Relying on AI in music analysis: An evaluation of predictive consistency

Jackie Zhou[1], Cameron J. Anderson[1], Michael Schutz[1,2]

[1] Department of Psychology, Neuroscience & Behaviour, McMaster University
[2] School of the Arts, McMaster University

MAPLE Lab — Music Acoustics Perception Learning

Visit our website at **maplelab.net** !
facebook.com/maplelab.mcmaster
twitter.com/maple_lab

Background

How much time do you spend on Spotify skipping through songs in your recommended playlist?

Music Information Retrieval (MIR) and Music Emotion Recognition (MER)

- Interdisciplinary field of research: music psychology, music theory, computer science
- Extracting musical features in audio to predict emotion (Kim et al. 2010)

Past studies involving MIR

- Perceptual studies: feature extraction
- Machine learning paradigms
- Efficacy unknown

MIRtoolbox
librosa

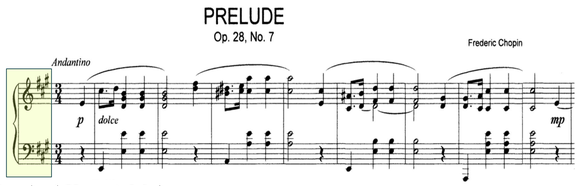
24 Preludes Sets

- Generalizable to music we listen to
- Balanced in major and minor modes
- Many performed variations

Research Question and Hypothesis

The efficacy of MIR systems have yet to be thoroughly analyzed, thus, to evaluate the consistency of MIR algorithms, we created a data set and compared algorithmically predicted mode to other measures of mode.

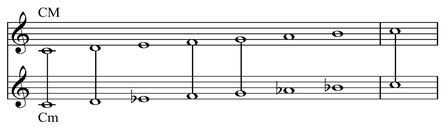
Methods

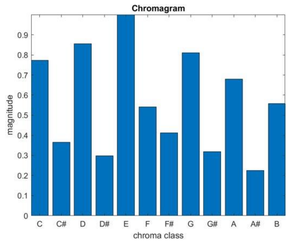


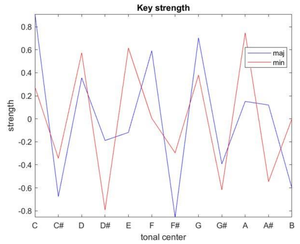
Nominal Key: A Major

Musical Modality

Structural syntax (specific grouping of notes) that contribute to the emotional character of music



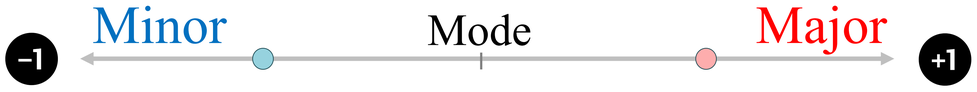




Perceptual Listening Mode-Vote Experiment

Major
Minor

+1
−1



Findings

MIRtoolbox 1.8.1

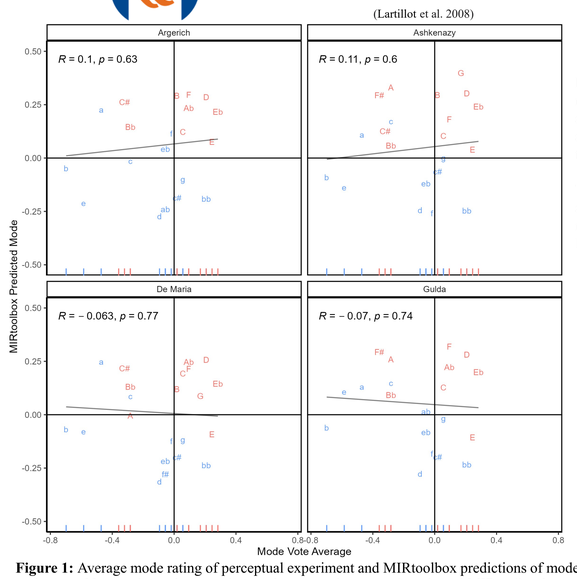


Figure 1: Average mode rating of perceptual experiment and MIRtoolbox predictions of mode in each piece. No correlation between listener's averaged perceived mode and MIRtoolbox's predicted mode, further revealing inconsistencies in MIRtoolbox's ability to predict perceptual information.

