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 by:

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

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 (for periodicity),
- Spectral peak picking with quadratic interpolation (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 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 $\Delta f = |f_j - f_i|$, $f_m = \min(f_i, f_j)$, and φ a unimodal function that peaks when Δf is around a fraction of the critical bandwidth.

We approximate the critical bandwidth with the ERB (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:

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

2) **Preprocessing**
   - High-pass at 20 Hz; low-pass at $\min(f_N, 12_{\mathrm{kHz}})$.
   - 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; 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.
   - 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. Expected Outcomes

1) **Validated algorithmic pipeline** for single-note analysis that robustly estimates f_0, extracts partials, and computes perceptually weighted dissonance/roughness profiles for two-note (and optional 3+ note) chords.
2) **Overtone tables per note** reporting (index n, frequency f_n, cents from nf_0, amplitude, perceptual weight), suitable for inclusion as figures/tables in the paper and as CSV exports.
3) **Interactive dissonance curves** $D(r)$ vs. cents with annotated minima and suggested rational approximations $\frac{p}{q}$, highlighting where perceived consonance is expected for the measured spectra.
4) **Tuning recommendations** that map spectral minima to musician-friendly interval labels (e.g., "near 7:4"), with cents offsets and confidence, for both harmonic and inharmonic sources.
5) **Inharmonicity characterisation** (when applicable), including fitted stiff-string coefficient B or an empirical deviation profile, enabling comparisons across instruments.

6) **Generalisation beyond Western tonality**, demonstrating how the method surfaces consonant relations in non-octave-centric or culturally distinct sound systems without assuming equal temperament.
7) **Benchmark dataset and scripts** comprising clean single-note recordings (and synthetic references), along with evaluation harnesses for reproducibility.
8) **Open implementation** (CLI examples) with clear documentation, enabling other researchers/musicians to run analyses and replicate figures.
9) **Ablation insights** quantifying the contribution of perceptual weighting, masking, and ERB normalisation to prediction quality and stability.
10) **Limitations write-up** describing failure modes (e.g., very sparse spectra, heavy noise, extreme inharmonicity) and guidance for future improvements.

IV. Evaluation Criteria

1) **Fundamental accuracy (synthetic ground truth)** Median absolute cents error of $\hat{f}_0$ on synthetic tones with known f_0 across SNRs ≥ 20 dB: target ≤ 3 cents; 95th-percentile ≤ 8 cents.
2) **Partial frequency/magnitude fidelity** For top-N partials (e.g., N=10), mean absolute frequency error $\leq 0.3\%$ and magnitude error ≤ 1.5 dB against synthetic ground truth; tracking $F_1 \geq 0.85$ across time.
3) **Inharmonicity fit quality** When a stiff-string model is appropriate, coefficient of determination $R^2 \geq 0.95$ for the fit $f_n \approx n f_0 \sqrt{1 + Bn^2}$; otherwise, report low residual structure in empirical deviations.
4) **Dissonance-judgment correspondence (listening test)** Rank correlation (Kendall τ or Spearman ρ) between predicted consonance ordering and listener ratings for interval sets drawn from the analysed spectra: target $\tau \geq 0.5$ (or $\rho \geq 0.7$).
5) **Minimum localisation & sharpness** For true/expected consonant relations, dissonance minima should lie within ± 7 cents of the recommended $\frac{p}{q}$ and exhibit a prominence (depth/width) exceeding a preset threshold relative to local noise.
6) **Naming/approximation accuracy** Fraction of cases where the top recommended rational $\frac{p}{q}$ matches the listener-preferred label within ± 10 cents: target $\geq 80\%$ on harmonic sources, $\geq 65\%$ on inharmonic sources.
7) **Robustness to mild pitch drift** Under controlled vibrato or drift within ± 50 cents peak-to-peak (≤ 100 -cent constraint across the note), stability of recommended intervals: median variation ≤ 5 cents.
8) **Noise robustness** Degradation curves for SNR $\in \{30, 20, 10\ \text{dB}\}$: $\hat{f}_0$ and minima localisation should remain within the targets of (1) and (5) down to 20 dB; graceful degradation at 10 dB.
9) **Ablation performance** Removing perceptual weighting, masking, or ERB normalisation should measurably worsen (4) and (5). Each ablated system must be reported with effect sizes and CIs.
10) **Computational efficiency** End-to-end runtime per two-note analysis on a standard laptop CPU (e.g., 30-60 s audio): target ≤ 2 s with N=20 partials and 1–2 cent sweep resolution.
11) **Reproducibility** Exact reproduction of figures/tables from scripts in a clean environment; checksum or hash match for exported CSVs; environment spec provided.
12) **Usability & reporting** Completeness of outputs (overtone table, dissonance plot, interval table, confidence metrics) and clarity of documentation/tutorial; short user study or heuristic checklist to ensure a musician can interpret results without acoustics background.

References

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