Emotion Perception in Music — Contribution of Primary Music Cues: Critique and Remodeling

Bingying Liu

Master in Interdisciplinary Data Science at Duke University

Background Information

Collecting **music emotion data** is a very subjective process, since people have different understandings of one piece of music under different contexts and associate music with their memories and life experience. In the research field of music emotion, researchers often separate **extra-musical factors** with **factors inherent in music** that cause emotions to be perceived by listeners.

T. Eerola, in order to study the latter aspect, conducted an optimized fractional factorial study on 7 primary cues and collected emotional ratings by **46 participants on a total of 200 musical samples** according to 4 perceived emotional characters (happy, sad, scary and peaceful). In his paper "Emotional expression in music: contribution, linearity, and additivity of primary musical cues", Eerola argued that many primary cues contributed to emotion in a linear and non-interactive fashion. However, my exploration of this dataset yields a different result, contradicting his linearity of cue levels argument.

Experiment Design

A full factorial experiment is an experiment whose design consists of two or more factors, each with discrete possible values or "levels", and whose experimental units take on all possible combinations of these levels across all such factors (wiki). If T. Erola is going to conduct a full factorial experiment, he needs to manipulate 6 × 2 × 5 × 5 × 4 × 3 × 4 = 14400 musical samples. Instead, he did an **optimized fractional factorial study** with 200 musical samples that he claimed could "allow the research resources to be concentrated on particular questions, thereby minimizing redundancy and maximizing the statistical power". However, I doubt if some interaction effects are effaced in this reduction process.

Research Question

- **How do primary cues contribute to different emotional expressions in music?**
- **Can we predict both happiness and sadness rating for music given single set of primary cues?**
- **Critiques of T. Eerola's original model**

Data Set

Outcome Variables	Description
Sad, Happy, Scary and Peaceful	Emotion ratings for sadness, happiness, scariness and peacefulness

Predictor Variables	Description
Register (6 scalar levels)	Whole piece was transposed so that the average pitches of the melody were the following: F3, B3, F4, B4, F5, B5/ 53, 59, 65, 71, 77 and 83 in MIDI pitch
Mode (2 factor levels)	1 = Major, 2 = Minor
Tempo (5 scalar levels)	Average number of non-simultaneous onsets per second (1.2, 2, 2.8, 4.4 and 6 NPS)
Sound Level (5 scalar levels)	-10, -5, 0, +5, +10 dB
Articulation (4 scalar levels)	Duration of a note relative to its interonset interval
Timbre (3 scalar levels)	The estimation of brightness: 1 = flute, 2 = horn, 3 = trumpet
Music Structure (4 factor categories)	| 1 = Sad, 2 = Happy, 3 = Scary, 4 = Peaceful

Although predictors all seem to be categorical variables, we treat register, tempo, sound level, articulation and timbre as **numerical variables** since they can be increasing or decreasing linearly (also the researcher chose certain factor levels so that linearity is achieved.)

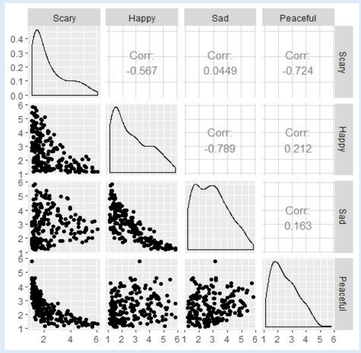
Mode and Music Structure are two **categorical variables**.

Exploratory Data Analysis

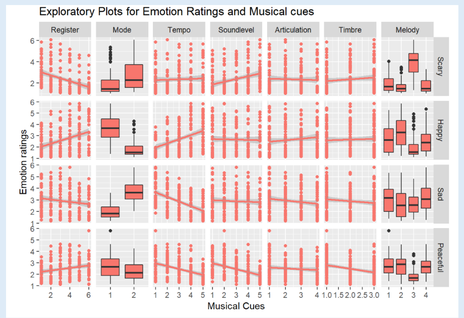
1. Correlation between perceived emotions

Happy is weakly correlated with peaceful (corr=0.212) and correlated negatively with Sad (corr=-0.789) as well as Scary (corr=-0.567).

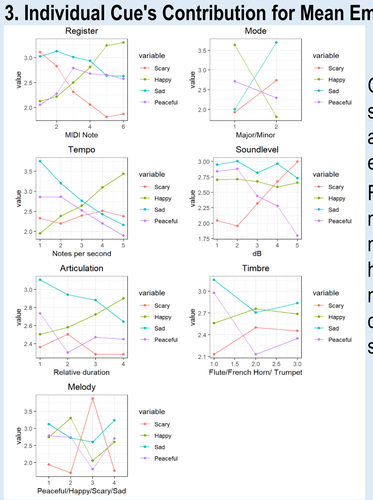
Sad is weakly correlated with peaceful (corr=0.163) and shows almost no correlation with scary. Peaceful is significantly correlated with scary in a negative fashion.



