MODELLING ROUGHNESS AND HARMONICITY FROM MUSICAL AUDIO

Joshua Frank and Peter M C Harrison
Centre for Music and Science, University of Cambridge

BACKGROUND + AIMS

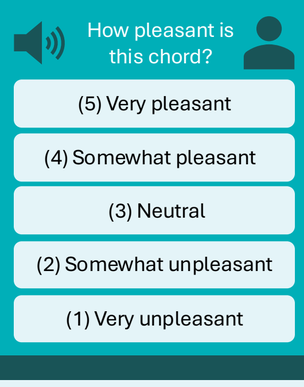
- Roughness + harmonicity strongly predict aesthetic responses to chords [1].
- **Problem:** Existing models generalise poorly to audio inputs [1].
- **Aim:** Develop improved models and benchmark them against existing techniques (MIRToolbox, [7]).

ROUGHNESS
An unpleasant "grating" sensation in sound; occurs due to nearby partials interacting, resulting in periodic loudness fluctuations called beating.

HARMONICITY
The degree to which a chord's spectral content resembles a single harmonic complex tone. Higher harmonicity = more pleasant.

VALIDATION DATASETS

- Composite dataset of stimuli and behavioural data from [2, 3, 4, 5].
- Behavioural data captures aesthetic evaluation of chords.
- In total we have >23,000 ratings for model validation.

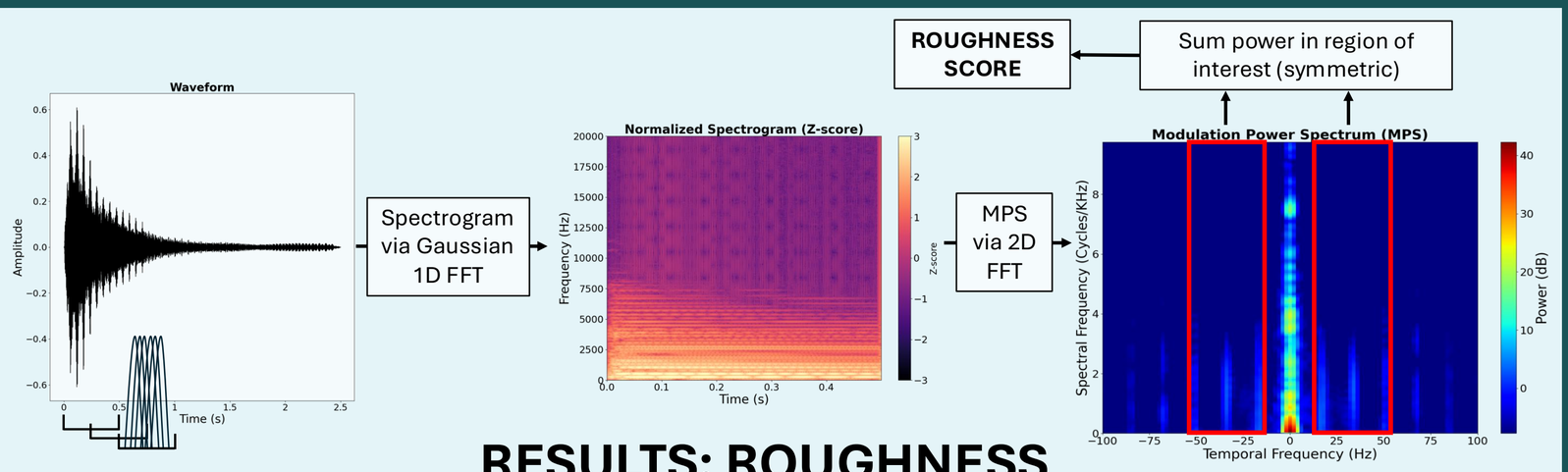


Dataset	N chord types	N participants	Timbre
Bowling *et al.* [2]	298	30	Piano
Frank & Harrison [3]	68	106	Shepard tones
Lahdelma & Eerola [4]	15	410	Piano
Parncutt *et al.* [5] A (Piano)	19	21	Piano
Parncutt *et al.* [5] B (OCT)	19	21	Octave-complex tones

METHOD: ROUGHNESS

- We adapt the *soundsig* Python package to extract the modulation power spectrum (MPS) from audio files.
- MPS: 2D Fourier Transform on a spectrogram, quantifies energy at different temporal modulation rates (TMR; i.e., beating rates) [10].
- Sum energy over a range of TMR values to get overall roughness estimate.
- Optimised range (from behavioural data): 11 – 65 Hz (consistent with prior research [6]).

