Influence of Instrument, Harmonic Motion, and Voice Leading on Listeners' Identification of Music as "Classical" or "Popular"

Gunho Jung*
gjung@andrew.cmu.edu

08 December 2019

Abstract

In this research paper, we examine how respondents rate the music to be Classical vs Popular by utilizing Classical and Popular Ratings data (Jimenez and Rossi 2012). First, we examine which experimental factors (Instrument, Harmony, and Voice) had the strong influences on ratings by interpreting the coefficients and t-values of the selected hierarchical models. Then, we look at how driving factors for predicting Classical vs Popular ratings are similar and different by comparing the statistically significant covariates from the updated hierarchical models. Lastly, we identify how self-declared musicians are different from non-musicians for the Classical music by interpreting the interaction terms between Selfdeclare and other covariates after differently dichotomizing the Selfdelcare variable. We find that Instrument has the strongest influences on both Classical and Popular, followed by Harmony and Voice. Also, we detect that there are both similarities and differences in the driving factors. Lastly, we consistently identify that self-declared musicians less rely on the Instrument type than the non-musicians and the musicians give higher classical ratings than the non-musicians if the musical piece followed the harmonic motion of I-V-VI. These interpretations were slightly sensitive to the dichotomization.

1 Introduction

Music, an art of sound in time that expresses ideas and emotions in significant forms through the element of rhythm, melody, harmony, and color (Dictionary.com 2019), can be found everywhere across the globe and distinctively influenced by the surrounding environments, cultures, time periods, locations, and etc. Among these various influences, in this paper, we will pay special attentions to the following experimental factors to the musical piece: *"Instrument, Harmonic Motion, and Voice Leading"*, and how they have influences on the listeners' distinction of the musical piece as "Classical" or "Popular" (how classical/popular does the music sound in a scale of 1-10). Specifically, we will be answering to the primary experiment question of **[1] Which experimental factor, or combination of factors, has the strongest influences on ratings?** by utilizing the evidences from the subsequent questions:

- Does Instrument exert the strongest influence among the three design factors (Instrument, Harmonic Motion,Voice Leading), as the researchers suspect?
- Among the levels of Harmonic Motion does I-V-VI have a strong association (the strongest) with classical ratings? Does it seem to matter whether the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits?

*Master's in Statistical Practice, Carnegie Mellon University

- Among the levels of Voice Leading, does contrary motion have a strong (the strongest) association with classical ratings?

Then, we will move onto answering to the following questions from the researchers: **[2] Are there differences in the things that drive Classical vs Popular ratings?** and **[3] Depending on whether the listeners are musicians or non-musicians, are there any differences in the way that they identify Classical music?**

2 Methods

Overview of Data

The data for the analyses are provided by Ivan Jimenez, a composer and musicologist visiting the University of Pittsburgh, and a student Vincent Rossi (2012) and contain 36 musical stimuli to 70 listeners, recruited from the population of undergraduates at the University of Pittsburgh. We excluded `X, first12, Instr.minus.Notes, X1stInstr, X2ndInstr` columns from the dataset, which were not utilized / irrelevant in this experiment.

Variables relevant to Classical and Popular ratings	
Variable Name	Description
Classical	How classical does the stimulus sound? (scale 1-10)
Popular	How popular does the stimulus sound? (scale 1-10)
Subject	Unique subject ID
Harmony	Harmonic Motion (4 levels)
Instrument	Instrument (3 levels)
Voice	Voice Leading (3 levels)
Selfdeclare	Are you a musician? (scale 1-6)
OMSI	Score on a test of musical knowledge
X16.minus.17	Auxiliary measure of listener's ability to distinguish classical vs popular music
ConsInstr	How much did you concentrate on the instrument? (scale 0-5)
ConsNotes	How much did you concentrate on the notes? (scale 0-5)
PachListen	How familiar with Pachelbel's Canon in D? (scale 0-5)
ClsListen	How much do you listen to classical music? (scale 0-5)
KnowRob	Have you heard Rob Paravonian's Pachelbel Rant? (scale 0-5)
KnowAxis	Have you heard Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music? (scale 0-5)
X1990s2000s	How much do you listen to pop and rock from the 90's and 2000's? (scale 0-5)
X1990s2000s.minus.1960s1970s	Difference between X1990s2000s variable and a similar variable referring to 1960's and 1970's pop and rock
CollegeMusic	Have you taken music classes in college? (0 = no, 1 = yes)
NoClass	How many music classes have you taken?
APTheory	Did you take AP Music Theory class in High School? (0 = no, 1 = yes)
Composing	Have you done any music composing? (scale 0-5)
PianoPlay	Do you play piano? (scale 0-5)
GuitarPlay	Do you play guitar? (scale 0-5)

Figure 1: Variable definitions for Classical and Popular ratings data from Jimenez and Rossi (2012)

2.1 Data Cleaning / Imputation & Exploratory Data Analysis

Before directly looking at the statistical evidences to answer to the research questions, we first conducted an univariate exploratory data analysis on all the variables from the original dataset to (1) identify and decide how to handle the missing values, (2) understand the distributions and look for if any transformations were necessary or not. For the response variables `Classical, Popular`, we eliminated the entire row if the value was missing, considering that it was hard to estimate what the actual values would be. Also, for the rows with either Classical or Popular values of 19, we converted them to 10, assuming that they were mistakes from inputting the data manually. However, for each of all the other variables, we conducted median value imputations if there were any missing values. After the imputations on missing values, we looked at the distributions of the continuous variables and determined whether we should consider transforming some variables or not. After understanding the distribution of each variable, we moved onto conducting a multivariate exploratory data analysis to identify any interesting relationships among variables. We created pairs plot to look at the two-dimensional relationship among the continuous variables and created boxplots to look at the relationship between the response variables vs categorical variables.

2.2 Experimental Factors (Instrument, Harmony, Voice) and Classical Ratings

While acknowledging the relationships between one of the response variable `Classical` vs predictors and predictor vs other predictors from multivariate EDA, we specifically focused on three experimental factors `Instrument, Harmony, Voice`. Referring back to our primary experiment question of *"which experimental factor, or combination of factors, has the strongest influences on ratings?"*, we decided to fit a multiple linear regression model of `Classical` as a response variable and `Instrument, Harmony, Voice` as predictors. As a first step, we considered for all potential interactions of all orders and conducted a series of variable selections through utilizing statistical evidences from stepAIC, stepBIC, and partial F-test (ANOVA). Once we selected variables for the multiple linear regression model, we tested whether the random intercept was needed in the model (Restricted Likelihood Ratio Tests (exactRLRT)), considering that the experiment had approximately 36 ratings from each participant. Then, we improved this hierarchical model by first adding one or more new random effect terms (to account for "personal biases") and decided which random effect was the best based on the AIC and BIC values (lowest). As a last step, with this selected hierarchical model, we looked for statistical evidences that could answer to the subsequent questions pertained to these experimental factors.

2.3 Different driving factors for Classical vs Popular Ratings

After getting a clear idea of which experimental factors had the strongest influences on Classical ratings, we shifted gears to answer to the second experiment question of *"are there differences in the things that drive Classical vs Popular ratings?* As a first step, we fitted all the other predictors into the existing hierarchical model and conducted stepAIC and stepBIC to select additional covariates (fixed effects) that were statistically significant to predict Classical ratings. Then, we utilized an automatic method called `fitLMER.fnc` to identify if the selected random effects would be different after adding more fixed effects. Based on the result of the automatic method, we finalized our model to predict Classical ratings. Similarly, we followed exactly same approaches to determine the predictors to predict Popular ratings. With these two selected hierarchical models (one for predicting Classical ratings and the other for predicting Popular ratings), we compared how these selected fixed effects and random effects were similar and different from each other as a way of providing answers to the posed experiment question.

