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

Tammy Huang*
tiamanh@stat.cmu.edu

December 8, 2019

Abstract

We address the question of how various musical features affect listeners' identification of music as classical or popular. We examined data on the classical and popular ratings of 70 listeners from the undergraduate program at University of Pittsburgh, collected by Jimenez et al. (2012). Particularly, instrument, harmonic motion, and voice leading are the three main experimental factors, and their effects on ratings are the main focus of this study. Using histograms, boxplots, and regression analysis, the effects of various musical features on listeners' identification of music were explored. We find out that instrument exerts the greatest influence on both classical and popular ratings among all experimental factors. While harmonic motion I-V-VI exerts the greatest influence on classical ratings among all harmonic motion, the effect of voice leading feature contrary motion on classical ratings may not significantly differ from the effects of other voice leading features. We also found out that musicians and non-musicians tend to respond differently to instrument and harmonic motion when rating classical music. Moreover, while popular ratings are mostly influenced by the three main experimental factors, classical ratings are also influenced by factors such as the respondent's familiarity with instrument, notes, composing, or specific composers and music pieces. Interestingly, we found that the effect of instrument on classical and popular ratings can be completely different for different listeners.

1 Introduction

Composers and musicologists are curious about how people's perception of music as classical or popular is associated with various musical features. Particularly, they are interested in the effect of instrument, harmonic motions, and voice leading on music identification. How do various musical features affect our identification of music as popular music or popular music?

To address this question, we explored the data on classical and popular ratings collected by composer Ivan Jimenez and his student Vincent Rossi in 2012, to explore the relationship between music identification and various musical features. Particularly, instrument, harmonic motion, voice leading are the three main experimental factors, which are the focus of our analysis.

In addition to answering the main question posed above, we will also address the following questions:

- What experimental factors (instrument, harmonic motion, voice leading) exert the strongest influence on classical and popular ratings?

*Department of Statistics & Data Science, Carnegie Mellon University

- Does instrument exert the strongest influence among the three main experimental factors?
- Does harmonic motion I-V-VI have the strongest influence on classical rating? Does the effect differ for respondents who are familiar with Rob Paravonian's Pachelbel Rant or Axis of Evil's Comedy bit?
- Does voice leading contrary motion have the strongest association with classical ratings?

- What are the differences between variables that exert the greatest influence on classical and popular ratings?
- What are the differences in the way that musicians and non-musicians identify classical music?

2 Methods

The data for this study come from the designed experiment conducted by Ivan Jimenez and Vincent Rossi in 2012. Ivan Jimenez, a composer and musicologist, conducted a designed experiment to measure the influence of instrument, harmonic motion, and voice leading on the identification of music as classical or popular. 70 participants were recruited from the undergraduate program at University of Pittsburgh for this experiment, with each participant presented with 36 music excerpts (or stimuli) from various instruments, harmonic motion, and voice leading. Instrument, harmonic motion, and voice leading are the three main experimental factors of interest. The information on the three main experimental factors, along with information on other musical features, were collected in the experiment.

The data from all participants in the designed experiment are represented in the data available to us, and the following variables were measured from each participant:

Y_1 = Classical = rating on how classical does the stimulus sound (on a scale of 1 to 10)
Y_2 = Popular = rating on how popular does the stimulus sound (on a scale of 1 to 10)
x_1 = Subject = the unique subject ID of each participant
x_2 = Harmony = categorical variable = harmonic motion, classified into 4 levels:
I-V-VI, I-IV-V, I-V-IV, and IV-I-V
x_3 = Instrument = categorical variable = instrument of the stimulus, classified into 3 levels:
electric guitar, piano, and string quartet
x_4 = Voice = categorical variable = voice leading, classified into 3 levels:
contrary motion, parallel 3rds, and parallel 5ths
x_5 = Selfdeclare = self-evaluation of the identification as a musician, on a scale of 0 to 6
x_6 = OMSI = score on a test of musical knowledge
x_7 = X16.minus.17 = auxiliary measure of the listener's ability to distinguish classical and popular music
x_8 = ConsInstr = self evaluation on how much the participant concentrate on the instrument while listening, on a scale of 0 to 5

x_9 = ConsNotes = self evaluation on how much the participant concentrate on the notes while listening, on a scale of 0 to 5

x_{10} = Instr.minus.Notes = differences between self-evaluation on concentration on instrument and on notes

x_{11} = PachListen = self-evaluation on the familiarity with Pachelbel's Canon in D, on a scale of 0 to 5

x_{12} = ClsListen = self-evaluation on the exposure to classical music, on a scale of 0 to 5

x_{13} = KnowRob = self-evaluation on the familiarity with Rob Paravonian's Pachelbel Rant, on a scale of 0 to 5

x_{14} = KnowAxis = self-evaluation on the familiarity with Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music, on a scale of 0 to 5

x_{15} = X1990s2000s = self-evaluation on the familiarity with pop and rock from the 90's and 2000's, on a scale of 0 to 5

x_{16} = X1990s2000s.minus.1960s1970s = difference between self-evaluation on the familiarity with pop and rock from the 90's and 00's, and from the 60's and 70's, on a scale of 0 to 5

x_{15} = CollegeMusic = dummy variable = whether the participant has taken music classes in college (0 or 1)

x_{16} = NoClass = number of classes ever taken by the participant

x_{17} = APTheory = dummy variable = whether the participant has taken AP Music Theory class in high school

x_{18} = Composing = dummy variable = whether the participant has ever composed music

x_{19} = PianoPlay = dummy variable = whether the participant plays piano

x_{20} = GuitarPlay = dummy variable = whether the participant plays guitar

x_{21} = X1stInstr = self-evaluation on the familiarity of the first instrument, on a scale of 0 to 5

x_{22} = X2stInstr = self-evaluation on the familiarity of the second instrument, on a scale of 0 to 5

The data are available in the file `ratings.csv` on Canvas, along with the PowerPoint presentation and the 36 stimuli recordings, provided by Jimenez et al. (2012).

Data cleaning is performed from the original dataset for further analysis. Due to errors in the process of data collection, there are some problems presented in the dataset available to us, including missing values, miscoded values, as well as abuse of categorical variables as numeric. In our data cleaning procedures, variables X1stInstr and X2ndInstr were removed from the variables of interest due to the problem of missing values, the values of variables ConsIntr, X16.minus.17, Instr.minus.Notes, and X1990s2000s.minus.1960s1970 were rounded to the nearest integer for corrections of miscoded values, and all variables other than OMSI and NoClass are converted into categorical variables. Moreover, only entries of complete cases are evaluated for further data analysis. The specific details of data cleaning are attached in appendix 4.1 (see page 21).

For our analysis we relied on data visualization from histograms and boxplots, as well as regression analysis on both simple linear models and hierarchical linear models. Variable combinations, transformations, and variable selections with stepwise method and automated method by fitLMER are employed in regression modeling. All data visualization and regression analysis are conducted using the R language and environment for statistical computing (R Core Team, 2017).

3 Results

3.1 Exploratory Data Analysis and Variable Transformation

We can roughly observe the association between all provided predictor variables and ratings through boxplots or scatterplots. For the association between provided predictor variables and classical ratings, we found out that the boxplots of most variables demonstrate some significant relationships. Particularly, variable Instrument, X16.minus.17, ConsNotes, and Composing may exert some significant impact on classical rating, as there are clearly observable differences between the means and interquartile ranges across different levels (see Figure 1).

On the other hand, the associations between most provided predictor variables and popular ratings are not so clear. The means and interquartile ranges of most predictor variables do not demonstrate significant differences. The only predictor variable that particularly demonstrates significant impact through data visualization is Instrument (see Figure 2).

Noted that for both classical and popular ratings, variable Instrument demonstrates the most obvious impact on ratings through data visualization. From the boxplot in the upper left corner of Figure 1, we see that music excerpts played in string quartet tend to receive the highest classical ratings, whereas music excerpts played in electric guitar tend to receive the lowest classical ratings. From Figure 2, we see that music excerpts played in electric guitar tend to receive the highest popular ratings, whereas music excerpts played in string quartet receive the lowest popular ratings.

we also discovered that the distribution of variable OMSI is severely right-skewed and in need of transformation. According to the suggestion of the Box-Cox method, we decided to perform logarithm transformation on OMSI. As we can see from Figure 3, the distribution of OMSI after logarithm transformation resembles a normal distribution, which is a desirable feature for predictor variables.

