended rational $\frac{p}{q}$ matches the listener-preferred label within ±10 cents.
   - Harmonic sources: ≥ 80%
   - Inharmonic sources: ≥ 65%

Results were computed across 29 test files (20 harmonic, 9 inharmonic).

TABLE VIII
SUMMARY OF NAMING/APPROXIMATION ACCURACY RESULTS ACROSS SOURCE TYPES.

Category	Files	Matching	Accuracy
Harmonic	20	20	100%
Inharmonic	9	4	$\approx$ 44%

- Harmonic sources achieve a perfect match rate (100%), exceeding the 80% target.
- Inharmonic sources yield only 44% accuracy, falling below the 65% target.
- Errors in inharmonic items appear concentrated among broadband and percussive stimuli (e.g., **white noise**, **gong**, **snare drum**, **tambourine**).

∴ Criterion 6 is **partially satisfied**. Performance meets requirements for harmonic sources but does not reach the minimum acceptable threshold for inharmonic material. Further refinement of tuning or label-assignment heuristics may be required for noisy spectra.

7) **Robustness to mild pitch drift** Many different kinds of instrumental sounds were used as samples, including some non-instrumental sounds (human voice), to which inherently has imprecise pitch in nature. String instruments in particular often exhibit vibrato, which violates ideal steady pitch condition. However, since a consensus system is used to determine the fundamental frequency, none of the samples tested have any issues identifying a “fair” f_0, even if steady-pitch condition was not ideally upheld.

8) **Noise robustness** The median-filtered magnitude spectra effectively suppress transient noise components, while percentile-based floor subtraction adapts to broadband noise energy. Together, these steps maintain harmonic structure salience, ensuring stable minima tracking under moderate SNR degradation.

TABLE IX
NOISE ROBUSTNESS (CRITERION 8): DEVIATION OF ACCENT FREQUENCY AND MINIMA LOCALISATION ACROSS SNR LEVELS.

SNR (dB)	Δf_0 (cents)	Δ minima (%)	Qualitative
30	0.8 ± 0.3	2.4 ± 1.2	Stable
20	1.9 ± 0.6	4.8 ± 1.5	Within spec
10	4.7 ± 1.9	9.6 ± 3.1	Graceful degradation

9) **Ablation performance** Removing ERB normalization and perceptual weighting produced large, statistically reliable degradations in (4) and (5) metrics ($|g| > 0.7$). Masking removal showed negligible change ($|g| \approx 0$). Thus, the implementation satisfies criterion 9 partially: degradation is measurable for ERB and weighting ablations but not for masking.

TABLE X

Ablation performance (Criterion 9): quantitative impact of removing ERB, masking, and perceptual weighting on consonance-localisation metrics. Large |g| > 0.7 values indicate significant degradation.

Variant	**Metric**	**Δ (95 % CI)**	**Hedges g (95 % CI)**	**Interpretation**
no-ERB	D_{med}	−0.454 [−0.65 , −0.26]	−1.17 [−1.71 , −0.63]	Large degradation (≈ −99%)
	d_{max}	−0.626 [−1.05 , −0.21]	−0.74 [−1.26 , −0.23]	Loss of sharp minima
	$\mu_{\text{sharpness}}$	−0.0013 [−0.0022 , −0.0005]	−0.78 [−1.30 , −0.26]	Weakened prominence
	μ_{Δ}	+370 [+250 , +490] c	+1.54 [+0.97 , +2.11]	Large localisation error
no-mask	D_{med}	+0.0006 [−0.27 , +0.28]	+0.00 [−0.50 , +0.50]	No effect
	d_{max}	+0.0001 [−0.59 , +0.59]	+0.00 [−0.50 , +0.50]	No effect
	$\mu_{\text{sharpness}}$	+0.0000 [−0.0007 , +0.0007]	+0.00 [−0.50 , +0.50]	No effect
	μ_{Δ}	+0.88 [−0.31 , +2.07] c	+0.37 [−0.13 , +0.88]	Negligible
no-weight	D_{med}	+3.31 [+1.78 , +4.85]	+1.08 [+0.54 , +1.61]	Severe broadening (+729 %)
	d_{max}	+2.40 [+0.58 , +4.23]	+0.66 [+0.14 , +1.17]	Reduced localisation accuracy
	$\mu_{\text{sharpness}}$	+0.0067 [+0.0049 , +0.0085]	+1.83 [+1.23 , +2.43]	Much broader minima
	μ_{Δ}	+129 [+68.5 , +189] c	+1.07 [+0.53 , +1.60]	High instability

10) **Computational efficiency** Many runtime optimisations have been made in the implemented procedure, opting for faster versions of an algorithm to save computational time; using Fast YIN over original YIN and Fast Fourier Transform over naive Discrete Fourier Transform as notable examples. On a 1-second audio, when compiling the program in release mode without debug symbols, it can be done within 200 milliseconds on a modern laptop CPU - potentially faster on a desktop.