2.4 Differences between Musicians and Non-musicians for the Classical Music

Last but not least, referring back to the researchers' final experiment question, we decided to analyze how self-declared musicians were different from non-musicians with respect to the Classical music. We first created two different versions of dataframes with variations in the `Selfdeclare` column values only. For the first dataframe, if the subject's `Selfdeclare` value was greater than 2, we assigned this subject to be Musician and if the value was less than 3, we assigned the subject to be Non-musician. For the second dataframe, if the subject's `Selfdeclare` value was greater than 3, we assigned the subject to be Musician and if the value was less than 4, we assigned the subject to be Non-musician. This way, we were able to set the stage whether the results were sensitive to how we dichotomized the `Selfdeclare` variable. Then, while holding the random effects fixed (excluding the random effects), we fitted two different multiple linear regression models (one model for each dataframe) with the predictors selected previously and their interactions with this updated `Selfdeclare` variable. Then, we conducted variable selections by utilizing the results from stepAIC. We fitted two final models with these selected interaction terms, identified which of these interactions were statistically significant, and analyzed (1) how self-declared musicians were different from non-musicians by looking at the coefficient estimates of these interaction terms and (2) whether the results of statistically significant interaction terms were sensitive to how we dichotomized `Selfdeclare` variable.

3 Results

In this section, we will re-address the experiment questions that we posed from the introduction and then answer them by referencing to the statistical evidences that we detected. First, we are going to interpret the results of both univariate and multivariate exploratory data analysis. Then, we will move onto analyzing experimental factors' influences on Classical ratings. Furthermore, we will compare how driving factors (including three main experimental factors) were different in terms of predicting classical vs popular ratings. Lastly, we will identify whether there were any differences between self-declared musicians vs non-musicians for the classical music particularly.

3.1 Exploratory Data Analysis

As a first step, we conducted an univariate EDA after cleaning the data (dropping rows with missing classical and popular ratings and excluding 5 columns that were irrelevant in this experiment) and imputing the missing values with the median values within each column. We created histograms for the continuous variables to understand the distributions. We treated Classical and Popular variables to be continuous but treated all the other ordinal variables (scaled) to be categorical. In total, there were four continuous variables: `Classical, Popular, OMSI, NoClass`.

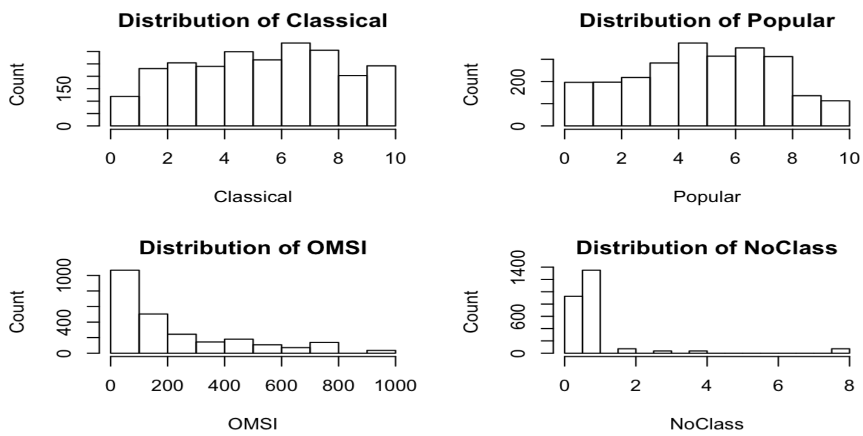


Figure 2: Univariate EDA on continuous variables

According to Figure 2, we detected that both Classical and Popular variables were not heavily skewed whereas both OMSI and NoClass variables were heavily skewed to the right due to very few extremely high values. We decided to only log-transform the OMSI variable in order to make the distribution to roughly resemble normal distribution while maintain the interpretability. We did not conduct any transformation on NoClass variable.

Moving onto the multivariate EDA, we created pairs plot to identify two-dimensional relationships among these four continuous variables and created boxplots to identify two-dimensional relationships between the two response variables vs categorical variables.

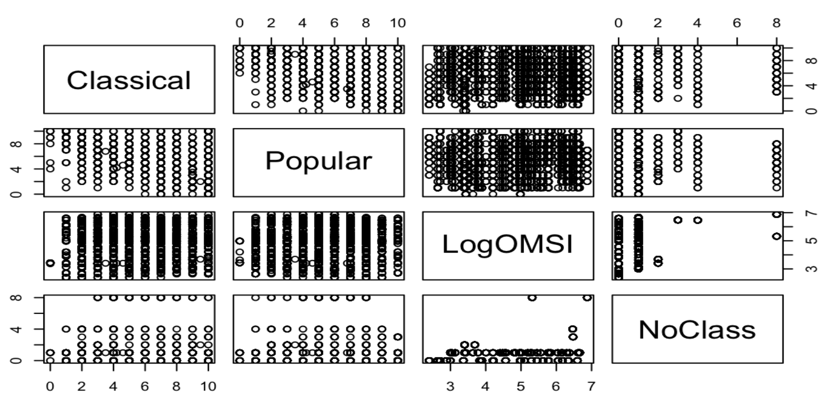


Figure 3: Multivariate EDA on continuous variables

From the pairs plot (Figure 3), we did not identify any interesting relationship among the continuous variables. One potential relationship that we need to keep in mind is between `NoClass` and `LogOMSI`. Even though the relationship was not apparent, we could get an idea that as the number of music class taken increases, LogOMSI value also tends to increase.

From the boxplots of response variables vs categorical variables (refer to Appendix A. Figure 6 and 7 on pg. 15), an interesting relationship that we identified was that if the instrument was string quartet, the classical ratings was the highest, followed by piano and guitar. On the other hand, if the instrument was guitar, the popular ratings was the highest, followed by piano and string quartet. Furthermore, we detected that if the music followed harmonic motion of I-V-VI, the classical rating was the highest while all the other harmonic motions had similar classical ratings. On the other hand, if the music followed harmonic motion of I-IV-V or I-V-IV, the popular rating was higher than the musical with either of I-V-VI or IV-I-V harmonic motion. In summary, from the multivariate EDA, we detected that both instrument and harmony variables formed relationships with Classical and Popular response variables.

3.2 Experimental Factors and Classical Ratings

After identifying two-dimensional relationships between the response variables and predictors, we decided to dive more deeply into looking at the influences of experimental factors (`Instrument, Harmony, Voice`) on one of the response variable `Classical`. In order to answer to our first experiment question of "*Which experimental factor, or combination of factors, has the strongest influences on ratings?*", we selected the hierarchical model (with these experimental factors) that best predicted the Classical ratings through conducting a series of variable selections for both fixed and random effects.

The chosen model had `Instrument, Harmony, Voice` and interaction between `Harmony` and `Voice` as the fixed effects and for each subject, the model had `Instrument, Harmony` as the random effects, while allowing Instrument and Harmony to be correlated. With this hierarchical model, we conducted marginal, conditional, and random effects residual analysis and did not detect any strong violations of the assumptions (refer to Appendix B.1 Figure 8 on pg. 16).

Since the model satisfied the assumptions, we moved forward to look for the influences of the experimental factors on classical ratings by looking at the statistical summary of the chosen model.