3.2 Effect of Instrument, Harmony, and Voice on Classical and Popular Ratings

The effects of experimental factors on classical and popular ratings are explored through regression analysis. Models on classical and popular ratings were proposed to account for the effect of instrument, harmonic feature, and voice leading. To choose the optimal model for each rating, a full model was first proposed, which includes all three main experimental factors and interactions of all orders. Variable selections were then performed by stepwise method, using both AIC and BIC as the criterion. Further analysis of the regression summary on p-value, confidence intervals, and effect sizes was conducted to decide which variables should be included in the final model.

We also considered adding random effects into the final model for both classical and popular ratings. As there are 70 participants (or subjects) in the dataset, we wonder whether there might be presence of subject bias on ratings. To account for such possibility, random effects were introduced to each model, with a random intercept to indicate the random variations on the ratings, and random slopes on the selected predictor variables to indicate the random variations of the effect from these variables on the ratings across different participants. The random effect model was then compared with the fixed effect only model through ANOVA. According to the result, the optimal

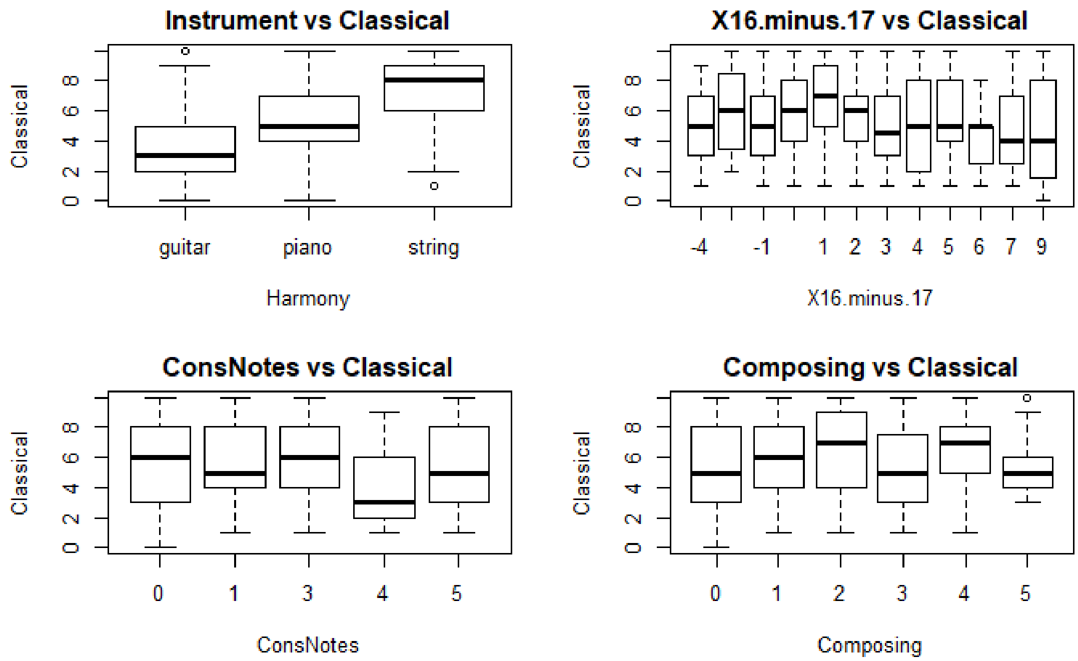

Figure 1: Boxplots of Instrument, X16.minus.17, ConsNotes, and Composing against Classical Rating

models for both ratings are the models with necessary random effects, which significantly decrease the AIC or BIC of models with only fixed effects. The specific details of constructing the optimal models for classical and popular rating from the three main experimental factors are listed in appendix 4.2 in page 21-24.

3.2.1 Model for Classical and Popular Rating from Main Experimental Factors

Our proposed model for classical ratings contains the three main experimental factors and interaction between Instrument and Harmony. According to the regression summary for the fixed effects (see Figure 4), the t-values of Instrument levels piano and string, Harmony level I-V-VI, and many levels of the interaction between Harmony and Voice exceed the (-2, 2) interval. This discovery implies that these variables are significant at a 5% level. To be more specific, compared to music excerpts played in electric guitar, harmonic motion I-V-VI, voice leading contrary motion, music excerpts played in piano, string, harmonic motion I-V-VI, and harmonic motion IV-I-V with voice leading parallel 3rds will receive higher classical ratings, music excerpts played in harmonic motion IV-I-V with voice leading parallel 5ths will receive lower classical ratings.

The proposed model for classical ratings also includes random variations on ratings across participants, and random variations on the effect of Harmony and Instrument on ratings. Classical ratings vary across participants with a variance around 1.69, and the effect of instrument piano and string varies across subjects with a variance 1.92, 3.68 respectively. The effect of Harmony also varies across participants, with a variance around 0.09, 1.75, and 0.16 for harmonic motion I-V-IV, I-V-VI, and IV-I-V. It is worth noted that the variance of random effect for Instrument is

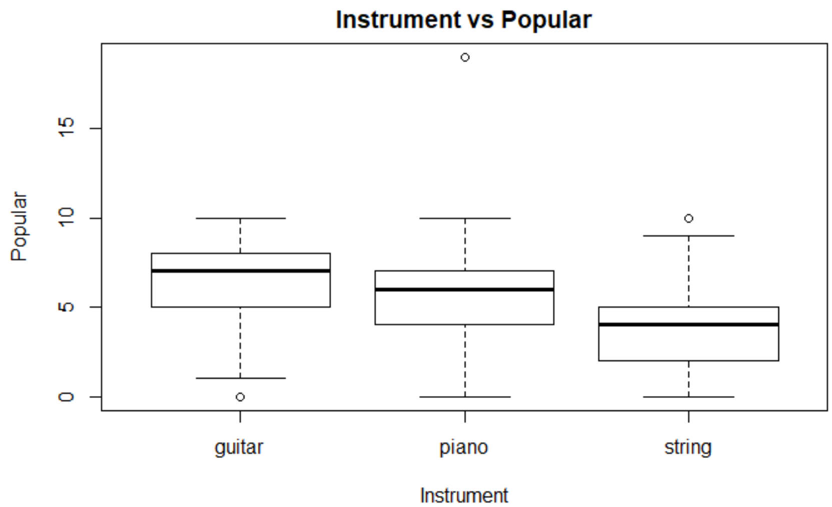

Figure 2: Boxplot of Instrument against Popular Rating

of significant magnitude, that Instrument can exert either positive or negative effect on classical ratings.

Our proposed model for popular rating contains the three main experimental factors. From the regression summary of the final model (Figure 5), the significant predictor variables at a 5% threshold are levels of Instrument (both string and piano), Harmony levels of I-V-VI and IV-I-V, and Voice level parallel 5th. Specifically, compared to music excerpts played in electric guitar, harmonic motion I-V-VI, and voice leading contrary motion, music excerpts played in piano, string, harmonic motion I-V-VI, and harmonic motion IV-I-V with voice leading parallel 3rds will receive lower popular ratings, whereas music excerpts played in voice leading parallel 5ths will receive higher classical ratings.

The proposed model for popular ratings also includes a random intercept across Subject, and a random slope of Instrument. The random effects suggest that popular ratings vary across participants with variance around 1.27, the effect of Instrument piano on classical rating varies across participants with variance around 1.72, and the effect of Instrument string quartet on classical rating varies across participants with variance 2.55. Similar to the scenario in the model for classical ratings, the variance of random effect of Instrument is of considerable magnitude, that the effect of instrument may be either positive or negative on popular rating across participants.

3.2.2 Instrument Exerts the Greatest Influence on Ratings among All Experimental Factors

We examined the influence of Instrument on classical and popular ratings based on our proposed models. It is discovered that for both classical and popular rating, Instrument exerts the most significant effect.

