International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Graphical Microanalysis of Pitch and Dynamics

Arka Ashis Gupta

Student, N/A, Independent

Abstract

We present a theoretical framework for graphical microanalysis of pitch and dynamics in mixed or polyphonic audio. The framework specifies objects, mappings, and invariances in perceptual coordinates without committing to a particular estimator. Pitch is represented on the Equivalent Rectangular Bandwidth (ERB) -rate axis, with a short-term, psychoacoustically justified smoothing operator to improve analytic legibility while preserving musically relevant modulations. Dynamics modulate interpretability via a monotone "visibility" weight compatible with simultaneous masking. Figures use an explicit, invertible CIE L*a*b* encoding so quantitative values are recoverable from graphics. We formalize reference pitch regions, glissando, and vibrato on the smoothed path, and we state analytic thresholds and invariances up front to discipline interpretation. The framework is score-agnostic, calibration-optional, and intended for descriptive microanalysis across styles.

Keywords: Graphical microanalysis, ERB-rate scale, Perceptual smoothing, Dynamics visualization, Continuous Pitch representation.

INTRODUCTION

Graphical microanalysis seeks to read musically meaningful structure directly from time–frequency representations: the stability of a pitch region, the onset and slope of a glide, the presence and depth of vibrato, and the interpretive role of dynamics. In practice, however, raw instantaneous pitch traces are visually faithful yet analytically fragile. Small estimator jitter, beat-level interference, and microscopic fluctuations dominate the eye, while dynamics are often rendered decoratively, without a principled link to what can be reliably inferred about pitch. What is missing is a theory that specifies, independent of any specific signal-processing pipeline, what counts as the "same" pitch in a display, when a glide is analytically real, how vibrato should be recognized as modulation rather than noise, and how loudness should modulate the confidence with which pitch is read.

This paper provides such a theory. We define a descriptive framework for graphical microanalysis that operates entirely in perceptual coordinates and remains agnostic to the choice of estimator. The framework addresses interpretation, not extraction: it prescribes how to read and compare pitch–dynamics curves once they exist, but it does not constrain how those curves are obtained.

The coordinate system is the ERB-rate axis $E(f) = 21.4\log_{10}(1 + 4.37\,f/1000)$, which linearizes auditory frequency resolution and supports tolerance statements in units aligned with human hearing. Musical intervals are reported in cents via the usual invertible map; when the manuscript moves between ERB-rate, Hz, and cents, we make this explicit so that all thresholds and claims are transparently portable.

Because raw instantaneous paths are analytically hostile, we adopt a short-term, psychoacoustically justified smoothing operator acting on the ERB-rate pitch path. This operator represents auditory temporal integration rather than algorithmic convenience: its role is to attenuate sub-phonemic fluctuations while preserving musically relevant modulations such as ordinary vibrato and notational glissandi. Throughout, categorical judgments are made on the smoothed path with declared thresholds. This "perceptual smoothing" regularizes the display without binding the framework to any particular front end.

Within this setting we formalize the basic analytic primitives. A **reference pitch region** is a time span on which the smoothed ERB-rate path remains within a narrow band; sameness of pitch is defined by fixed ERB-rate and cents tolerances. A glissando is a span where the smoothed cents path admits a robust linear trend whose slope exceeds a reporting threshold and whose fit is sufficiently coherent. Vibrato is a bounded modulation around a local mean with dominant rate in the canonical musical band and with depth above a nominal floor. These definitions are normative: they discipline interpretation and make graphical judgments reproducible across excerpts and readers.

Dynamics enter as visibility weights that modulate interpretability. Rather than fixing a specific loudness model, we require only that an admissible weight be monotonically related to short-term perceived loudness and reflect the sublinear contribution of energy outside the critical band of the current pitch. Figures may scale opacity or lightness so that the display itself communicates when pitch judgments are less reliable. To make dynamics quantitatively recoverable, we pair this with an invertible CIE L*a*b* encoding in which lightness carries the monotone transform of while chroma is fixed; captions disclose the mapping so that numerical values can be recovered from the graphics.