Fixed Effects Statistical Summary			
Fixed Effects	Estimate	Std. Error	t value
Intercept	4.255	0.223	19.092
Instrumentpiano	1.369	0.171	8.004
Instrumentstring	3.117	0.235	13.243
HarmonyI-V-IV	0.155	0.154	1.009
HarmonyI-V-VI	1.140	0.214	5.330
HarmonyIV-I-V	-0.177	0.151	-1.171
Voicepar3rd	-0.270	0.151	-1.789
Voicepar5th	-0.236	0.151	-1.565
HarmonyI-V-IV:Voicepar3rd	-0.366	0.214	-1.711
HarmonyI-V-VI:Voicepar3rd	-0.680	0.214	-3.181
HarmonyIV-I-V:Voicepar3rd	0.528	0.214	2.475
HarmonyI-V-IV:Voicepar5th	-0.189	0.214	-0.881
HarmonyI-V-VI:Voicepar5th	-0.425	0.214	-1.988
HarmonyIV-I-V:Voicepar5th	0.120	0.213	0.561

Figure 4: Statistical Summary of the Fixed Effects from the selected model

Figure 4 illustrates the fixed effects' estimates, standard errors, and t values. The intercept represents the baseline of our model, in which when the musical piece had the guitar as an instrument, followed harmonic motion of I-IV-V, and had the voice as Contrary. For the fixed effect

`Instrument`, while holding other variables constant, on average, if the musical piece was played in piano, the scale of Classical rating is expected to increase by 1.369, and if the musical piece was played in string quartet, the scale of Classical rating is expected to increase by 3.117 from the baseline of our model. Similar interpretations applied for two other fixed effects `Harmony, Voice`. For the fixed effect of the interaction term between `Harmony` and `Voice`, for example, while fixing the Voice to be "Parallel 3rd" and all the other fixed effects constant, on average, if the musical piece followed the harmonic motion of I-V-IV, the scale of Classical rating is expected to decrease by 0.211 (-0.366 + 0.155).

Random Effects Statistical Summary				
Groups	Name	Variance	Std.Dev.	Correlation
Subject	Intercept	2.550	1.597	
	Instrumentpiano	1.641	1.281	-0.39
	Instrumentstring	3.479	1.865	-0.57 0.66
	HarmonyI-V-IV	0.051	0.226	0.62 -0.60 -0.42
	HarmonyI-V-VI	1.600	1.265	-0.05 -0.26 -0.42 0.21
	HarmonyIV-I-V	0.011	0.103	0.17 -0.33 -0.03 -0.21 0.07
Residual		2.371	1.540	

Figure 5: Statistical Summary of the Random Effects from the selected model

Figure 5 illustrates the random effects' variances, standard deviations, and correlations. Referring back to our random effects (`Instrument + Harmony | Subject`), we observed that the variance of the random effect `Instrument` was relatively larger than the variance of the random effect `Harmony`. This suggested that the random effect of Instrument varied more than the random effect of Harmony did. When we compared the variances of these two random effects to the estimated residual variances, we detected that the variance of `Instrument = string quartet` was larger than the variance of the estimated residuals, meaning that including Instrument as the random effect for each `Subject` was statistically valid. For the variance of the `Harmony`, none of the categories had variances greater than the variance of the estimated residuals, indicating that adding Harmony as the random effect was not significant. However, we maintained these two random effects, because these two random effects had significantly lower AIC and BIC values than the model with only Instrument as the random effect(refer to Appendix B.2 Figure 9 on pg. 16).

Referring back to the first subsequent question of the first experiment question, we detected that it was `Instrument` that was most influential (as the researchers suspected) with the largest fixed effect coefficient estimate and t-value. Moving onto the second subsequent question, for the `Harmony` factor, it was particularly influential when the musical piece followed the harmonic motion of I-V-VI, because the coefficient estimate and the t-value was the highest among all the other harmonic motions. Furthermore, the harmonic motion of I-V-VI seemed to matter with the Pachelbel Rant when the respondent was greatly familiar with it as the coefficient and the t-value were the largest (refer to Appendix B.3. Figure 10 on pg. 17). To answer to the third subsequent question, we did not detect any statistical evidences that `Voice` was influential, because the 95% confidence intervals of the coefficient estimates contained 0 for all categories. In other words, among the levels of Voice Leading, we did not detect any evidences that the contrary motion had the strongest association with classical ratings.

Finally, the interaction between `Harmony` and `Voice`, particularly between harmonic motion of I-V-VI and Voice of Parallel 3rd was shown to have some moderate influence on the classical ratings. Last but not least, for the random effects, we found that `Instrument` varied greatly among the subjects whereas `Harmony` did not as the variances of all categories were less than the estimated residual variances.

3.3 Different driving factors for Classical vs Popular Ratings

While maintaining these three experimental factors and the interaction term, we moved onto identifying whether there were any additional variables that were statistically significant so that we could answer to the second experiment question of *"are there differences in the things that drive Classical vs Popular ratings?"*

3.3.1 Selecting additional fixed, random effects for Classical ratings

To additionally select covariates for the new model, we fitted all the other predictors to our model and selected the variables by using the result of the automatic methods stepAIC and stepBIC. Specifically, these additional factors were determined by the overlap between the stepAIC and stepBIC. The overlapping additional factors were: `Selfdeclare, X16.minus.17, ConsInstr, PachListen, ClsListen, KnowRob, KnowAxis, X1990s2000s, X1990s2000s.minus.1960s 1970s, APTheory, Composing, PianoPlay`. However, we decided to eliminate the `X1990s2000s` variable from the model, because this variable was strongly correlated with `X1990s2000s.minus.1960s1970s`, and X1990s2000s.minus.1960s1970s variable can be calculated from the calculation of X1990s2000s and imaginary X1960s1970s variable. All the other variables were both statistically and musically made sense to be included to predict the response variable. In summary, for the additional factors to predict Classical ratings (for the fixed effects of our hierarchical model), we decided to include `Selfdeclare, X16.minus.17, ConsInstr, PachListen, ClsListen, KnowRob, KnowAxis, X1990s2000s.minus.1960s1970s, APTheory, Composing, PianoPlay`.

Now that we confirmed the additional factors, we decided to examine if there were any changes in the random effects. In order to validate the random effects, we conducted random effects selection method using fitLMER.fnc function, once with setting method to be AIC and another with method to be BIC. The result suggested random effects to be `(Instrument|Subject) + (Harmony|Subject)`. However, since our previous selected model had the random effects to be `(Instrument + Harmony | Subject)`, we decided to compare the AIC values for the model with new random effects to previously selected random effects (refer to Appendix C.1 Figure 11 on pg. 17). The AIC value for the previous random effects was less than the new random effects, giving us an evidence that keeping the original random effects was better. Therefore, in summary, for the final model for predicting Classical ratings, we selected the fixed effects of `Instrument, Harmony, Voice, Selfdeclare, X16.minus.17, ConsInstr, PachListen, ClsListen, KnowRob, KnowAxis, X1990s2000s.minus.1960s1970s, APTheory, Composing, PianoPlay` and random effects of `(Instrument + Harmony|Subject)`.

With the newly selected driving factors to predict Classical ratings, we conducted marginal, conditional, and random effects residual analysis and did not detect any strong violations of the assumptions (refer to Appendix C.2 Figure 12 on pg. 18).

3.3.2 Experimental factors vs Popular ratings

For the Popular ratings, we repeated the same procedures that we conducted for the Classical ratings. As a first step, we started with identifying the influence of `Instrument, Harmony, Voice` on Popular ratings after conducting variable selections based on the lowest AIC and BIC values.