The optimal model for classical rating suggests that Instrument exerts the greatest influence among all main experimental factors and their interactions. From the regression output in Figure 4, we see that not only does Instrument exert statistically significant effect on classical ratings, the

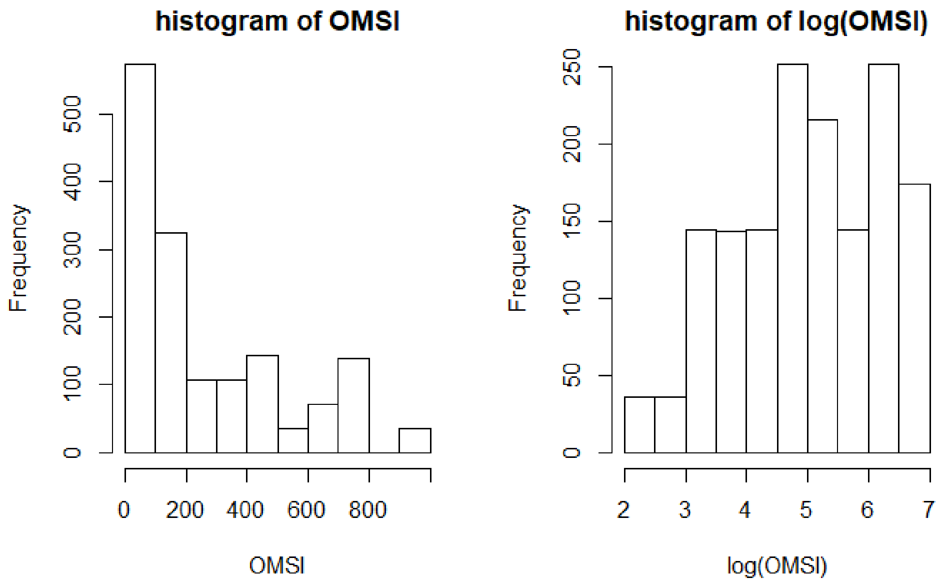

Figure 3: Histograms of OMSI and log(OMSI)

effect exerted is also of the greatest magnitude among all predictor variables. For one, levels of Instrument are significant at a 5% level. The t-values of Instrument levels piano and string both exceed the (-2, 2) interval, which means that the effects of these two levels are significantly different from the effect of the baseline case, namely, when the music is played in electric guitar.

We also found that levels of Instrument have the greatest effect size among all predictor variables for classical ratings. We separated the coefficient of Instrument level electric guitar from the intercept, and listed the coefficients of each level of Instrument in Figure 6. The effect sizes of these levels are around 3.81, 5.45 and 7.40 respectively, which implies that music played in electric guitar, piano, and string quartet is expected to receive 3.81, 5.45, and 7.40 classical rating respectively, when the music is played in harmonic motion I-V-VI and voice leading contrary motion. Also, as levels of Instrument has considerable effect size (all above 3.8) and small standard error (all below 0.3), the confidence intervals for these levels are strictly positive, which means that the effects brought by all instruments are unambiguous. To conclude, Instrument exerts the greatest effect on classical ratings, among all main experimental factors.

The optimal model for popular rating also suggests that Instrument plays the most significant role in affecting ratings. For one, levels of Instrument are significant at the 5% level. According to the regression output in Figure 5, the t-value of both Instrument levels piano and string are smaller than -2, which means their effects on popular rating significantly differ from the effect brought by the baseline instrument level of electric guitar. Also, the coefficients of Instrument have the greatest effect size among all predictor variables. Again, we separated the coefficient of all levels of Instrument and listed in Figure 7. As we can see, the effect sizes of these levels around 6.84, 5.69, and 3.82 respectively, which implies that music played in electric guitar, piano, and string quartet is expected to receive 6.84, 5.69, and 3.82 popular rating respectively, given that the music is played in harmonic motion I-V-VI and voice leading contrary motion. The confidence intervals for these levels are both strictly positive (with the considerable effect sizes and small standard errors), such

```
Random effects:
 Groups   Name             Variance Std.Dev.
 Subject  (Intercept)      1.68937  1.2998
          Instrumentpiano  1.92062  1.3859
          Instrumentstring 3.68496  1.9196
          HarmonyI-V-IV    0.08731  0.2955
          HarmonyI-V-VI    1.74687  1.3217
          HarmonyIV-I-V    0.16079  0.4010
 Residual                  2.44098  1.5624
Number of obs: 1541, groups:  Subject, 43

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                 3.8034     0.2477  15.352
Instrumentpiano             1.6532     0.2329   7.099
Instrumentstring            3.5877     0.3085  11.630
HarmonyI-V-IV               0.2127     0.2001   1.063
HarmonyI-V-VI               1.2662     0.2808   4.510
HarmonyIV-I-V              -0.3023     0.2039  -1.483
Voicepar3rd                -0.3101     0.1945  -1.594
Voicepar5th                -0.2038     0.1950  -1.045
HarmonyI-V-IV:Voicepar3rd  -0.4298     0.2754  -1.560
HarmonyI-V-VI:Voicepar3rd  -0.7074     0.2760  -2.563
HarmonyIV-I-V:Voicepar3rd   0.7514     0.2754   2.728
HarmonyI-V-IV:Voicepar5th  -0.2103     0.2760  -0.762
HarmonyI-V-VI:Voicepar5th  -0.5236     0.2761  -1.896
HarmonyIV-I-V:Voicepar5th   0.3356     0.2754   1.218
```

Figure 4: Regression Summary for the Proposed Model on Classical Ratings from Main Experimental Factors

```
Random effects:
 Groups   Name             Variance Std.Dev.
 Subject  (Intercept)      1.273    1.128
          Instrumentpiano  1.716    1.310
          Instrumentstring 2.546    1.596
 Residual                  3.037    1.743
Number of obs: 1541, groups:  Subject, 43

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       6.84206    0.21294  32.131
Instrumentpiano  -1.14711    0.22761  -5.040
Instrumentstring -3.02320    0.26644 -11.347
HarmonyI-V-IV     0.02312    0.12554   0.184
HarmonyI-V-VI    -0.25238    0.12563  -2.009
HarmonyIV-I-V    -0.24880    0.12545  -1.983
Voicepar3rd       0.19603    0.10877   1.802
Voicepar5th       0.23207    0.10877   2.134
```

Figure 5: Regression Summary for the Proposed Model on Popular Ratings from Main Experimental Factors

```
                          Estimate Std. Error
Instrumentguitar            3.8034     0.2478
Instrumentpiano             5.4566     0.2904
Instrumentstring            7.3911     0.2901
```

Figure 6: Effect Size of Instrument in the Classical Model

	Estimate	Std. Error
Instrumentguitar	6.8366	0.2301
Instrumentpiano	5.6914	0.2302
Instrumentstring	3.8177	0.2300

Figure 7: Effect Size of Instrument in the Popular Model

	Estimate	Std. Error
HarmonyI-IV-V	3.8034	0.2478
HarmonyI-V-IV	4.0161	0.2784
HarmonyI-V-VI	5.0696	0.3456
HarmonyIV-I-V	3.5011	0.2603

Figure 8: Effect Size of Harmony in the Classical Model

that the effects are unambiguous. To conclude, Instrument exerts the greatest effect on popular ratings, among all main experimental factors.

3.2.3 Harmonic Motion I-V-VI Exerts the Greatest Influence on Classical Ratings Among All Levels of Harmonic Motion

Through our proposed models, we found out that harmonic motion I-V-VI exerts the most significant effect on classical ratings, among all levels of harmonic motion.

For classical ratings, harmonic motion exerts significant and considerable impact. Our proposed model of classical ratings sets harmonic motion I-IV-V as the baseline, with the significance of harmonic motion I-V-IV, I-V-VI, and IV-I-V tested in the regression output. According to the result (see Figure 4), harmonic motion I-V-VI is the only level in Harmony with t value significantly greater than 2 (around 4.51), which affirms the statistical significance of this level. The interaction between harmonic motion I-V-VI and voice leading parallel 3rds is also significant at a 5% level. This means that compared to music excerpts played in electric guitar, with harmonic motion I-IV-V and voice leading contrary motion, music excerpts played in harmonic motion I-V-VI tend to receive higher classical ratings.

The significant impact of harmonic motion I-V-VI can also be observed from the effect sizes of levels of Harmony. Factoring out all Harmony levels from intercept (see Figure 8), the coefficients for Harmony levels I-IV-V, I-V-IV, I-V-VI, and IV-I-V are observed to be around 3.80, 4.02, 5.07, and 3.50 respectively, indicating that when the stimulus is played in electric guitar with voice leading contrary motion, the expected classical ratings for each harmonic motion is 3.80, 4.02, 5.07, and 3.50. Among all harmonic levels, harmonic motion I-V-VI has the greatest effect size (5.07), and therefore exerts the greatest influence on classical ratings.