2. Emotion ratings and Cue Levels



3. Individual Cue's Contribution for Mean Emotion Rating



Graphs on the left hand side show clear patterns across the levels for four emotions.

For instance, major mode corresponds to music with higher happiness rating and minor mode corresponds to higher sadness rating.

Modelling

Stage 1: Simple Linear Regressions

T. Eerola seemed to make a fatal mistake by not checking residual plots for all the regression models he made. The residual versus fitted value plot above shows a quadratic trend (LOESS function is not a horizontal line) and also a fanning problem for outcome variable happy.

The same problem exists in all four regressions for emotions. Even though later on Eerola tried quadratic and even cubic numerical variables such as Register and Tempo, the fact that non-constant variance assumption wasn't met was the biggest problem hindering him to make correct conclusions. His linearity of cue level argument was challenged.

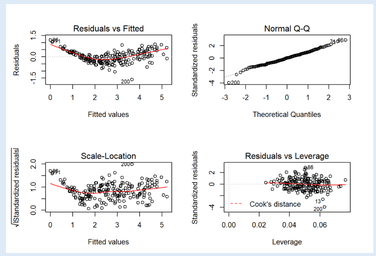


Figure: Problem of T. Eerola's first model

Stage 2: Log Encoding and Quadratic Terms

To satisfy the non-constant variance assumption, I logarithmzed all the outcome variables.

To correct quadratic trends in residual plots for emotional expression log(scary) and log(peaceful), I also added quadratic terms.

Specifically, I added register^2 in scary model and register^2, timbre^2 and tempo^2 in peaceful model. Careful check of interaction plots shows no interesting interaction between musical cues. I'm satisfied with these models since most predictors have really significant p-values and r-squares for these set of models are between 0.879 to 0.933.

```
##
## Calls:
## Model 1: lm(formula = loghappy ~ Register + as.factor(Mode) + Tempo +
##     Articulation + Timbre + as.factor(Melody), data = music_emotion)
## Model 2: lm(formula = logscary ~ Register + Register2 + as.factor(Mode) +
##     Tempo + Soundlevel + Timbre + as.factor(Melody), data = music_emotion)
## Model 3: lm(formula = logsad ~ Register + as.factor(Mode) + Tempo + Soundlevel +
##     Articulation + Timbre + as.factor(Melody), data = music_emotion)
## Model 4: lm(formula = logpeaceful ~ Register + Register2 + as.factor(Mode) +
##     Tempo + Tempo2 + Soundlevel + Articulation + Timbre + Timbre2 +
##     as.factor(Melody), data = music_emotion)
##
##                                Model 1    Model 2    Model 3    Model 4
##                              ---------- ---------- ---------- ----------
##                              loghappy   logscary   logsad     logpeaceful
## ------------------------------------------------------------------------
##   (Intercept)                 0.515***   0.432***   1.496***   2.116***
##                              (0.044)    (0.069)    (0.048)    (0.094)
##   Register                   0.087***  -0.316***  -0.022***   0.221***
##                              (0.005)    (0.033)    (0.003)    (0.027)
##   as.factor(Mode): 2/1       -0.692***   0.339***   0.620***  -0.167***
##                              (0.018)    (0.023)    (0.018)    (0.019)
##   Tempo                       0.116***   0.036***  -0.135***   0.047
##                              (0.006)    (0.008)    (0.006)    (0.034)
##   Articulation                0.072***                -0.068***  -0.029**
##                              (0.008)                (0.008)    (0.009)
##   Timbre                     -0.022*     0.100***  -0.034**   -0.929***
##                              (0.011)    (0.014)    (0.011)    (0.082)
##   as.factor(Melody): 2/1      0.185***  -0.101**   -0.163***  -0.045
##                              (0.025)    (0.032)    (0.025)    (0.026)
##   as.factor(Melody): 3/1     -0.293***   0.748***  -0.178***  -0.477***
##                              (0.025)    (0.032)    (0.025)    (0.026)
##   as.factor(Melody): 4/1     -0.053*    -0.052      0.030     -0.077**
##                              (0.026)    (0.034)    (0.027)    (0.028)
##   Register2                                0.026***           -0.024***
##                                           (0.005)            (0.004)
##   Soundlevel                               0.111***  -0.020**  -0.104***
##                                           (0.008)    (0.007)  (0.007)
##   Tempo2                                                      -0.026***
##                                                               (0.006)
##   Timbre2                                                      0.190***
##                                                               (0.020)
## ------------------------------------------------------------------------
##   sigma                       0.126      0.163      0.129      0.133
##   R-squared                   0.933      0.901      0.906      0.879
##   F                         334.108    191.074    204.327    113.534
##   p                           0.000      0.000      0.000      0.000
##   N                         200        200        200        200
## ========================================================================
```