From the result of the two, we consistently identified that Instrument was statistically significant. However, the model summary suggested that Harmony, Voice and interaction between Harmony and Voice were not statistically significant to predict Popular ratings. Referring back to our first experiment question's first subsequent question, we found an evidence that Instrument exerted the strongest influence among the three design factors as the researchers suspected while all the other experimental factors did not have influence on the Popular ratings. However, considering that this was a designed experiment with specific intention to measure the influence of these experimental factors, we decided to keep all three variables.

Moving onto the random effects, after conducting a series of restricted likelihood ratio test and AIC & BIC values comparisons, we found that the model with the lowest AIC and BIC values was with the random effects of `(Instrument|Subject) + (Harmony|Subject)`. Referring back to the random effects for predicting Classical ratings, we found a similarity that both Instrument and Harmony were chosen for the random effects, and also found a difference that for predicting the Popular ratings, the random effects of Instrument and Harmony being uncorrelated was preferred to the random effect of Instrument and Harmony being correlated. However, musically, it made more sense to have Instrument and Harmony to be correlated within the random effects, and therefore decided to choose random effects for predicting Popular ratings to be same as what we selected for predicting Classical ratings. In other words, we decided to select `(Instrument + Harmony|Subject)` to be the random effects instead. As a side note, the chosen random effects actually was the second best option as its AIC and BIC values were second lowest (still a reasonable selection).

With this hierarchical model, we also conducted marginal, conditional, and random effects residual analysis and did not detect any strong violations of the assumptions (refer to Appendix C.3 Figure 13 on pg. 19).

3.3.3 Additional driving factors vs Popular ratings

To additionally select covariates for the hierarchical model to predict Popular ratings, we followed the same approach of **Section 3.3.1 Selecting additional fixed, random effects for Classical ratings**. The overlapping additional factors of AIC and BIC results `Selfdeclare, LogOMSI, X16.minus.17, ConsInstr, ConsNotes, PachListen, KnowAxis, X1990s2000s.minus.1960s 1970s, CollegeMusic, PianoPlay` were included in the new model, while keeping the experimental factors fixed. With these fixed effects, we moved onto examine if there were any changes in the random effects as well. By following the same previous approaches and justifications, we finalized our random effects to be `(Instrument + Harmony | Subject)`. This model's marginal, conditional, and random effects residual analysis also did not strongly violate any assumptions (refer to Appendix C.4 Figure 14 on pg. 19).

3.3.4 Comparing driving factors for Classical vs Popular Ratings

To recap, for the final model to predict Classical ratings, we selected the fixed effects of `Instrument, Harmony, Voice, Harmony:Voice, Selfdeclare, X16.minus.17, ConsInstr, PachListen,` ClsListen, KnowRob, KnowAxis, X1990s2000s.minus.1960s1970s, APTheory, Composing, PianoPlay and random effects of (Instrument + Harmony | Subject). For the final model to predict Popular ratings, we selected the fixed effects of Instrument, Harmony, Voice, Selfdeclare, LogOMSI, X16.minus.17, ConsInstr, ConsNotes, PachListen, KnowAxis, X1990s2000s.minus.1960s1970s, CollegeMusic, PianoPlay and random effects of (Instrument + Harmony | Subject). All of these models satisfied the assumptions and both statistically and musically made sense.

To identify how driving factors for Classical ratings were similar and different from those of Popular ratings, we specifically compared the newly added factors and the random effects term. For the random effects, we did not detect any differences, because both models consistently had Instrument and Harmony as the random effects for each subject. For the fixed effects, both models had Selfdeclare, X16.minus.17, ConsInstr, PachListen, KnowAxis, X1990s2000s.minus.1960s 1970s, PianoPlay as the driving factors, but Classical ratings model uniquely had additional ClsListen, KnowRob, APTheory, Composing driving factors whereas Popular ratings model uniquely had additional LogOMSI, ConsNotes, CollegeMusic driving factors. Therefore, to answer to the second experiment question, we concluded that there were both similarities and differences in the driving factors for predicting Classical vs Popular ratings.

3.4 Musicians vs Non-musicians for the Classical Music

Our final experiment question was *"Depending on whether the listeners are musicians or non-musicians, are there any differences in the way that they identify Classical music?"* In order to answer to this question, we dichotomized the Selfdeclare variable into two different ways, fitted two different models (selected fixed effects + interactions between Selfdeclare variable and all the other fixed effects), conducted variable selections through the automatic method (while fixing on the random effects), interpreted statistically significant interactions' coefficient estimates, and compared which of these interaction terms were shown in both models vs shown in one model but not in the other model.

3.4.1 Musicians vs Non-Musicians: first dichotomization

For the model with first dichotomization (assigning the subject to be musicians if the Selfdeclare value was greater than 2 and non-musicians if not), we detected that interactions between Selfdeclare and Instrument, Harmony, ConsInstr were shown to be statistically significant with p-values less than 0.05 (refer to the left image of Appendix D. Figure 15 on pg. 20). From the result of the statistically significant interaction terms, while holding other variables constant, for those respondents who defined themselves as Musicians, if the instrument played in the musical piece was either piano or string quartet, on average, they were less likely to give higher Classical ratings for the musical piece than the Non-musicians. In other words, when compared to the Non-musicians, if the musical piece was played in either piano or string quartet, the Musicians will less likely to define the music to be classical. This corresponded to the finding of Ivan Jimenez and Vincent Rossi that Instrument was more influential for ratings of non-musicians than ratings of musicians. Furthermore, while holding other variables constant, if the musical piece followed the harmonic motion of I-V-VI, on average, the musicians will more likely to give higher Classical ratings than the non-musicians. Last but not least, while holding other variables constant, on average, for the respondents who concentrated more on the instrument (greater than or equal to the rating of 3), musicians were less likely to give higher Classical ratings than the non-musicians. In other words, the more musicians concentrated on the instrument, the less likely for them to define the musical piece to be classical. This also corresponded to what we had observed from the interpretation of the interaction term between musicians and Instrument.

3.4.2 Musicians vs Non-Musicians: second dichotomization

For the model with second dichotomization (assigning the subject to be musicians if the Selfdeclare value was greater than 3 and non-musicians if not), we detected that interactions between `Selfdeclare` and `Instrument, Harmony, LogOMSI, X16.minus.17, ConsInstr` were shown to be statistically significant with p-values less than 0.05 (refer to the right image of Appendix D. Figure 15 on pg. 20). From the result of the statistically significant interaction terms, even if we dichotomize the `Selfdeclare` variable slight differently, we consistently found that while holding other variables constant, if the instrument played in the musical piece was either piano or string quartet, on average, musicians were less likely to give higher Classical ratings for the musical piece than the Non-musicians. Also, while holding other variables constant, if the musical piece followed the harmonic motion of I-V-VI, on average, the musicians will more likely to give higher Classical ratings than the non-musicians. For the interaction between LogOMSI and Selfdeclare variable, while holding other variables constant, on average, the higher the score a musician gets on OMSI (higher score percentage), the more likely he or she rates the musical piece to be classical than the non-musician. For the interaction between X16.minus.17 and Selfdeclare variable, while holding other variables constant, on average, the musicians, who have less abilities to distinguish classical vs popular music (less than measure of 4), are more likely to give higher Classical ratings than the non-musicians. Finally, while holding other variables constant, on average, for the respondents who paid moderate concentration on the Instrument (rating of 3), musicians were less likely to give higher Classical ratings than the non-musicians, which also partially corresponded to what we had observed from the first dichotomization.