We also explored whether the effect of harmonic motion I-V-VI on classical rating differ if the respondents are familiar with Rob Paravonian's Pachelbel Rant or Axis of Evil's Comedy bit. To investigate the significance of this association, we designed a model that includes all variables in our proposed model for classical ratings, variables KnowRob and KnowAixs, and two-way interactions between all levels of Harmony and KnowRob and KnowAxis. The regression summary of this model was investigated for significant interactions. We found out that for harmonic motion I-V-VI, the only significant interaction at 5% level is the interaction with KnowRob5, with a coefficient around 1.67 (see Figure 9). This implies that when the music stimulus is played with harmonic motion

	Estimate	Std. Error	t value
HarmonyI-IV-V:KnowRob1	-0.89287	0.62672	-1.425
HarmonyI-V-IV:KnowRob1	-0.86649	0.67368	-1.286
HarmonyI-V-VI:KnowRob1	0.08204	0.76645	0.107
HarmonyIV-I-V:KnowRob1	-0.62281	0.66482	-0.937
HarmonyI-IV-V:KnowRob5	-0.22506	0.55981	-0.402
HarmonyI-V-IV:KnowRob5	-0.61003	0.60213	-1.013
HarmonyI-V-VI:KnowRob5	1.66630	0.68476	2.433
HarmonyIV-I-V:KnowRob5	-0.13147	0.59367	-0.221
HarmonyI-IV-V:KnowAxis1	1.75751	1.27346	1.380
HarmonyI-V-IV:KnowAxis1	3.23396	1.36906	2.362
HarmonyI-V-VI:KnowAxis1	0.19924	1.55743	0.128
HarmonyIV-I-V:KnowAxis1	2.26953	1.35080	1.680
HarmonyI-IV-V:KnowAxis5	0.95521	0.44789	2.133
HarmonyI-V-IV:KnowAxis5	0.78660	0.48155	1.633
HarmonyI-V-VI:KnowAxis5	-0.17474	0.54779	-0.319
HarmonyIV-I-V:KnowAxis5	0.60376	0.47507	1.271

Figure 9: Interactions between Harmony and KnowRob, KnowAxis

I-V-VI, a 1.67 points increase in classical rating is expected if the respondent is very familiar with Pachelbel Rant. Thus, we conclude that the effect of harmonic motion I-V-VI differs for respondents familiar with Pachelbel Rant.

3.2.4 The Effect of Voice Leading Contrary Motion on Classical Ratings Compared to Other Levels of Voice Leading

We found out that voice leading may not exert statistically significant effect on classical ratings through our proposed model. According to the regression summary of the model on classical ratings in Figure 4, our proposed model sets voice leading contrary motion as the baseline, with neither of the remaining voice leading levels (parallel 3rds and parallel 5ths) demonstrating t value exceeds the -2 to 2 threshold. The only significant variables related to voice leading are the interactions between voice leading parallel 3rds and harmonic motion I-V-IV, and between voice leading parallel 3rds and harmonic motion IV-I-V. From this, we conclude that the effect of different levels of Voice on classical ratings does not significantly differ between each other at a 5% significance level.

We can also examine the effect of Voice through the coefficients of its levels. Factoring out all levels of voice leading from the intercept, the coefficients for Harmony levels contrary motion, parallel 3rds, and parallel 5ths are observed to be around 3.80, 3.50, 3.60 respectively (see Figure 10). This indicates that when the music stimulus is played in electric guitar with harmonic motion I-IV-V, music played in voice leading contrary motion will receive higher classical ratings than music played in other voice leading. However, we can also see that the coefficients of all three levels of voice leading are fairly close to each other, and therefore their effects on classical ratings may not be significantly different.

Overall, voice leading contrary motion may exert the greatest influence on classical ratings among all levels of Voice, but the effect on classical ratings brought by contrary motion may not significantly differ from the effect brought by other voice leading.

	Estimate	Std. Error
Voicecontrary	3.8034	0.2478
Voicepar3rd	3.4933	0.2478
Voicepar5th	3.5996	0.2480

Figure 10: Effect Size of Voice in the Classical Model

3.3 Differences between Significant Predictor Variables for Classical and Popular Ratings

In addition to the effect of three main experimental factors, we also considered the effects of other predictor variables provided in the dataset by Jimenez et al. (2012). With our proposed models for classical and popular ratings from the three main experimental factors, other predictor variables are added to the proposed models, with variable selection by stepwise method with both AIC and BIC as the criterion.

Random effects were also evaluated for both classical and popular ratings. Random variation on the ratings across participants, and random variations on the effect of significant experimental factors are also added to the model for the possibility of further improvement. A further examination of model summary and residual plots is then conducted to verify the validity of models.

Lastly, the automated method is performed to verify model selections. The chosen automated method, fitLMER, is performed on the full model with all possible predictor variables, to verify the result of our model selection. The specific information of the final models for classical and popular ratings from all predictor variables is listed below, with detailed descriptions on the construction of final models listed in appendix 4.3 in page 24-28.

3.3.1 Variable Exclusion and Transformation

Prior to variable selections, some variables were either excluded from considerations, or transformed before putting into both models.

For variable exclusions, some variables are by definition correlated with each other, and therefore excluded from considerations. Variable Instr.minus.Notes denotes the differences between variable ConsInstr and ConsNotes. If all three variables are included in the models, colinearity will be induced to the model. Variable X1990s2000s.minus.1960s1970s is excluded from variables of interest for the same reason: the variable will be correlated with variable X1990s2000s, and it is very difficult to interpret without data on the preferences for music during 1960s and 1970s.

Particularly, variables AP Theory and CollegeMusic are excluded from variables of interest for classical rating. As we can see from the boxplots of AP Theory and CollegeMusic against classical ratings, the effects of these two variables on classical ratings are very vague, with similar means and largely overlapping interquartile ranges (see Figure 11). From above, we decided to exclude the AP Theory and CollegeMusic from the model on classical rating.

It is also found out that the distribution of variable OMSI is severely right-skewed, and the variable is thus log-transformed before modeling.

In sum, the predictor variables other than experimental factors for both models are Selfdeclare, log(OMSI), X16.minus.17, ConsInstr, ConsNotes, PachListen, ClsListen, knowRob, KnowAxis, X1990s2000s, NoClass, AP Theory, CollegeMusic, Composing, PianoPlay, and GuitarPlay.

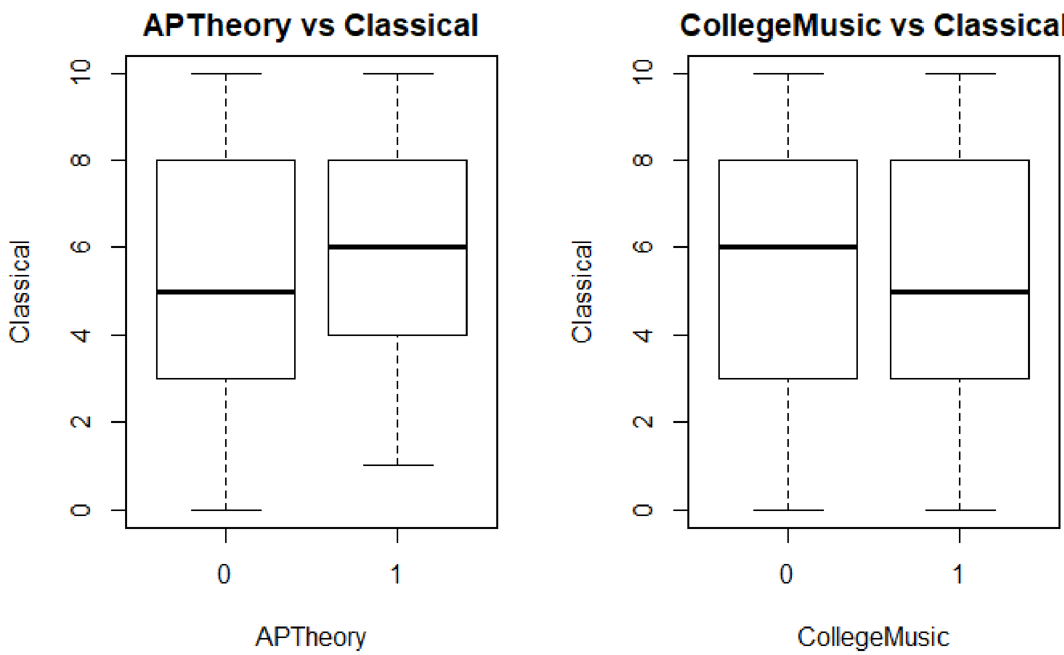

Figure 11: Boxplots of APTheory and CollegeMusic against Classical Rating

3.3.2 Variable Selection

Fixed effects of both models were selected through stepwise method from all possible predictor variables. For model on classical ratings, both AIC and BIC methods recommend to include variables X16.minus.17, ConsInstr, ConsNotes, PachListen, ClsListen, X1990s2000s, and Composing, with AIC method suggesting the inclusion of two additional variables, Selfdeclare and KnowRob. After analyzing regression summary, these two variables are included in the fixed effects. For model on popular ratings, both AIC and BIC methods recommend the inclusion of three main experimental effects.