Conclusions

- **Happiness**

An increase in **register** of 1 level is associated with a multiplicative change of $e^{0.087} \approx 1.09$ (95% CI: 1.08, 1.10) in happiness rating.
Changing **mode** from major to minor is associated with a multiplicative change of $e^{-0.69} \approx 0.50$ (95% CI: 0.48, 0.51) in happiness rating.
An increase in **tempo** of 1 level is associated with a multiplicative change of $e^{0.12} \approx 1.13$ (95% CI: 1.11, 1.14) in happiness rating.
An increase in duration of a note (**articulation**) of 1 level is associated with a multiplicative change of $e^{0.07} \approx 1.07$ (95% CI: 1.06, 1.09)in happiness rating.
An increase in **brightness of timbre** of 1 level is associated with a multiplicative change of $e^{-0.02} \approx 0.98$ (95% CI: 0.96, 1.00) in happiness rating.
Changing **melody** from **sad to happy** is associated with a multiplicative change of $e^{0.19} \approx 1.2$ (95% CI: 1.15, 1.26) in happiness rating.
Changing **melody** from **sad to scary** is associated with a multiplicative change of $e^{-0.29} \approx 0.75$ (95% CI: 0.71, 0.78) in happiness rating.
Changing **melody** from **sad to peaceful** is associated with a multiplicative change of $e^{-0.05} \approx 0.95$ (95% CI: 0.90, 1.00) in happiness rating.
The fact that F statistic is pretty big shows that sound level has little association with happy mood.

Mode seems extremely important for happy music. And if we are to rank importance for other predictors, then the order would be melody, tempo, register, articulation and timbre. (conclusions of sadness, scariness and peacefulness could be seen in paper).

Multivariate Multiple Regression—Happiness and Sadness

Multivariate multiple regression (MMR) is the method of modeling multiple response variables with a single set of predictor variables. I would like to model both happiness and sadness level as a function of register, mode, tempo, sound level and etc.

MMR is very similar with running separate regressions with each outcome variables. The only difference lies in hypotheses tests for regression parameters and confidence interval for prediction. Let's run an MMR for loghappy and logsad.

Summary results are the same as running separate regressions. But coefficients from two models covary when we check the variance-covariance matrix and we should consider covariance when determining if a predictor is jointly contributing to both models.

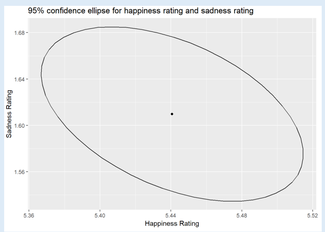
The effect of sound level doesn't seem to be significant in loghappy. Thus, we need to formally test whether to include sound level in multivariate model using Anova command.

```
## Type II MANOVA Tests: Pillai test statistic
##                Df test stat approx F num Df den Df    Pr(>F)
## Register        1   0.60046   142.02      2    189 < 2.2e-16 ***
## as.factor(Mode) 1   0.90441   896.17      2    189 < 2.2e-16 ***
## Tempo           1   0.73787   266.01      2    189 < 2.2e-16 ***
## Soundlevel      1   0.06085     6.12      2    189  0.002651 **
## Articulation    1   0.33973    48.42      2    189 < 2.2e-16 ***
## Timbre          1   0.11303    12.04      2    189 1.195e-05 ***
## as.factor(Melody)  3   1.05912    71.29      6    380 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sound level appears to be jointly significant for the multivariate models though in individual loghappy model it seems not as important as other predictors. P-value for sound level is 0.002651.

Let's try to predict happiness and sadness rating for a piece of music!

Register = 83 in MIDI pitch
Mode = major
Tempo = 4.4 NPS
Sound level = -5dB
Articulation = staccato
Timbre = horn
Melody = peaceful



This is a **95% confidence ellipse for happiness and sadness rating**, which is equivalent to confidence interval in 1d setting.

Black dot in the center is our predicted value for happy and sad.

We're 95% confident that the true values of happy and sad when Register = 6, Mode = 1, Tempo = 4, Sound level = 2, Articulation = 4, Timbre = 2 and Melody = 4 are within area of the ellipse.

Happy and sad are negatively correlated. Predicting higher rating of happiness means predicting lower rating of sadness, and vice versa.

Critiques of T. Eerola's Models

T. Eerola didn't check residual plots for all the regression models he made. Residual plots of his final model (plain linear regression) has non-constant variance problem. I remodeled the same set of data by observing patterns in exploratory process and carefully checking residual plots.

Eerola overfit this data set in the second stage by quadratic/cubic encoding all numerical variables. An optimized fractional factorial design might in its nature exclude interaction effects between predictor variables.

Citation:

- *"Emotional expression in music: contribution, linearity, and additivity of primary musical cues", Tuomas Eerola, Anders Friberg, and Roberto Bresin , https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726864/*
- *"Getting started with Multivariate Multiple Regression", Clay Ford, https://data.library.virginia.edu/getting-started-with-multivariate-multiple-regression/*
- *"Factorial Experiment", https://en.wikipedia.org/wiki/Factorial_experiment*