3.4.3 Summary of Different Dichotomizations of Musicians vs Non-Musicians

From above, we detected that the interactions between `Selfdeclare` and `Instrument, Harmony, ConsInstr` were shown to be statistically significant no matter how we dichotomized the Selfdeclare variable. However, we additionally found that the interactions between `Selfdeclare` and `LogOMSI, X16.minus.17` were statistically significant when we dichotomized the Selfdeclare variable with threshold of 3. In other words, we found an evidence that the result of statistically significant interaction terms were very slightly sensitive to how we dichotomize the variable. Therefore, we reached to a conclusion that for the Classical ratings, people who self-identify as “Musicians” may be influenced by the things that do not influence “Non-Musicians”, and these different influences can be very slightly sensitive to how we define the respondents to be either Musicians and Non-Musicians.

4 Discussion

After extensively analyzing the Classical vs Popular ratings data, we were able to find evidences for (1) which experimental factor had the strongest influence on Classical & Popular ratings (2) which additional driving factors were significant for predicting Classical vs Popular ratings and (3) whether there were any differences in the way that self-declared musicians vs non-musicians identified Classical music.

4.1 Summary of Experimental Factors on Classical vs Popular Ratings

Referring back to the first experiment question and the result of **section 3.2**, we detected that Instrument had the strongest influence on the Classical ratings, followed by the Harmonic Motion. Specifically, while holding other variables constant, on average, if the musical piece was played by a string quartet, the respondents tended to rated the music as Classical the highest, followed by piano and guitar. If the musical piece followed the harmonic motion of I-V-VI, on average, the respondents tended to rate the music as Classical the highest, and it seemed to matter if the respondent was greatly familiar with the Pachelbel Rant. We did not detect any statistical evidences to support that Voice was influential. Furthermore, we found that the interaction between Harmony and Voice was shown to have some moderate influence on the classical ratings. Last but not least, we selected Instrument and Harmony (correlated) to be the random effects. Particularly, Instrument varied greatly among the respondents whereas Harmony did not; however, we decided to keep Harmony as the additional random effect, because it resulted in much lower both AIC and BIC values. Similarly, referring back to the result of **section 3.3.2**, we also detected that Instrument had the strongest influence on the Popular ratings, whereas Harmony and Voice were not statistically significant. For the random effects, for each respondent, we also selected Instrument and Harmony (correlated).

4.2 Summary of Different Driving Factors on Classical vs Popular Ratings

While maintaining these three experimental factors, we moved onto identifying whether there were any additional variables that were statistically significant, so that we could answer to the second experiment question. Referring to the result of **section 3.3.1**, we detected that the additional statistically significant factors to predict Classical ratings were Selfdeclare, X16.minus.17, ConsInstr, PachListen, ClsListen, KnowRob, KnowAxis, X1990s2000s.minus.1960s1970s, APTheory, Composing, and PianoPlay. For the Popular ratings (refer to **section 3.3.3**), Selfdeclare, LogOMSI, X16.minus.17, ConsInstr, ConsNotes, PachListen, KnowAxis, X1990s2000.minus.1960s1970s, CollegeMusic, PianoPlay were the additional statistically significant factors. We found that there were common driving factors for both Classical and Popular ratings, and also there were some unique driving factors for Classical vs Popular ratings. No matter which additional factors were included in each model, the random effects of Instrument and Harmony for each respondent remained the same.

4.3 Summary of Musicians vs Non-musicians for the Classical Music

Finally, with these selected driving factors, in order to answer to the last experiment question, we identified whether there were any differences in the way that the self-declared musicians vs non-musicians rated the Classical music. Referring back to the result of **section 3.4**, we found that while holding other variables constant, when compared to the non-musicians, the self-declared musicians were less likely to rate the music to be more Classical if the musical piece was played in string quartet or piano. Also, the musicians will give higher classical ratings if the music followed the harmonic motion of I-V-VI. If the musicians paid close attention to the instrument, they were less likely to give higher Classical ratings than the non-musicians. Furthermore, no matter how we defined the respondents to be (either musicians or non-musicians) based on their Selfdeclare values, these interpretations remained consistently the same. Even though we did detect additional statistically significant interaction terms when we dichotomized the Selfdeclare variable with threshold of 3, we concluded that these influences were very slightly sensitive to how we dichotomize.

4.4 Strengths, Weaknesses, and Future Experiment Directions

Even though we found various statistical evidences to provide answers to three main experiment questions (both musically and statistically made sense), we need to acknowledge that these evidences and interpretations could be different if we handled better with the missing values in the beginning. The main drawback of our missing value imputation strategy was that for each column, we simply imputed the median value. This strategy could be problematic, because it could reduce the variance of the imputed variables and did not take into consideration of relationships between variables (correlations). For instance, we observed that Harmony and Voice were not statistically significant to predict Popular ratings and random effect of Harmony was not significant for predicting both Classical and Popular ratings due to its variance being lower than the estimated residual variances. If we handled the missing values more carefully, such as conducting regression imputation, we could have preserved better relationships among variables and could have answered to the experiment questions differently. Also, the variables that we selected as additional driving factors were mainly based on the results of the automatic method, meaning that there is a possibility that more or less variables were actually significant to predict the response variables.

For the future experiment, we recommend the researchers to reevaluate these interpretations and answers after carefully re-handle the missing values. Also, if possible, we recommend adding additional column that indicates whether the musical piece was actually Classical or Popular. Even though we had one column called `X16.minus.17`, which was an auxiliary measure of listener's ability to distinguish classical vs popular music, since we did not have any information whether the music was actually classical or popular, we could not deeply analyze the significance of this variable. Therefore, if we implement these suggested future experiment directions/recommendations, the answers to the three main experiment questions will be even more convincing and accurate.

References

R Core Team (2017), *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ivan Jimenez, Vincent Rossi (2012). The Influence of Timbre, Harmony, and Voice Leading on Listeners' Distinction between Popular and Classical Music. Pittsburgh, PA, USA.

Definition of music. Dictionary.com. (2019). URL https://www.dictionary.com/browse/music.