Random effects on main experimental factors were introduced for both models, with the significance tested by ANOVA. For model on classical ratings, the inclusion of random effects on Instrument and Harmony improves the model with only fixed effects. For model on popular ratings, the inclusion of random effect on Instrument improves the model with only fixed effects. A further examination of the random effects model affirms the validity of the above variable selections.

The models proposed above was further affirmed by the fitLMER automatic method. The function fitLMER first conducts backward selection on the fixed effects, and then forward selection on the random effects. Both methods give the same result in variable selection, which confirms the validity of variables included in the final models. The specific details on the model construction for classical and popular ratings, along with the regression summaries of the final models, can be found in appendix 4.3 on page 24-28.

3.3.3 Summary on the Final Model on Classical and Popular Rating

All variables of the model on classical ratings are listed below.

$$
\begin{aligned}
Y &= \text{Classical} = \text{rating on how classical does the stimulus sound (on a scale of 1 to 10)} \\
x_1 &= \text{Subject} = \text{the unique subject ID of each participant} \\
x_2 &= \text{Harmony} = \text{categorical variable} = \text{harmonic motion, classified into 4 levels:} \\
&\quad \text{I-V-VI, I-IV-V, I-V-IV, and IV-I-V} \\
x_3 &= \text{Instrument} = \text{categorical variable} = \text{instrument of the stimulus, classified into 3 levels:} \\
&\quad \text{electric guitar, piano, and string quartet} \\
x_4 &= \text{Harmony:Instrument} = \text{interaction between Instrument and Harmony} \\
x_5 &= \text{Selfdeclare} = \text{self-evaluation of the identification as a musician, on a scale of 0 to 5} \\
x_6 &= \text{X16.minus.17} = \text{auxiliary measure of listener's ability to distinguish classical and} \\
&\text{popular music} \\
x_7 &= \text{ConsInstr} = \text{self evaluation on how much the participant concentrate on the instrument} \\
&\text{while listening, on a scale of 0 to 5} \\
x_8 &= \text{ConsNotes} = \text{self evaluation on how much the participant concentrate on the notes} \\
&\text{while listening, on a scale of 0 to 5} \\
x_9 &= \text{PachListen} = \text{self-evaluation on the familiarity with Pachelbel's Canon in D, on a scale} \\
&\text{of 0 to 5} \\
x_{10} &= \text{ClsListen} = \text{self-evaluation on the exposure to classical music, on a scale of 0 to 5} \\
x_{11} &= \text{KnowRob} = \text{self-evaluation on the familiarity with Rob Paravonian's Pachelbel Rant,} \\
&\text{on a scale of 0 to 5} \\
x_{12} &= \text{X1990s2000s} = \text{self-evaluation on the familiarity with pop and rock from the 90's and} \\
&\text{2000's, on a scale of 0 to 5} \\
x_{13} &= \text{Composing} = \text{dummy variable} = \text{whether the participant has ever composed music}
\end{aligned}
$$

All variables of the model on popular ratings are listed below.

$$
\begin{aligned}
Y &= \text{Popular} = \text{rating on how classical does the stimulus sound (on a scale of 1 to 10)} \\
x_1 &= \text{Subject} = \text{the unique subject ID of each participant} \\
x_2 &= \text{Harmony} = \text{categorical variable} = \text{harmonic motion, classified into 4 levels:} \\
&\quad \text{I-V-VI, I-IV-V, I-V-IV, and IV-I-V} \\
x_3 &= \text{Instrument} = \text{categorical variable} = \text{instrument of the stimulus, classified into 3 levels:} \\
&\quad \text{electric guitar, piano, and string quartet}
\end{aligned}
$$

Overall, the final model for classical ratings includes significantly more predictor variables than the final model for popular ratings. While the final model of the popular ratings contains only the three main predictor variables, the final model for classical ratings contains 10 more variables:

Selfdeclare, X16.minus.17, ConsInstr, ConsNotes, PachListen, ClsListen, knowRob, X1990s2000s, Composing, and the interaction between Harmony and Instrument.

Patterns are observed in the extra variables included in the model for classical ratings. Specifically, people more inclined to declare themselves as musicians are expected to give lower classical rating compared to non-musicians; people concentrate on instrument tend to give higher classical rating than people who do not, and people concentrate on notes tend to give lower classical rating than people who do not; people who are more familiar with Canon tend to give higher classical rating, whereas people who reported to listen to more classical music tend to give lower classical ratings. The regression summary of the model is attached in appendix 4.3 in page 26. In sum, there are significantly more factors that influence classical ratings than popular ratings, and these variables seem to be music knowledge related.

The effects of three main experimental factors differ for classical and popular ratings. As we can see from Figure 12, most effects of the main experimental factors on classical and popular ratings have opposite signs, which means these variables exert opposite effect on classical and popular rating. For classical ratings, music excerpts played in piano and string quartet are expected to receive high classical ratings compared to music played in electric guitar; music excerpts played in harmonic feature I-V-IV and I-V-VI are expected to receive higher rating, and music excerpts played in feature IV-I-V are expected to receive lower ratings compared to those played in I-IV-V. Music excerpts played with voice leading parallel 3rds and parallel 5ths are expected to receive lower classical ratings compared to those played with contrary motions. For popular ratings, music excerpts played in piano and string quartet are expected to receive lower popular ratings compared to music played in electric guitar. Compared to music played in harmonic feature I-VI-V and leading voice feature contrary motions, music played in harmonic feature I-V-VI and IV-I-V will receive lower popular ratings, and music played in harmonic feature I-V-IV, voice features parallel 3rds and parallel 5ths will receive higher popular ratings. Overall, the effects of Instrument, Harmony, and Voice are almost opposite for classical and popular ratings.

Random effects also differ for classical and popular ratings. For one, in addition to the random intercept and random effect of Instrument, the classical model also includes the random effect of Harmony. This means that the effect of instrument significantly varies across subjects for both popular and classical ratings, the effect of harmonic motion varies significantly across subjects for only classical ratings.

For another, random effect of Instrument differs for classical and popular ratings. As we can see from Figure 13, the variation on the effect of Instrument has a greater variance for classical rating than for popular rating. This implies that there is more variation in the effect of instrument across participants for classical ratings than for popular ratings. Moreover, the variation on popular ratings across participants has a greater variance than for classical ratings, which implies that there is a stronger presence of subject variation for popular ratings than for classical ratings.

To conclude, while the effect of Instrument varies more for classical ratings than for popular ratings, the effect of Harmony only significantly varies across subjects for classical ratings.

3.4 Differences in the Responses to Classical Music Identification between Musicians and Non-musicians

All respondents were first dichotomized into two groups, musician and non-musicians, for further analysis of the differences of responses. The dichotomization is based on responses to variable Selfdeclare. Specifically, if a participant scores himself or herself 3 or higher out of 5 in Selfdeclare,

Classical	Estimate	Std. Error
(Intercept)	1.71445	1.83952
Instrumentpiano	1.65278	0.23594
Instrumentstring	3.58760	0.31229
HarmonyI-V-IV	0.21211	0.20060
HarmonyI-V-VI	1.26653	0.28435
HarmonyIV-I-V	-0.30233	0.20387
Voicepar3rd	-0.31008	0.19499
Voicepar5th	-0.20403	0.19543

Popular	Estimate	Std. Error
(Intercept)	6.84209	0.21499
Instrumentpiano	-1.14711	0.23031
Instrumentstring	-3.02322	0.26960
HarmonyI-V-IV	0.02313	0.12576
HarmonyI-V-VI	-0.25235	0.12585
HarmonyIV-I-V	-0.24880	0.12567
Voicepar3rd	0.19601	0.10896
Voicepar5th	0.23205	0.10896

Figure 12: Regression Summary of Main Experimental Effects from the Classical and Popular Model

Random effects for Classical Model:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	0.67664	0.8226
	Instrumentpiano	1.98027	1.4072
	Instrumentstring	3.78442	1.9454

Random effects for Popular Model:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1.308	1.144
	Instrumentpiano	1.767	1.329
	Instrumentstring	2.617	1.618
Residual		3.048	1.746

Figure 13: Regression Summary of Random Effects on Instrument from the Classical and Popular Model

Interaction	Estimate	Std. Error	t value
HarmonyI-V-VI:Musician	1.24310	0.39636	3.136
Instrumentpiano:Musician	-0.65110	0.22363	-2.911
Instrumentstring:Musician	-1.18372	0.22271	-5.315

Figure 14: Regression Summary Significant Interactions with Musician in Classical Model

then we consider that participant as fairly confident in claiming the self-identification as a musician, or non-musicians if otherwise. In this way, the numbers of observations from the musician group and the non-musician group are roughly equal (714 and 827 respectively), which facilitates our further analysis.