SPECTROGRAM
A 3D representation of a sound over time, broken down into frequency/pitch components, and showing how the amplitudes of each component change over time.



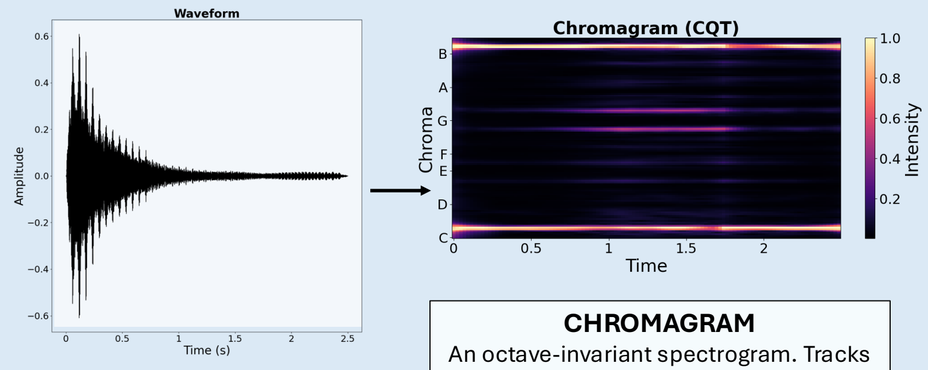
RESULTS: ROUGHNESS

- We compare MPS to existing models [7] chosen for their performance in a prior review [1].
- MPS yields the best (most strongly negative) correlation on average, and for all but one dataset.

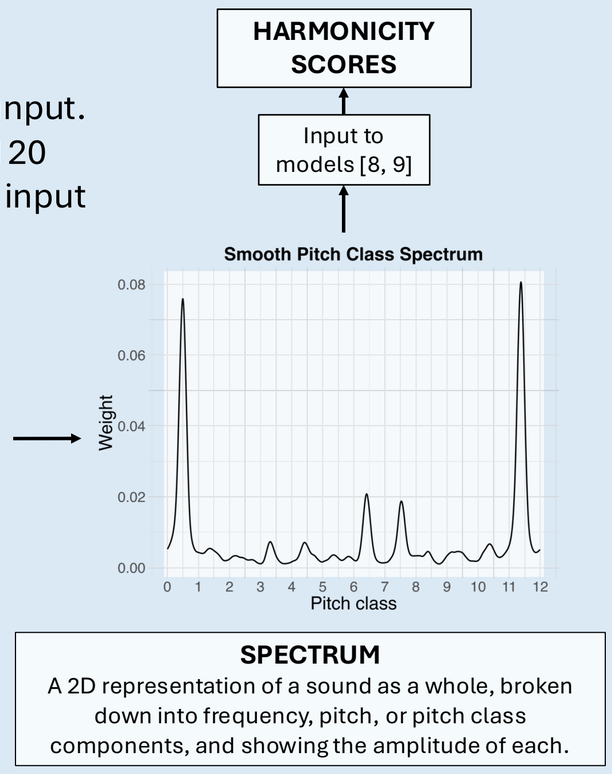
Model	Bowling *et al.* [2]	Frank & Harrison [3]	Lahdelma & Eerola [4]	Parncutt *et al.* [5] A	Parncutt *et al.* [5] B	Weighted avg. correlation
MPS	***-.73***	***-.88***	***-.89***	-.59	***-.76***	***-.76***
Sethares 1	-.57	-.82	-.75	***-.73***	-.57	-.62
Sethares 2	-.45	-.85	-.84	-.55	-.52	-.54
Vassilakis	-.32	-.84	-.53	.28	-.45	-.40

METHOD: HARMONICITY

- Existing harmonicity models [8, 9] do not support audio input.
- Our approach: generate high-resolution chromagrams (120 chroma bins), convert them to pitch-class spectra, then input them to existing models.



