Modelling roughness and harmonicity from musical audio

Joshua Frank (jbf43@cam.ac.uk)
Peter M C Harrison (pmch2@cam.ac.uk)
Centre for Music and Science, University of Cambridge

Abstract for ICCCM 2025

Background

Music psychologists commonly extract low-level psychoacoustic features from musical data. When considering musical chords, roughness and harmonicity are key predictors of aesthetic responses. These can be modelled successfully from symbolic data, but existing techniques to model them from audio data perform poorly by comparison (Harrison & Pearce, 2020). Modelling directly from audio is desirable for naturalistic large-scale musical data, as these are more often more easily accessible in audio than symbolic form.

Aims

We aim to develop audio-based models of psychoacoustic roughness and harmonicity that outperform existing implementations and approach the performance of symbolic models.

Methods

We construct two novel models, validating them on four behavioural datasets. For roughness, our model uses the modulation power spectrum, tuned to find an appropriate range of amplitude modulation rates which are perceptually relevant for aesthetic appraisals. For harmonicity, our model uses high-resolution chromagrams as a bridge to input audio data into Harrison & Pearce's (2018) model.

Results

Our models reach or exceed the performance of existing audio models, and perform nearly as well as the best-in-class symbolic models. For the roughness model, we find an optimal range of 11 to 65 Hz modulation rate for predicting aesthetic responses. This contrasts with findings of differences in the 30 to 200 Hz range between consonant and dissonant intervals (Arnal *et al.*, 2015) but accords well with the literature on roughness perception (e.g., Plomp & Levelt, 1965).

Conclusions

We offer two new models for the extraction of psychoacoustic roughness and harmonicity from audio data that reach or exceed the performance of prior audio-based models and approach the performance of best-in-class symbolic models. We also find

correlational evidence that the aesthetic response to roughness is minimally influenced by high temporal modulation rates.

References

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A., & Poeppel, D. (2015). Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology, 25*(15): 2051-2056. https://doi.org/10.1016/j.cub.2015.06.043

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

Frank, J., & Harrison, P. M. C. Modelling individual differences in chord pleasantness. PsyArXiv Preprint. https://doi.org/10.31234/osf.io/5hrbm_v1

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 International Society for Music Information Retrieval Conference*, 160–167. https://doi.org/10.48550/arXiv.1807.00790

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

Lahdelma, I., & Eerola, T. (2016). Mild Dissonance Preferred Over Consonance in Single Chord Perception. *i-Perception, 7*(3). https://doi.org/10.1177/2041669516655812

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

Plomp, R., & Levelt, W. J. M. (1965). Tonal Consonance and Critical Bandwidth. *The Journal of the Acoustical Society of America, 38*: 548–560. https://doi.org/10.1121/1.1909741