Interactions with Musician were introduced in our final model in section 3.3 to investigate the differences in responses between musicians and non-musicians. A full model with all variables in the final model, Musician, and the two-way interactions between Musician and each variable was proposed. The significance of interactions is then evaluated by the reported t values in the regression summary. All significant interactions from the regression summary are listed in Figure 14. According to the output, only interactions of Instrument and Harmony with Musicians are significant at the 5% level. As we can see from the table in Figure 14, the only interactions with t-value exceeds the range of -2 to 2 are the interactions between Music and variables Instrument and Harmony.

From above, We conclude that the effect of Instrument and Harmony on classical ratings significantly differ between musicians and non-musicians. Specifically, musicians tend to give lower classical rating for music played in both piano and string quartet, compared to non-musician, they also tend to give higher classical for music with harmonic feature I-V-VI than non-musicians.

We also wonder whether the classical ratings vary across musicians and non-musicians. To test the significance of this variation, a random intercept was added to the model to reflect the variations of the expected classical ratings between musicians and non-musicians. However, we discovered that the inclusion of the random intercept is not necessarily, as the inclusion of the random effect increases the BIC statistics of the previous model by 5 (see Figure 15). Thus, we conclude that while musicians and non-musicians have different responses to Instrument and Harmony when rating classical music, the mean ratings given by the two groups may not significantly differ.

It is also worth noted that the result of regression analysis is sensitive to dichotomization. If we choose another threshold of Selfdeclare response to dichotomize respondents, the significance and effect sizes of variables will be different. However, we believed the dichotomization listed above yields a representative result. The result of the above dichotomization ensures similar sample sizes between musicians and non-musicians, and the result from the above dichotomization is similar to the result if we instead choose Selfdeclare response 2 as the threshold. Detailed descriptions on the sensitivity of dichotomization is listed in appendix 4.4, in page 28-29.

4 Discussion

The data on the ratings of 70 participants from University of Pittsburgh provided by Jimenez et al. (2012) displays a wide range of responses to various musical features, with a particular emphasis on the three main experimental factors: Instrument, Harmony, and Voice.

In our analysis, we found out that instrument exerts the greatest influence on both classical

```
Models:
lmer_test_music: Classical ~ (Harmony + Instrument + Voice + X16.minus.17
    + ConsInstr +  ConsNotes + PachListen + ClsListen + KnowRob + X1990s2000s
    + Composing) * Musician + Harmony:Voice + (Harmony + Voice | Subject)
lmer_test_music_r: Classical ~ (Harmony + Instrument + Voice + X16.minus.17 +
    ConsInstr + ConsNotes + PachListen + ClsListen + KnowRob + X1990s2000s +
     Composing) * Musician + Harmony:Voice + (Harmony + Voice | Subject) +
     (1 | Musician)
                  Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer_test_music   85 6396.6 6850.5 -3113.3   6226.6
lmer_test_music_r 86 6396.3 6855.5 -3112.1   6224.3 2.3048      1      0.129
```

Figure 15: ANOVA Output for Testing the Inclusion of Random Effect on the Classical Rating Across Musician

and popular ratings. Among all harmonic motions, I-V-VI exerts the greatest influence on classical ratings, whereas the effect of voice leading contrary motion on classical ratings may not significantly differ from other voice leading features. Also, the effects of instrument and harmonic motion on classical ratings tend to differ across musicians and non-musicians. Moreover, while popular ratings are mostly influenced by the three main experimental factors, classical ratings are also influenced by factors such as the respondent's familiarity with instrument, notes, composing, or specific composers and music pieces.

There are some interesting discoveries in our analysis. For one, we found out that instrument exerts the greatest influence for both classical and popular ratings. This may imply that among all musical features included in the design experiment, instrument is the most distinguishable feature in music identification for most participants.

We also found out that the effect of instrument varies greatly across participants for both classical and popular ratings. In fact, given the considerable size of variance, the effect of instrument on ratings can be completely different for different participants. The effect of the same instrument may be associated with an increase in classical rating for some participants, while a decrease for others, as the coefficient for variable Instrument can be either positive or negative in the model for classical ratings with random effects.

Furthermore, we found out that the final model on classical ratings contains many more predictor variables on music knowledge, compared to the final model on popular ratings. According to the classical rating model, classical ratings will be significantly affected by whether the respondents are familiar with instruments, notes, composing, or specific composers and music pieces. This discovery may imply that the ability to distinguish classical music requires more specific music knowledge, whereas such knowledge is not necessary for the identification of popular music.

Our study was also limited by the rating data reported by Jimenez et al. (2012). The analysis is based on data collected from 70 undergraduate students from University of Pittsburgh, and therefore may only be representative of University students, if not only undergraduate students from University of Pittsburgh. People with different backgrounds would have different extent of exposure to classical and popular music, and therefore would likely to present different responses.

Another limitation of this study comes from the handling of missing data. Only complete cases from variables of interest are included in this study. However, missing data may induce bias in our estimates, unless the missingness is completely at random and unrelated to any of the predictor variables included in our analysis. Further analysis of the reason behind missingness can be conducted, with some corresponding imputation methods to handle missing data.

In summary, we found out that instrument exerts the greatest influence on ratings among all predictor variables for both classical and popular ratings. While harmonic motion I-V-VI is the level of Harmony that exerts the greatest influence on classical ratings, the effect of voice leading contrary motion does not significantly differ from the effect of other voice leading features. Musician and non-musicians tend to respond differently to instrument and harmony when rating classical music. Other than the three main experimental factors, there are significantly more variables that exert influence on classical ratings compared to popular ratings. The significance of instrument in affecting both ratings may imply that instrument is the most distinguishable feature in music identification for most participants, and the considerable number of predictors for classical ratings may imply that the identification for classical music requires more specific music knowledge than for popular music. While our model provides sufficient explanatory power for musical response for undergraduate students at University of Pittsburgh, it may not be readily applied to the broader population, as the experiment is specifically designed for a specific population. Further study should be conducted to investigate the association between various music features and musical identification for a model that applies to the general public.

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/.

Jimenez, I., Rossi,V. (2012), *The Influence of Timbre, Harmony, and Voice Leading on Listeners' Distinction between Popular and Classical Music*. Pennsylvania: University of Pittsburgh. Retrieved December 6, 2019 (https://canvas.cmu.edu/courses/11853/files/folder/hw10).

Appendix

4.1 Data Cleaning

We first found out that some of the variables have considerable amount of missing entries. Particularly variable X1stInstr and X2ndInstr, each with 1512 and 2196 NAs respectively. Considering the fact there are more than half of the entries are NAs for these two variables, I decided to exclude the two from the variables of interest.

Moreover, some variables, including variables ConsIntr, X16.minus.17, Instr.minus.Notes, as well as X1990s2000s.minus.1960s1970s, take more discrete values than expected. For instance, in the case of variable Constr, even though this variable indicates a scale of 0-5 on whether the participant pays attention to notes while listening to music, the variable contains values like 4.33, 2.33, and 3.67. Out of suspicion for errors in data collection, we rounded the values of these variables to integers for further data analysis.

Furthermore, all variables except NoClass and OMSI are converted into factor variable, as these variables take on discrete values, with no observable monotone patterns demonstrated in their levels.

There are many missing data in most of the variables. we decided to consider only the complete cases, as imputation tend to induce errors. After deleting all observations with NAs, we still have 1541 entries out of the 2310 entries from the original dataset.

4.2 Model Construction for Classical and Popular Ratings From Main Experimental Effects

4.2.1 Classical Rating

I first proposed a full model with all the individual variables and interactions between these variables (with both two-way and three-way interactions). Then I perform variable selections with both AIC and BIC as the criterion. It turns out that while AIC prefers the model with all three main effects with the interaction term between Harmony and Voice, BIC prefers the most parsimonious model, with just the two main effects as the predictors.

Looking at the regression summary (Figure 16), I found out that some levels of the interaction term have p-values smaller than 0.05, which means that they are statistically significant at 5% level, and should be included in the model.