Appendix

A. Multivariate EDA on Response Variables vs Categorical Variables

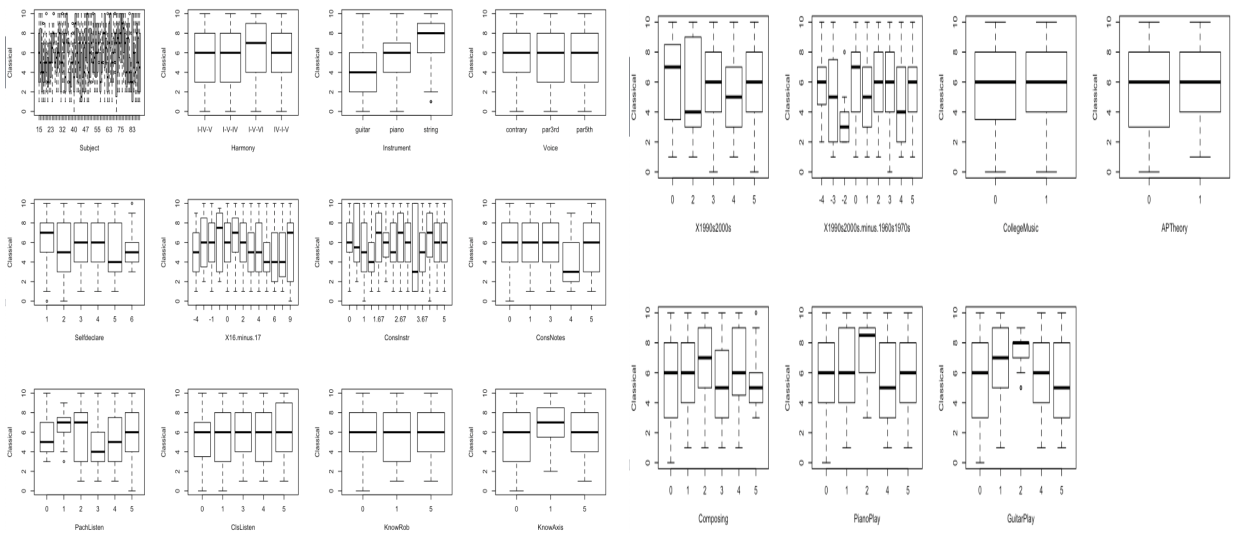


Figure 6: Multivariate EDA on Classical Ratings vs Categorical Variables

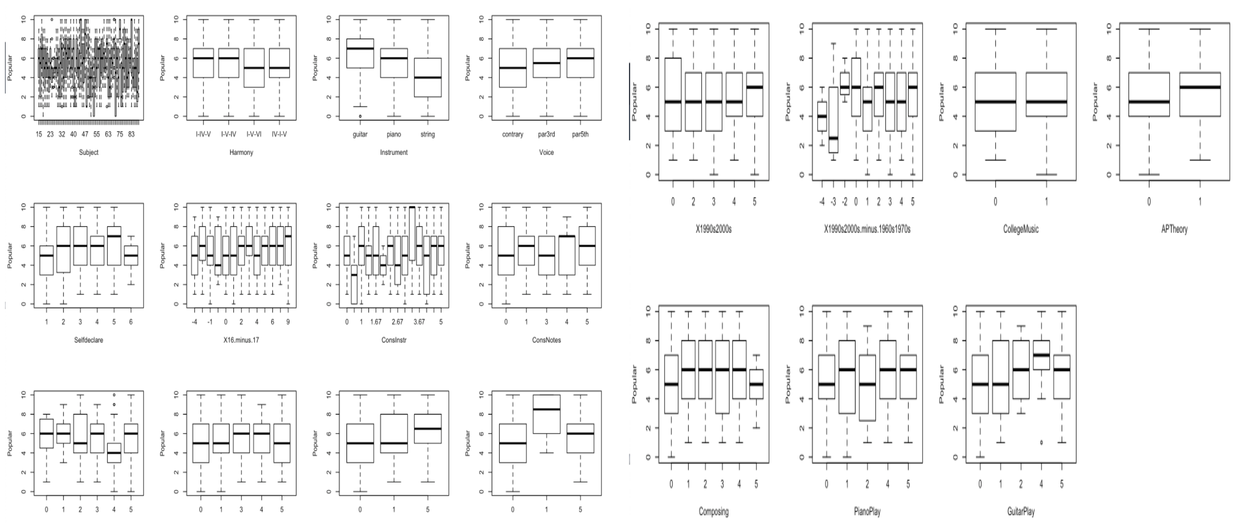


Figure 7: Multivariate EDA on Popular Ratings vs Categorical Variables

B.1. Residual Analysis of Hierarchical Model: Experimental Factors vs Classical

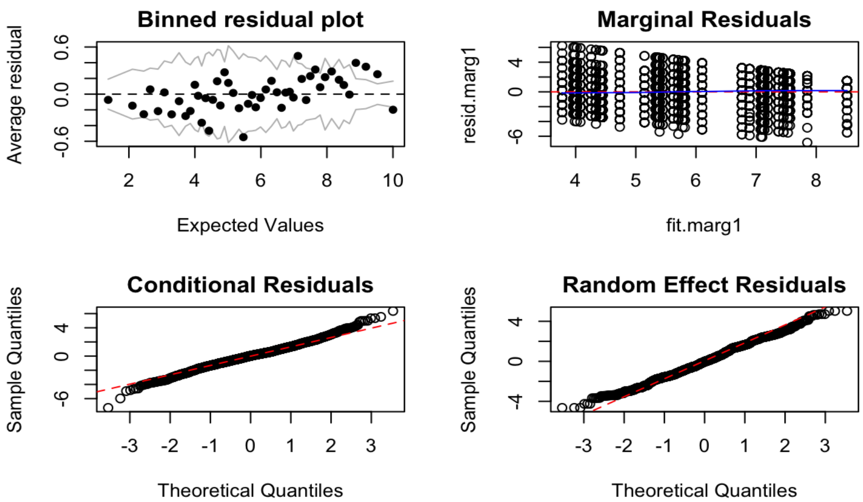
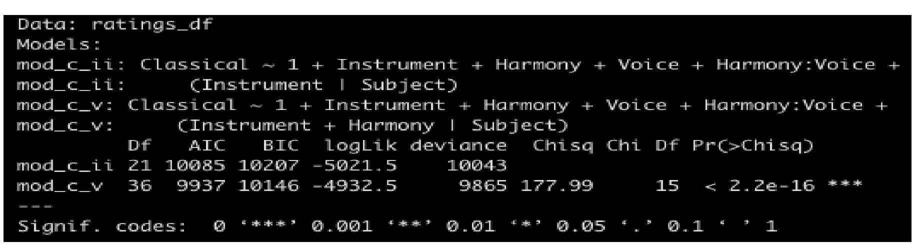


Figure 8: Residual analysis of chosen hierarchical model - experimental factors vs classical ratings

For the chosen model, we first conducted average residual analysis. From the first plot above, we noticed that the average residuals were evenly distributed along average residual = 0 line. However, we detected a slight linear pattern between expected values vs average residuals. From the second plot, just as we observed from the average residual plot, we detected that the marginal residuals were evenly scattered, but formed a slight linear pattern. The linear pattern was not problematic, because we should expect the marginal residuals to be correlated. Moving onto the conditional residuals, we did not detect any group structure and most of the residuals roughly aligned well along the normalized line. Even though there were some conditional residuals deviating from the normalized line, overall, the conditional residuals were homoskedastic and normally distributed. For the random effect residuals, we did not detect any outliers in the random effects (η's), because most of the random effect residuals were closely located to others. Also, the normality of the random effects were adequately met. In summary, since our hierarchical model does not strongly violate assumptions, we concluded that our model with three experimental factors seemed adequate.

B.2. Random Effects Selection: Experimental Factors vs Classical



```
Data: ratings_df
Models:
mod_c_ii: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
mod_c_ii:     (Instrument | Subject)
mod_c_v: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
mod_c_v:     (Instrument + Harmony | Subject)
         Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
mod_c_ii 21 10085 10207 -5021.5    10043
mod_c_v  36  9937 10146 -4932.5     9865 177.99     15  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 9: Random effects selection based on the lowest AIC and BIC values

B.3. Harmony & Pachelbel rants/comedy bits: Experimental Factors vs Classical

```
HarmonyI-V-IV:KnowRob1     0.15641    0.35819   0.437
HarmonyI-V-VI:KnowRob1     0.90890    0.59680   1.523
HarmonyIV-I-V:KnowRob1     0.39034    0.35400   1.103
HarmonyI-V-IV:KnowRob5    -0.15364    0.30661  -0.501
HarmonyI-V-VI:KnowRob5     1.94111    0.50693   3.829
HarmonyIV-I-V:KnowRob5    -0.09555    0.30160  -0.317
HarmonyI-V-IV:KnowAxis1    1.69916    0.80416   2.113
HarmonyI-V-VI:KnowAxis1   -0.67588    1.33855  -0.505
HarmonyIV-I-V:KnowAxis1    0.92252    0.79422   1.162
HarmonyI-V-IV:KnowAxis5   -0.11749    0.26880  -0.437
HarmonyI-V-VI:KnowAxis5   -0.83202    0.44708  -1.861
HarmonyIV-I-V:KnowAxis5   -0.24097    0.26537  -0.908
```

Figure 10: Interaction between Harmony and Pachelbel rants vs comedy bits

To validate whether harmonic motion seemed to matter whether the respondent was familiar with one or the other (or both) of the Pachelbel rants/comedy bits, we decided to add interactions between `Harmony` and `KnowRob` and `Harmony` and `KnowAxis` and looked at the coefficient estimates and the t-values. According to Figure 10, we detected that only the interaction between harmonic motion of I-V-VI and KnowRob of 5 (Pachelbel Rants) had strong influence on the classical ratings with coefficient estimates of 1.941 and t-value of 3.829. We did not detect any of the interactions between Harmony and KnowAxis (Pachelbel comedy bits) statistically significant since the 95% confidence intervals for all the interactions contained 0.

C.1. Random Effects Selection after selecting driving factors for Classical ratings