CHROMAGRAM
An octave-invariant spectrogram. Tracks the energy of pitch classes over time.



SPECTRUM
A 2D representation of a sound as a whole, broken down into frequency, pitch, or pitch class components, and showing the amplitude of each.

RESULTS: HARMONICITY

- We test these pipelines against one existing model [7].
- Other models exist but have been argued to be inappropriate for chords [1].
- Both [8] and [9] outperform the existing alternative on average and on all but one dataset.
- The model of [8] shows the best overall performance.

Model	Bowling *et al.* [2]	Frank & Harrison [3]	Lahdelma & Eerola [4]	Parncutt *et al.* [5] A	Parncutt *et al.* [5] B	Weighted avg. correlation
Milne [9]	.54	.66	.46	.21	***.72***	.55
Harrison & Pearce [8]	***.67***	***.76***	.18	***.39***	.15	***.65***
MIRToolbox Inharmonicity	.19	-.03	***.50***	-.07	-.15	.15

EXPLANATORY MODELLING POWER

- Multiple regression over the full dataset: combined MPS and chromagram-based version of [8] outperform the combined best MIRToolbox methods.

Model	Adjusted R^2	β Roughness	β Harmonicity
MPS + Harrison & Pearce	.63	-0.57	0.31
Sethares 1 + MIR Inharmonicity	.40	-0.63	0.03 (n.s.)

DISCUSSION

- We compare and validate several methods for extracting roughness and harmonicity from musical audio.
- MPS is the best-performing roughness model, and chromagram-based implementation of [8] is the best-performing harmonicity model.
- In combination, these models provide a robust increase in explanatory power compared to prior models.
- Effective and computationally efficient method for psychoacoustic feature extraction from musical audio.

NEXT STEPS

- How sensitive are models to other parameters? E.g., manipulations of stimulus length, register, etc.
- How well do the models behave on real music?
- Compare MPS method to a similar approach recently proposed in [10].

REFERENCES

[1] Harrison, P. M. C., & Pearce, M. T. (2020). Simultaneous Consonance in Music Perception and Composition. *Psychological Review, 127*(2): 216–244. https://doi.org/10.1037/rev0000169
[2] Bowling, D. L., Purves, D., & Gill, K. Z. (2018). Vocal similarity predicts the relative attraction of musical chords. *Proceedings of the National Academy of Sciences, 115*(1): 216–221. https://doi.org/10.1073/pnas.1713206115
[3] Frank, J., & Harrison, P. M. C. Modelling individual differences in chord pleasantness. PsyArXiv Preprint. https://doi.org/10.31234/osf.io/5hrbm_v1
[4] Lahdelma, I., & Eerola, T. (2016). Mild Dissonance Preferred Over Consonance in Single Chord Perception. *i-Perception, 7*(3). https://doi.org/10.1177/2041669516655812
[5] Parncutt, R., Engel, I., & Radovanovic, L. (2023). Consonance and Dissonance of Simultaneous Trichords in Western Music: A Listening Experiment to Test Models of Harmonicity and Roughness. *Psychological Topics, 32*(1): 13-34. https://doi.org/10.31820/pt.32.1.2
[6] Plomp, R., & Levelt, W. J. M. (1965). Tonal Consonance and Critical Bandwidth. *Journal of the Acoustical Society of America, 38*(4): 548–560. https://doi.org/10.1121/1.1909741
[7] Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for Music Information Retrieval. In C. Preisach, H. Burkhardt, L. Schmidt- Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications*, 261–268. Berlin: Springer.
[8] Harrison, P. M. C., & Pearce, M. T. (2018). An Energy-Based Generative Sequence Model For Testing Sensory Theories of Western Harmony. *Proceedings of the 19th ISMIR Conference*, 160–167. https://doi.org/10.48550/arXiv.1807.00790
[9] Milne, A. J. (2013). A computational model of the cognition of tonality (PhD thesis). The Open University, Milton Keynes, England.
[10] Anikin, A. (2025). Acoustic estimation of voice roughness. *Attention, Perception, & Psychophysics, 87*: 1771–1787. https://doi.org/10.3758/s13414-025-03060-3

ACKNOWLEDGMENTS

UKRI Economic and Social Research Council