Then we consider whether the model can be improved by adding random effects. First a random intercept was added to the model to reflect the variations on classical ratings across participants. This random intercept model was compared with the previous AIC model through ANOVA, and according to the result shown in Figure 17, the random intercept model has a very small p value, significantly lower AIC and BIC statistics. This means that we are confident at a 5% level to reject the AIC level and accept the random intercept model for classical ratings.

Next we consider the possibility of random slopes for predictor variables in the model. The random variation across participants was considered for each variable individually, and the significance of variation is evaluated through ANOVA. We found out that the random effect of Instrument and Harmony improves the random intercept model by significantly decreasing the AIC and BIC statistics. We then add both random effect of Instrument and Harmony into the model, and compare this model with all the proposed model through ANOVA. According to the output (see Figure 18), the model with both random effects has the lowest AIC and BIC statistics, and therefore should be considered the optimal model.

```
Coefficients:               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   3.8032     0.2156  17.642  < 2e-16
Instrumentpiano               1.6554     0.1416  11.692  < 2e-16
Instrumentstring              3.5861     0.1412  25.404  < 2e-16
HarmonyI-V-IV                 0.2223     0.2827   0.786   0.4318
HarmonyI-V-VI                 1.2619     0.2833   4.454 9.03e-06
HarmonyIV-I-V                -0.3023     0.2822  -1.071   0.2842
Voicepar3rd                  -0.3101     0.2822  -1.099   0.2720
Voicepar5th                  -0.1917     0.2827  -0.678   0.4978
HarmonyI-V-IV:Voicepar3rd    -0.4394     0.3995  -1.100   0.2715
HarmonyI-V-VI:Voicepar3rd    -0.7139     0.4002  -1.784   0.0747
HarmonyIV-I-V:Voicepar3rd     0.7566     0.3995   1.894   0.0584
HarmonyI-V-IV:Voicepar5th    -0.2223     0.4002  -0.556   0.5786
HarmonyI-V-VI:Voicepar5th    -0.5314     0.4002  -1.328   0.1845
HarmonyIV-I-V:Voicepar5th     0.3235     0.3995   0.810   0.4181
```

Figure 16: Summary of AIC Model for Classical Rating from Main Experimental Factors

```
Models:
model_AIC: Classical ~ Instrument + Harmony + Voice + Harmony:Voice
lmer.1: Classical ~ Voice + Instrument + Harmony + Harmony:Voice + (1 | Subject)
          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model_AIC 15 6910.6 6990.7 -3440.3   6880.6
lmer.1    16 6523.8 6609.2 -3245.9   6491.8 388.83      1  < 2.2e-16
```

Figure 17: ANOVA Output for AIC Model and Random Intercept Model for Classical Rating

```
Models:
lmer.1: Classical ~ Voice + Instrument + Harmony + Harmony:Voice + (1 |
     Subject)
lmer.2: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1 +
     Instrument | Subject)
lmer.4: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1 +
     Voice | Subject)
lmer.3: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1 +
     Harmony | Subject)
lmer.5: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1 +
     Instrument + Harmony | Subject)
       Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
lmer.1 16 6523.8 6609.2 -3245.9   6491.8
lmer.2 21 6271.9 6384.1 -3115.0   6229.9 261.865      5  < 2.2e-16
lmer.4 21 6533.4 6645.6 -3245.7   6491.4   0.000      0          1
lmer.3 25 6487.0 6620.5 -3218.5   6437.0  54.428      4  4.281e-11
lmer.5 36 6190.1 6382.3 -3059.0   6118.1 318.956     11  < 2.2e-16
```

Figure 18: ANOVA Output for Random Slope Models and Random Intercept Model for Classical Rating

Lastly, regression summary of the final model on classical ratings was evaluated. The model summary (see Figure 4) shows that all predictor variables (except Voice, an experimental factor that has to be included in the model) are significant at 5% level, with t-values exceed the (-2, 2) interval.

In sum, the model suggests that music excerpts played in piano and string quartet are expected to receive higher classical ratings compared to those played in electrical guitar. Whereas music played with leading harmonic features I-V-IV and I-V-VI are expected to receive higher classical ratings compared to those played with I-VI-V, music played with leading harmonic feature IV-I-V are expected to receive lower ratings. Also, music played with leading voice features parallel 3rds and parallel 5ths are expected to receive lower classical ratings compared to those with leading voice feature as contrary motion. Also, the mean classical rating varies across subject, so are the effects of Instrument and Harmony on classical ratings.

4.2.2 Popular Rating

Model on popular rating from main experimental factors is developed in the same method. From a full model with all main effect with interactions of all order, stepwise method was employed for variable selection. According to the result, both AIC and BIC criteria recommend the model with Instrument as the sole predictor. As the three main experimental variables are design variables that have to be included in the model, the model I proposed for popular rating with main experimental factors is the model that includes all experimental factors but no interaction.

Next, random effects were introduced to the initial model to see whether any there are any improvement. Through ANOVA (see Figure 19), it is discovered that the inclusion of both random intercept on the variation of rating across participants, and the random slope of Instrument on rating across participants, significantly decreases the AIC and BIC statistics. From this, we conclude that

```
Models:
lm_pop_full: Popular ~ Instrument + Harmony + Voice
lmer_pop: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
lmer_pop_2: Popular ~ Instrument + Harmony + Voice + (Instrument | Subject)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lm_pop_full  9 6910.5 6958.6 -3446.2   6892.5
lmer_pop    10 6513.0 6566.4 -3246.5   6493.0 399.54      1  < 2.2e-16
lmer_pop_2  15 6358.5 6438.6 -3164.2   6328.5 164.50      5  < 2.2e-16
```

Figure 19: ANOVA Output for Models on Popular Rating From Main Experimental Effects

```
                         Estimate Std. Error t value Pr(>|t|)
Selfdeclare2             -0.68859    0.47751  -1.442 0.149500
Selfdeclare3             -2.26053    1.03066  -2.193 0.028441
Selfdeclare4             -0.40603    0.43643  -0.930 0.352341
Selfdeclare5             -6.83172    2.09212  -3.265 0.001118
Selfdeclare6             -8.15218    2.83011  -2.881 0.004027
KnowRob1                  1.57500    0.78382   2.009 0.044676
KnowRob5                 -1.30113    0.61801  -2.105 0.035428
```

Figure 20: Summary of Variable Selfdeclare and KnowRob in the Fixed Effect Model on Classical Rating

the optimal model for popular rating contains the three main experimental factors as the fixed effects, and random effects with variation on the intercept and slope of Instrument.

Lastly, we examined the fixed effects in the final model through model summary in Figure 5. The model summary shows that all predictor variables (or some levels of predictor variables) are significant at 5% level, with t-values exceed the (-2, 2) interval.

4.3 Model Construction for Classical and Popular Ratings from All Variables of Interest

4.3.1 Classical Rating

We first selected the fixed effects through stepwise method from a full model including all main experimental effect with necessary interactions, and all other predictors of interest. Both AIC and BIC are chosen as the criterion for variable selection. As expected, AIC prefers a model with more variables than BIC. While both models include variables X16.minus.17, ConsInstr, ConsNotes, PachListen, ClsListen, X1990s2000s, and Composing other than the main experimental factors, AIC recommends the inclusion of two additional predictors, namely, Selfdeclare and KnowRob. The model summary of the AIC model was further investigated, and it is found out that some levels of these two variables are quite significant, with p values smaller than 0.05, relatively great effect sizes with small standard error (see Figure 20). From above, we concluded that Selfdeclare and KnowRob should be included in the model.

For random effects, the random variations of the effect of Harmony and Instrument across participants is recommended by AIC and BIC method, along with the random variations on the classical ratings across participants. Then we added the random effects of Instrument and Harmony

```
Models:
lm_fixed: Classical ~ Harmony + Instrument + Voice + Harmony:Voice + Selfdeclare +
     X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen +
     KnowRob + X1990s2000s + Composing
lmer_final: Classical ~ Harmony + Instrument + Voice + Harmony:Voice + Selfdeclare
    + X16.minus.17 + ConsInstr + ConsNotes + PachListen + ClsListen +
    KnowRob + X1990s2000s + Composing + (Harmony + Instrument | Subject)
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lm_fixed   55 6443.0 6736.7 -3166.5   6333.0
lmer_final 76 6099.4 6505.3 -2973.7   5947.4 385.51     21  < 2.2e-16
```

Figure 21: ANOVA Output to Choose between Model with Only Fixed Effects and with Random Effects for Classical Ratings

across Subject to see whether the model can be improved. The ANOVA result shows that the inclusion of the random effects significantly improve the model by reducing AIC and BIC statistics (see Figure 21).