```
Data: ratings_df
Models:
lmer_full_aic: Classical ~ Harmony + Instrument + Voice + Selfdeclare + X16.minus.17 +
lmer_full_aic:     ConsInstr + PachListen + ClsListen + KnowRob + KnowAxis +
lmer_full_aic:     X1990s2000s.minus.1960s1970s + APTheory + Composing + PianoPlay +
lmer_full_aic:     (1 | Subject) + (Instrument | Subject) + (Harmony | Subject) +
lmer_full_aic:     Harmony:Voice
lmer_full_aic2: Classical ~ Harmony + Instrument + Voice + Selfdeclare + X16.minus.17 +
lmer_full_aic2:     ConsInstr + PachListen + ClsListen + KnowRob + KnowAxis +
lmer_full_aic2:     X1990s2000s.minus.1960s1970s + APTheory + Composing + PianoPlay +
lmer_full_aic2:     (Instrument + Harmony | Subject) + Harmony:Voice
               Df    AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer_full_aic  90 9842.5 10366 -4831.3   9662.5
lmer_full_aic2 94 9839.4 10387 -4825.7   9651.4 11.144      4    0.02499 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 11: Random effects selection after adding driving factors for Classical ratings

C.2. Residual Analysis after adding driving factors for Classical

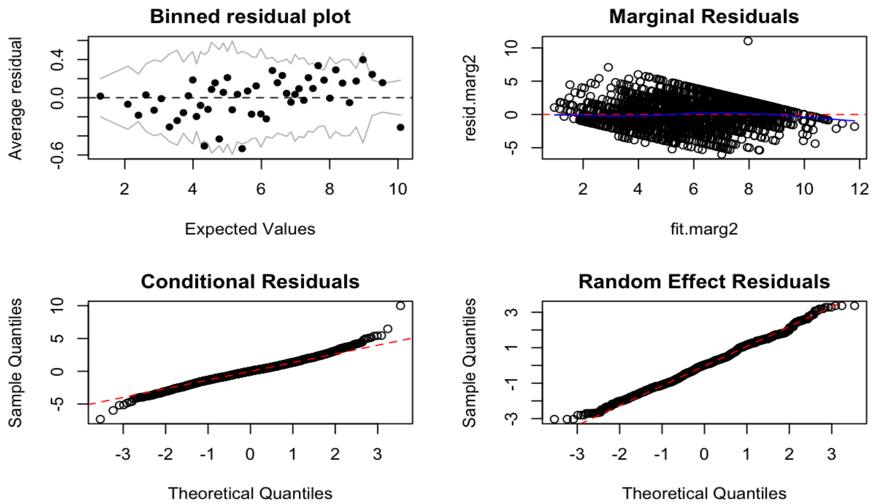


Figure 12: Residual analysis of hierarchical model after adding driving factors to predict Classical

For the newly chosen model, we first conducted average residual analysis. From the first plot above, we noticed that the average residuals were evenly distributed along average residual = 0 line. More average residuals were included within the confidence interval. When we looked at the distribution of the marginal residuals, just as we observed from previous marginal residuals before adding additional fixed effects predictors, we detected that the marginal residuals were evenly scattered, but formed a downward linear pattern. The linear pattern was not problematic, because we should expect the marginal residuals to be correlated. Moving onto the conditional residuals, we did not detect any group structure and most of the residuals roughly aligned well along the normalized line. Even though there were some conditional residuals deviating from the normalized line (potential outlier of the ϵ), overall, the conditional residuals were homoskedastic and normally distributed. For the random effect residuals, we did not detect any severe outliers in the random effects (η's), because most of the random effect residuals were closely located to others. Also, the normality of the random effects were adequately met. In summary, for the chosen model after adding additional covariates for the fixed effects, our new hierarchical model does not strongly violate assumptions and seemed adequate to predict the Classical ratings.

C.3. Residual Analysis of Hierarchical Model: Experimental Factors vs Popular

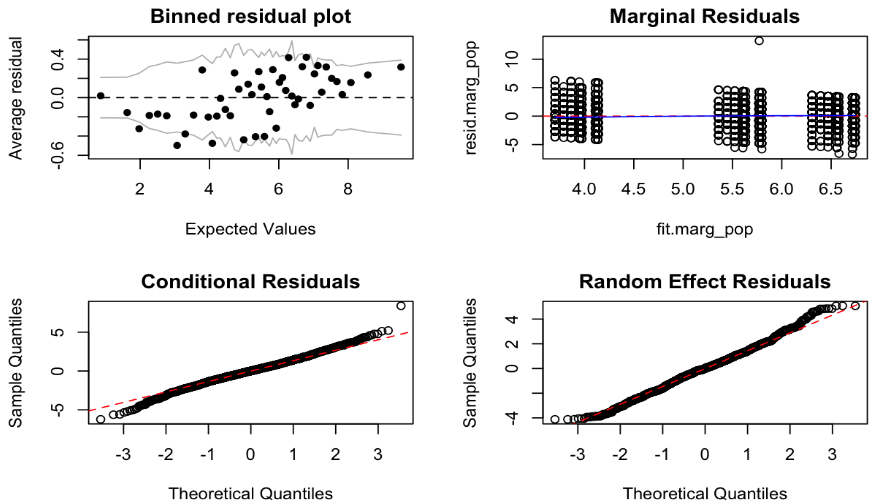


Figure 13: Residual analysis of chosen hierarchical model - experimental factors vs popular ratings

C.4. Residual Analysis after adding driving factors for Popular

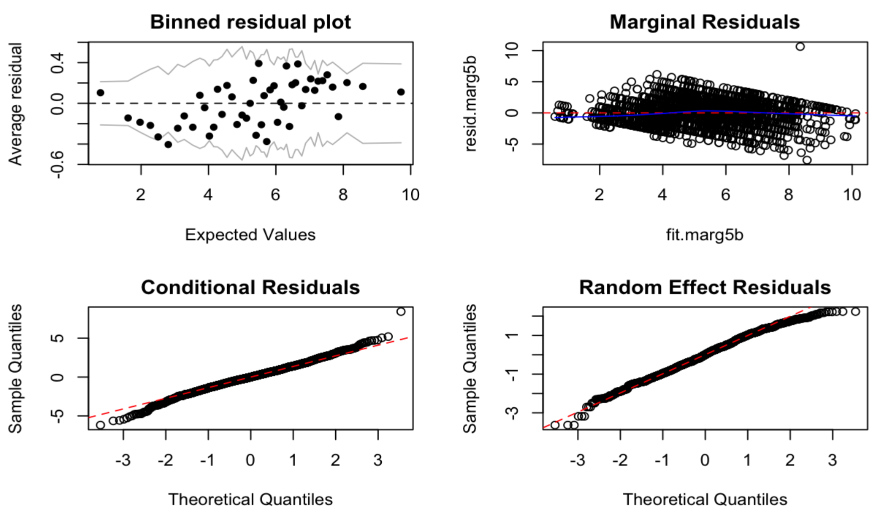


Figure 14: Residual analysis of hierarchical model after adding driving factors to predict Popular