librosa

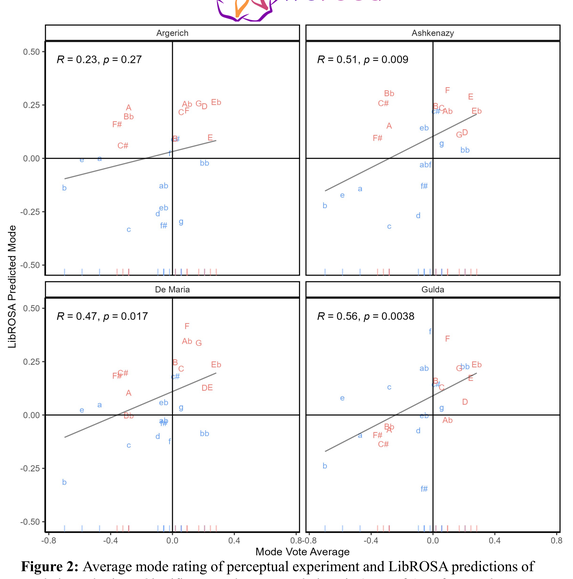


Figure 2: Average mode rating of perceptual experiment and LibROSA predictions of mode in each piece. Significant moderate correlations in 3 out of 4 performers between perceived mode and LibROSA's predicted mode.

Perceptual experiment mode average on the x-axis, computed mode average predictions on the y-axis.

The data are coloured and labeled by its nominal key. Separated by performers.

Mode
A Nominal Major
a Nominal Minor

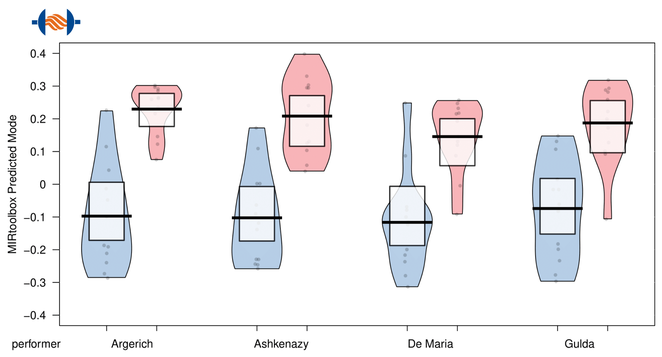


Figure 3: MIRtoolbox estimates of nominal mode. Boxplots with bean density comparing the relationship between MIRtoolbox estimates and nominal mode. Coloured by nominal mode and separated by performer.

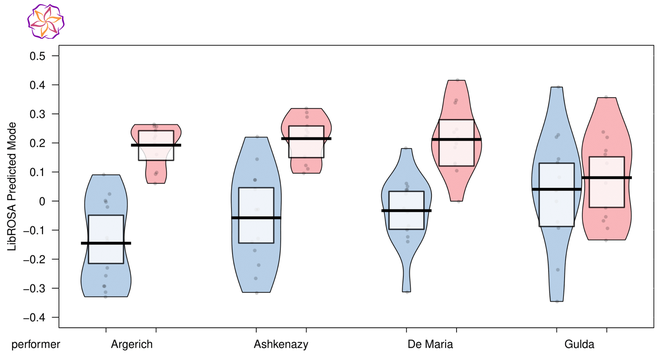


Figure 4: LibROSA estimates of nominal mode. Boxplots with bean density comparing the relationship between LibROSA estimates and nominal mode. Coloured by nominal mode and separated by performer.

Summary

- We compared algorithmic predictions of modality from two MIR systems (MIRtoolbox and LibROSA) to the composer-declared nominal mode and an experiment of listener's perceived mode.
- Results reveal no predictive consistency for MIRtoolbox's predictions of perceptual mode, whereas LibROSA's predictions of mode moderately correlated with perceptual mode for three of the four performance interpretations.
- Both systems provide a general predictive alignment with the nominal mode, with MIRtoolbox having a clearer, more distinct alignment.
- Results of this study highlight the differences between algorithms and may reveal specific use cases for each.
- Future work in examining piecewise differences may reveal interesting trends with inconsistent algorithmic predictions.

Acknowledgments

Ministry of Ontario Research and Innovation
Canada Foundation for Innovation / Fondation canadienne pour l'innovation
NSERC CRSNG
Social Sciences and Humanities Research Council of Canada / Conseil de recherches en sciences humaines du Canada
Canada

Selected References

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., ... & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In Proc. ismir (Vol. 86, pp. 937-952).

Lartillot, O., Toiviainen, P., Eerola, T. (2008). A Matlab Toolbox for Music Information Retrieval. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization.* Springer, Berlin, Heidelberg.

Schedl, M., Flexer, A. & Urbano, J. (2008) The neglected user in music information retrieval research. *J Intell Inf Syst* **41**, 523–539.