The framework is anchored by explicit invariances: time translation, monotone remapping via the ERB-rate coordinate, and amplitude scaling that affects visibility but not the pitch path itself. Just as importantly, we impose a discipline of decision-making: all categorical labels (stable region, glide, vibrato) must be reproducible from the same curve when using the same globally declared thresholds. This prevents post-hoc, excerpt-specific parameter shifts.

Our contributions are fourfold. First, we propose a perceptual-coordinate framework for pitch–dynamics microanalysis that is explicitly theory-only and estimator-agnostic. Second, we justify a minimal, short-term perceptual smoothing operator and make it the substrate for all categorical judgments. Third, we provide formal definitions of reference regions, glissando, and vibrato on the smoothed path, together with a compact parameter box that fixes analytic tolerances for reproducibility. Fourth, we specify an invertible visualization scheme for dynamics-as-visibility so that figures are not merely illustrative but measurably interpretable.

Contributions

1. A perceptual-coordinate framework for microanalysis on the ERB-rate axis with explicit, stated invariances and thresholds.
2. A perceptual smoothing operator that justifies minimal time-integration for analytic legibility while preserving musically relevant modulations.
3. Formal definitions of reference pitch regions, glissando, and vibrato on the smoothed path, plus a parameter box that fixes analytic tolerances.
4. An invertible CIE L*a*b* encoding for dynamics-as-visibility so values are recoverable from figures; score-agnostic and calibration-optional.

The remainder of the paper states the coordinate system and perceptual smoothing with a global parameter box, formalizes reference regions, glissando, and vibrato, introduces dynamics-as-visibility with an invertible color encoding, and discusses limitations and trade-offs. [3–6,8–12,15–17].

Signal model and coordinates

We represent instantaneous pitch as $p(t) = E(f(t))$in ERB-rate. Analysis proceeds on the smoothed path $p^\star(t) = (\mathcal{S}_\tau p)(t)$with $\tau \in [80,150]$ms, reflecting auditory integration. A reference pitch region on $[t_1, t_2]$ satisfies $\max_{t\in[t_1,t_2]} |p^\star(t) - \bar{p}^\star| \le \delta_E$, with $\bar{p}^\star$the interval mean and δ_Ea small ERB-rate tolerance (default 0.20 ERBr). For musical reporting we also use a cents tolerance δ_c(default 20 cents). On the smoothed cents path $q^\star(t) = 1200 \log_2 \frac{f^\star(t)}{f_{\mathrm{ref}}}$:

Glissando is any span where a robust linear fit yields $\left|\frac{dq^\star}{dt}\right| \ge 40$cents/s with $R^2 \ge 0.8$.

Vibrato is a bounded modulation with dominant frequency in 3–8 Hz and depth above a nominal 10–20 cents floor.

Invariances. Time translation; monotone remapping via $E(\cdot)$; amplitude scaling affects only visibility weights (below), not $p^\star$.

Parameter (defaults). $\tau = 100$ms; $\delta_E = 0.20$ERBr; $\delta_c = 20$cents.

Dynamics as visibility. Let $W(t) \in [0,1]$be a visibility weight that is monotonically related to short-term perceived loudness under simultaneous masking. We do not commit to a particular model; admissible Wmust satisfy: (i) local increases in energy produce nondecreasing W; (ii) energy outside the critical band of $f^\star(t)$contributes sublinearly. Figures may scale opacity by $W(t)$to indicate when pitch judgments are less reliable.

Color mapping

We encode the visibility weight W(t) (or any admissible loudness proxy) with color.

Normalization

For a chosen loudness measure L(t) (e.g., LUFS, phons, or log N), define $\ell(t) = \mathrm{clip}\,(\frac{(L(t)-L_{\min})}{L_{\max}-L_{\min}}, 0, 1)$, with $L_{\min} = P_5(L),\ L_{\max} = P_{95}(L)$ per file (or calibrated absolute bounds).

Perceptual color space

Map to CIE L*a*b*: $L^*(t) = 100\,\ell(t)^\gamma,\ a^*(t) = a_0,\ b^*(t) = b_0$, with fixed chroma ($a_0$,$b_0$) for amplitude and optional hue offsets per track ID for separability. Convert L*a*b*→sRGB for display.