D. Musicians vs Non-Musicians - Model Summary of the Interaction Terms

Instrumentpiano:Selfdeclare1	-0.60164	0.18769	-3.205	0.001366	**
Instrumentstring:Selfdeclare1	-0.71211	0.18603	-3.828	0.000133	***
HarmonyI-V-IV:Selfdeclare1	0.05229	0.21571	0.242	0.808490	
HarmonyI-V-VI:Selfdeclare1	1.20006	0.21570	5.564	2.94e-08	***
HarmonyIV-I-V:Selfdeclare1	0.07053	0.21551	0.327	0.743480	
Selfdeclare0:LogOMSI	-18.04977	10.47112	-1.724	0.084878	.
Selfdeclare1:LogOMSI	-37.12922	23.26262	-1.596	0.110600	
Selfdeclare1:X16.minus.17-2	NA	NA	NA	NA	
Selfdeclare1:X16.minus.17-1	274.21303	145.96051	1.879	0.060409	.
Selfdeclare1:X16.minus.17-0.5	NA	NA	NA	NA	
Selfdeclare1:X16.minus.170	291.07543	161.85694	1.798	0.072247	.
Selfdeclare1:X16.minus.171	170.13548	91.19966	1.866	0.062229	.
Selfdeclare1:X16.minus.172	NA	NA	NA	NA	
Selfdeclare1:X16.minus.173	172.00054	105.92955	1.624	0.104565	
Selfdeclare1:X16.minus.174	163.90706	103.12828	1.589	0.112113	
Selfdeclare1:X16.minus.175	NA	NA	NA	NA	
Selfdeclare1:X16.minus.176	NA	NA	NA	NA	
Selfdeclare1:X16.minus.177	NA	NA	NA	NA	
Selfdeclare1:X16.minus.179	NA	NA	NA	NA	
Selfdeclare1:ConsInstr0.67	NA	NA	NA	NA	
Selfdeclare1:ConsInstr1	NA	NA	NA	NA	
Selfdeclare1:ConsInstr1.33	NA	NA	NA	NA	
Selfdeclare1:ConsInstr1.67	-1.50190	9.85938	-0.152	0.878938	
Selfdeclare1:ConsInstr2	NA	NA	NA	NA	
Selfdeclare1:ConsInstr2.33	NA	NA	NA	NA	
Selfdeclare1:ConsInstr2.67	NA	NA	NA	NA	
Selfdeclare1:ConsInstr3	-69.06517	29.71739	-2.324	0.020205	*
Selfdeclare1:ConsInstr3.33	NA	NA	NA	NA	
Selfdeclare1:ConsInstr3.67	5.86544	11.17235	0.525	0.599634	
Selfdeclare1:ConsInstr4	NA	NA	NA	NA	
Selfdeclare1:ConsInstr4.33	-85.24732	34.76398	-2.452	0.014270	*
Selfdeclare1:ConsInstr5	-145.30135	71.64446	-2.028	0.042661	*
Selfdeclare0:ConsNotes1	31.42556	26.63013	1.180	0.238087	
Selfdeclare1:ConsNotes1	-61.85039	36.71033	-1.685	0.092153	.
Selfdeclare0:ConsNotes3	NA	NA	NA	NA	
Selfdeclare1:ConsNotes3	NA	NA	NA	NA	
Selfdeclare0:ConsNotes4	NA	NA	NA	NA	
Selfdeclare1:ConsNotes4	NA	NA	NA	NA	
Selfdeclare0:ConsNotes5	NA	NA	NA	NA	
Selfdeclare1:ConsNotes5	NA	NA	NA	NA	
Selfdeclare1:KnowAxis1	NA	NA	NA	NA	
Selfdeclare1:KnowAxis5	NA	NA	NA	NA	
Selfdeclare1:PachListen1	NA	NA	NA	NA	
Selfdeclare1:PachListen2	NA	NA	NA	NA	
Selfdeclare1:PachListen3	NA	NA	NA	NA	
Selfdeclare1:PachListen4	NA	NA	NA	NA	
Selfdeclare1:PachListen5	NA	NA	NA	NA	

Instrumentpiano:Selfdeclare1	-0.45503	0.23121	-1.968	0.049181	*
Instrumentstring:Selfdeclare1	-0.93183	0.22910	-4.067	4.91e-05	***
HarmonyI-V-IV:Selfdeclare1	0.10003	0.26583	0.376	0.706747	
HarmonyI-V-VI:Selfdeclare1	1.46089	0.26583	5.496	4.30e-08	***
HarmonyIV-I-V:Selfdeclare1	0.34274	0.26539	1.291	0.196674	
Selfdeclare0:LogOMSI	0.07247	0.16126	0.449	0.653177	
Selfdeclare1:LogOMSI	7.15671	2.20826	3.241	0.001208	**
Selfdeclare1:X16.minus.17-2	NA	NA	NA	NA	
Selfdeclare1:X16.minus.17-1	NA	NA	NA	NA	
Selfdeclare1:X16.minus.17-0.5	NA	NA	NA	NA	
Selfdeclare1:X16.minus.170	10.90286	2.91950	3.735	0.000192	***
Selfdeclare1:X16.minus.171	NA	NA	NA	NA	
Selfdeclare1:X16.minus.172	NA	NA	NA	NA	
Selfdeclare1:X16.minus.173	5.16433	2.19868	2.349	0.018913	*
Selfdeclare1:X16.minus.174	NA	NA	NA	NA	
Selfdeclare1:X16.minus.175	NA	NA	NA	NA	
Selfdeclare1:X16.minus.176	NA	NA	NA	NA	
Selfdeclare1:X16.minus.177	NA	NA	NA	NA	
Selfdeclare1:X16.minus.179	NA	NA	NA	NA	
Selfdeclare1:ConsInstr0.67	NA	NA	NA	NA	
Selfdeclare1:ConsInstr1	NA	NA	NA	NA	
Selfdeclare1:ConsInstr1.33	NA	NA	NA	NA	
Selfdeclare1:ConsInstr1.67	-3.35513	3.02245	-1.110	0.267079	
Selfdeclare1:ConsInstr2	NA	NA	NA	NA	
Selfdeclare1:ConsInstr2.33	NA	NA	NA	NA	
Selfdeclare1:ConsInstr2.67	NA	NA	NA	NA	
Selfdeclare1:ConsInstr3	-4.56363	1.99633	-2.286	0.022340	*
Selfdeclare1:ConsInstr3.33	NA	NA	NA	NA	
Selfdeclare1:ConsInstr3.67	NA	NA	NA	NA	
Selfdeclare1:ConsInstr4	NA	NA	NA	NA	
Selfdeclare1:ConsInstr4.33	-4.07202	2.33773	-1.742	0.081659	.
Selfdeclare1:ConsInstr5	-0.60627	1.43609	-0.422	0.672942	
Selfdeclare1:PachListen1	NA	NA	NA	NA	
Selfdeclare1:PachListen2	NA	NA	NA	NA	
Selfdeclare1:PachListen3	NA	NA	NA	NA	
Selfdeclare1:PachListen4	NA	NA	NA	NA	
Selfdeclare1:PachListen5	NA	NA	NA	NA	

Figure 15: Summary of selected interactions between Selfdeclare and fixed effects
Left: Dichotomization with Threshold of 2 // Right: Dichotomization with Threshold of 3