Lastly, we examined the final model through model summary and residual plots. Figure 22, all significant variables from the model summary of the final models are listed. As we can see, all predictor variables (or some levels of predictor variables) are significant at 5% level, as the t-values exceed the (-2, 2) interval. This discovery affirms the validity of the previous variable selection. For residual analysis, the codes attached below (see Figure 23) are used to examine the marginal, conditional, and random effect residual plots. All three plots demonstrate nice patterns, with the both marginal and conditional residuals centered around zero, and all three residuals demonstrate observable vertical patterns. Overall, the model seem valid.

Variable selection is further affirmed by the automated method, with function fitLMER. This method first backward selects the fixed effects from the full model with all possible predictor variables, then forward selects the random effects, and lastly verified the fixed effects with the random effects included. This method gives the same model as our proposed final model, therefore further verifying the our choice on the final model for classical ratings. The code used for the method is attached below.

```
lmer_fixed = lmer(Classical ~ Instrument + Harmony + Voice + Harmony:Voice +
(1 | Subject) + Selfdeclare + log(OMSI) + X16.minus.17 + ConsInstr + ConsNotes
+ PachListen + ClsListen + KnowRob + KnowAxis + X1990s2000s + NoClass + Composing
+ PianoPlay + GuitarPlay, data = df)

lmer_random <- fitLMER.fnc(lmer_fixed,ran.effects=c("(Voice|Subject)",
"(Instrument|Subject)", "(Harmony|Subject)"), method="AIC")
```

4.3.2 Popular Rating

The final model for popular rating is proposed in a similar manner. The full model is built upon the optimal model for popular rating from the three main experimental effects, with the inclusions of all other predictor variables of interest (Selfdeclare, log(OMSI),X16.minus.17, ConsInstr,

```
Fixed effects:
                            Estimate Std. Error t value
HarmonyI-V-VI                1.26653    0.28435   4.454
Instrumentpiano              1.65278    0.23594   7.005
Instrumentstring             3.58760    0.31229  11.488
Selfdeclare3                -2.05609    0.90204  -2.279
Selfdeclare5                -6.07880    1.83105  -3.320
Selfdeclare6                -7.04287    2.47694  -2.843
X16.minus.17-2               3.28136    0.95018   3.453
X16.minus.17-1               1.52309    0.62645   2.431
X16.minus.170                3.69890    0.58036   6.373
X16.minus.171                1.79859    0.54618   3.293
X16.minus.172                4.90847    1.09430   4.486
X16.minus.173                1.50984    0.60870   2.480
X16.minus.174               -2.36995    1.04116  -2.276
X16.minus.175                5.74854    1.80557   3.184
ConsInstr2                   2.94857    1.10278   2.674
ConsInstr4                   2.32443    1.07505   2.162
ConsNotes1                   2.18657    0.76911   2.843
ConsNotes3                   2.69698    0.36032   7.485
ConsNotes4                  -3.11151    0.58192  -5.347
PachListen3                 -4.41992    1.95240  -2.264
ClsListen1                  -2.55037    0.65669  -3.884
ClsListen5                  -4.36151    1.55603  -2.803
KnowRob5                    -1.07141    0.54085  -1.981
X1990s2000s2                 2.10707    0.91663   2.299
X1990s2000s3                -2.16734    0.73684  -2.941
Composing1                  -1.26150    0.60040  -2.101
Composing2                   1.99342    0.34968   5.701
Composing3                  -1.94468    0.53860  -3.611
HarmonyI-V-IV:Voicepar3rd -0.42917    0.27607  -1.555
HarmonyI-V-VI:Voicepar3rd -0.70786    0.27662  -2.559
HarmonyIV-I-V:Voicepar3rd  0.75182    0.27607   2.723
```

Figure 22: Regression Summary of Significant Variables in the Final Model on Classical Ratings

```
attach(df)
resid.marg <- r.marg(lmer_final)
resid.cond <- r.cond(lmer_final)
resid.reff <- r.reff(lmer_final)

sch <- as.numeric(Subject)
index <- sch
for (j in unique(sch)) {
  len <- sum(sch==j)
  index[sch==j] <- 1:len
}

new.data <- data.frame(index,resid.marg,Subject)
names(new.data) <- c("index","resid.marg","Subject")
ggplot(new.data,aes(x=index,y=resid.marg)) +
  facet_wrap( ~Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
  geom_hline(yintercept=0)

new.data <- data.frame(index,resid.cond,Subject)
names(new.data) <- c("index","resid.cond","Subject")
ggplot(new.data,aes(x=index,y=resid.cond)) +
  facet_wrap( ~Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
  geom_hline(yintercept=0)

new.data <- data.frame(index,resid.reff,Subject)
names(new.data) <- c("index","resid.reff","Subject")
ggplot(new.data,aes(x=index,y=resid.reff)) +
  facet_wrap( ~Subject, as.table=F) +
  geom_point(pch=1,color="Blue") +
  geom_hline(yintercept=0)
detach(df)
```

Figure 23: Codes for Residual Analysis for Final Model on Classical Rating

```
Models:
lm_pop_full: Popular ~ Instrument + Harmony + Voice
lmer_pop: Popular ~ Instrument + Harmony + Voice + (Instrument | Subject)
            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lm_pop_full  9 6910.5 6958.6 -3446.2   6892.5
lmer_pop    15 6358.5 6438.6 -3164.2   6328.5 564.04      6  < 2.2e-16
```

Figure 24: ANOVA Output to Choose the Best Model for Popular Ratings

ConsNotes, PachListen, ClsListen, knowRob, KnowAxis, X1990s2000s,NoClass, AP Theory, CollegeMusic, Composing, PianoPlay, and GuitarPlay). Stepwise variable selection is performed with both AIC and BIC as the criterion, with both methods recommend the model including only the three main experimental factors.

We then added the random effects into the model for improvement. We found out the model fail to converge if random effects on Harmony and Voice are introduced. On the other hand, the ANOVA output suggests that the inclusion of random variation on the effect of Instrument on rating across Subject improves the model with only fixed effect by decreasing the AIC and BIC statistics (see Figure 24). In this way, the model on popular ratings from all predictor variables happens to be the same as the model from the three main experimental factors.

We further verified the model by investigating the model summary and residual plots. Model summary (see Figure 5) suggests that all predictor variables (or levels of predictor variables) are statistically significant at 5% level. The residual plots for the model seem valid, with the mean of marginal and conditional residuals centered around 0, and no observable vertical patterns in all marginal, conditional, and random effect residuals.

Lastly, the model was further affirmed by the automated method fitLMER, which gives the same model as our choice. The codes used for residual plots and fitLMER for popular model are the same as those used for classical model.

4.4 A Different Dichotomization of Respondents to Musician and Non-Musician

If we instead set the threshold of dichotomization as 2 in the response to Selfdeclare, the result becomes different. From the table provided in Figure 25, the statistically significant interaction terms are the interaction between Musician and Instrument, X16.minus.17, and ConsNotes, and the result suggests that musicians tend to give lower score for music played in piano and string quartet, compared to non musician; musicians tend to give higher score given on unit increase in the ability of distinguishing classical and popular music, compared to non-musicians; musicians tend to give higher classical score than non-musicians given both group concentrate on notes when listening to music.

According to the above analysis, we can see that the result is indeed sensitive to the dichotomization. Even though both dichotomizations lead to the result of significance of interaction with Instrument, the above dichotomization indicates more significant interactions compared to the dichotomization provided in section 3.4.

Musician_2	-4.17752	1.06931	-3.907
HarmonyI-V-IV:Musician_2	0.12091	0.47362	0.255
HarmonyI-V-VI:Musician_2	1.37616	0.74291	1.852
HarmonyIV-I-V:Musician_2	0.70615	0.44716	1.579
Instrumentpiano:Musician_2	1.01332	0.38533	2.630
Instrumentstring:Musician_2	0.91799	0.38518	2.383
Voicepar3rd:Musician_2	0.50880	0.45566	1.117
Voicepar5th:Musician_2	0.30477	0.42384	0.719
X16.minus.170:Musician_2	2.22401	1.17939	1.886
X16.minus.171:Musician_2	3.26113	1.09894	2.968

Figure 25: Regression Summary of Interactions with Musician from A Different Dichotomization