Invertibility (recovering dB/phon from pixels)

$\ell(t) = \left(\frac{L^*(t)}{100}\right)^{\frac{1}{\gamma}}, L(t) = L_{\min} + \ell(t)\,(L_{\max} - L_{\min})$. Report ($\gamma$, $L_{\min}$, $L_{\max}$, a_0, b_0) in figure captions.

Audibility bounds

Frames with L(t) below a perceptual floor (e.g., <20 phons or below masked threshold) are rendered with reduced alpha; frames above a comfort ceiling (e.g., >100 phons or above P_{95}) are clipped and optionally tinted to indicate saturation.

Reliability and uncertainty

Interpretive confidence. Because this is a descriptive framework, we communicate uncertainty through visibility W(t), explicit thresholds (τ, δ_E, δ_c), and disclosure of when categorical decisions change under small perturbations of these thresholds. Authors using a specific estimator may optionally include estimator-internal proxies, but such proxies are not part of the framework.

Limitations

The framework is descriptive rather than transcriptive. Perceptual smoothing trades microscopic fidelity for macroscopic legibility; very rapid ornaments shorter than τ will be attenuated, and extremely shallow glides may fall below the reporting slope. In textures with strong inharmonic or noisy content, $W(t)$ may be low even when $f^{\star}(t)$ is defined. Absolute loudness requires calibration; simultaneous masking is treated axiomatically via $W(t)$, not by a specific model.

Conclusion

We presented a theory-only framework for graphical microanalysis of pitch and dynamics in perceptual coordinates. By working on the ERB-rate axis and adopting a short-term, psychoacoustically justified smoothing operator, the framework makes stable regions, glissandi, and vibrato legible without binding to an estimator. Dynamics modulate interpretability via a monotone visibility weight, and an invertible CIE L*a*b* encoding makes figures quantitatively recoverable. With fixed thresholds and stated invariances, the framework supports reproducible, score-agnostic microanalysis across styles.

Acknowledgements

I thank Mr. Nolan Miranda for his valuable mentorship; Ms. Shaiphali Saxena for steady guidance; and Mr. Yusuf Khan for mathematical support that strengthened the analytic core of this work. I am grateful to my family—my sister, Prithvisha; my parents and grandparents — for patience and belief. Finally, thanks to PomPom, whose quiet company made long nights tolerable. Any remaining errors are my own.

References

1. J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoustics, Speech, and Signal Processing, 1977.
2. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE, 1978.
3. B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," JASA, 1983.
4. B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hearing Research, 1990.
5. H. W. Kuhn, "The Hungarian method for the assignment problem," Naval Research Logistics Quarterly, 1955.
6. J. Munkres, "Algorithms for the assignment and transportation problems," Journal of the Society for Industrial and Applied Mathematics, 1957.
7. M. Slaney, Auditory Toolbox, Tech. Rep. #1998-010, Interval Research, 1998.
8. ISO 226:2003, Acoustics — Normal equal-loudness-level contours. ISO, 2003.
9. E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, 3rd ed., Springer, 2007.
10. ITU-R BS.1770-5, Algorithms to measure audio programme loudness and true-peak audio level, 2023.
11. EBU R 128, Loudness normalisation and permitted maximum level of audio signals, 2014 revision.
12. ISO 11664-4:2008, Colorimetry — Part 4: CIE 1976 L*a*b* colour space.
13. A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral

smoothness," IEEE Trans. Speech and Audio Processing, 2003.
14. A. P. Klapuri, "Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes," in Proc. ISMIR, 2006.
15. E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic Music Transcription: An Overview," IEEE Signal Processing Magazine, 2019.
16. D. Mürbe, M. Pabst, et al., "Effects of professional singing education on vocal vibrato," Journal of Voice, 2007.
17. ISO/CIE, "Colorimetry — Part 4: CIE 1976 L*a*b* colour space," ISO/CIE 11664-4:2019, 2019.
18. CIE, "Colorimetry (4th ed.)," CIE 015:2018, 2018.