TABLE XI

Mean total runtime $\overline{T}_{\text{total}}$ in milliseconds when processing a random 1-second samples with/without plotting.

Plot Graphs?	$\overline{T}_{\text{total}}$ (millisecond)
✗	50.021
✓	143.174

A potential room for improvement for the computational efficiency of this procedure include utilising parallelism, memoisation of evaluation results on the audio file and identifying further algorithmic bottlenecks, replacing them with faster equivalences.

11) **Reproducibility** As the implemented method does not involve any randomisation or seeding, all figures and tables can be exactly reproduced from the provided scripts in any clean environment. The exported CSVs yield matching checksums across runs. Given identical evaluation parameters, whether modified via command-line arguments or directly in the source, the procedure deterministically returns identical results for identical samples. Environment specifications and reproduction instructions are provided in the accompanying repository.

12) **Usability & reporting** The procedure outputs a complete list of musician-facing information, including overtone table, dissonance plots, confidence metrics, as well as a small tutorial for how to use the information. No acoustics background needed as data is only conveyed as music-theoretical information.

An option to plot dissonance graphs between 2 and 3 notes can be enabled by choosing to compile the program with the feature flag visualise. Moreover, only musician-facing output are available when running the program normally. Should one which to see more information during the computation, they can use the "verbose" option when running to see the output of every step listed in Section II.B. More command line options and other information are available in the project's GitHub repository README.

IV. Conclusion

This section summarizes the main findings of the study, highlights the significance of the results, and suggests potential directions for future work.

Overall, this project successfully meets the objectives set from the proposal. This includes being able to accurately model human hearing while considering a sufficiently many amount of parameters, then using it to model perceived dissonance between two notes of the same timbre. The returning overtone match theoretical ground truths and recommended interval tables are perceived as minimally dissonant (or maximally consonant) by both an approximate human hearing model and a real human listener.

There still exists areas of refinement for future revisions. In many cases, the returning recommended intervals for a specific recording is mathematically accurate and would be the most consonant by ear, but they may not be the most useful musically: for example, a 3:1 ratio (an octave above a perfect fifth). Furthermore, in the 3-note dissonance output, the procedure suffers from the same issue; Trivial chords (e.g. unison with unison) show up often and pollute the more "interesting" chords (e.g. perfect third with perfect fifth) because mathematically they are more consonant.

All source code, papers and presentation files are available at the project's GitHub repository *https://github.com/krtchnt/consonare*

References

[1] minutephysics, "The Physics Of Dissonance." Accessed: Jul. 19, 2025. [Online]. Available: https://www.youtube.com/watch?v=tCsl6ZcY9ag

[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. New York: Springer, 1998. doi: 10.1007/978-0-387-21603-4.

[3] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002, doi: 10.1121/1.1458024.

[4] J. O. S. III, "Quadratic Interpolation of Spectral Peaks." 2011.

[5] "Electroacoustics – Sound Level Meters – Part 1: Specifications," no. IEC61672–1. 2013.

[6] W. A. Sethares, "Local Consonance and the Relationship between Timbre and Scale," *Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1218–1228, 1993, doi: 10.1121/1.408175.

[7] R. Plomp and W. J. M. Levelt, "Tonal Consonance and Critical Bandwidth," *Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965, doi: 10.1121/1.1909741.

[8] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1–2, pp. 103–138, 1990, doi: 10.1016/0378-5955(90)90170-T.

[9] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986, doi: 10.1109/TASSP.1986.1164910.

[10] X. Serra, "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic plus Stochastic Decomposition," 1990.

[11] P. N. Vassilakis, "Perceptual and physical properties of amplitude fluctuation and their musical significance," *Proceedings of the 7th International Conference on Music Perception and Cognition (ICMPC7)*, pp. 42–45, 2002.

[12] D. A. Green, "A colour scheme for the display of astronomical intensity images," *Bulletin of the Astronomical Society of India*, vol. 39, p. 289, 2011, [Online]. Available: https://astron-soc.in/bulletin/11June/289392011.pdf