Effect of Harmony, Voice, Instrument on Listener's Identification of Musical Type

Jinqi Wu*
jinqiw@andrew.cmu.edu

07 December 2019

Abstract

In this article we explored the effect of Harmony Motion, Instrument, and Voice Leading on listen's identification of music as "classical" or "popular". We firstly built a random intercept model with three design variables by using AIC criterion and anova test. Then we added other independent variables and other random effects into our previous model of classical ratings and popular ratings to obtain our best model. By checking the coefficients and significance of our final model, we found that Instrument exert the strongest influence among the three design factors and I-V-VI in Harmony and Contrary Motion in Voice Leading have a strong association with classical ratings. Furthermore, we showed in this article how do musicians and non-musicians identify classical music in different ways by implementing one way anova test. Finally, we compared the the two models about classical ratings and popular ratings in coefficients and significance to show differences in things that drive classical vs. popular ratings.

1 Introduction

In 2012, Ivan Jimenez, a composer and musicologist visiting the University of Pittsburgh, and student Vincent Rossi, designed an experiment of measuring the influence of instrument, harmonic, and voice leading on listeners' identification of music as "classical" or "popular".

In this article, we want to use the data collected by Ivan Jimenez to further explore the relationship between people's identification of musical type and instrument type, harmonic type, and voice type. This problem is crucial because this can help us determine the key factors affecting people's identification of music.

In addition to answering the main question posted above, we will address the following questions as well:

- What experimental factor, or combination of factors, has the strongest influence on ratings?
  - 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?

*Department of Statistics & Data Science, Carnegie Mellon University

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

- Are there differences in the way that musicians and non-musicians identify classical music?

- Are there differences in the things that drive classical, vs. popular, ratings?

The data utilized in this article is taken from homework 10:It provides 36 musical stimuli to 70 listeners, recruited from the population of undergraduates at the University of Pittsburgh. Answer of each listener's rate of the musical sound in two scales(classical and popular) is recorded. The data also contains experimental variables(Instrument, Harmonic Motion, and Voice Leading) and other important variables.

2 Methods

In this report, the variables in the dataset are represented in the data available to us. They are shown in Table 1 and Table 2. In these two tables, the Classical and Popular are our response variables and Harmony, Instrument and Voice are our experimental factors.

`Classical`	`How classical does the stimulus sound?`
`Popular`	`How popular does the stimulus sound?`
`Subject`	`Unique subject ID`
`Harmony`	`Harmonic Motion (4 levels)`
`Instrument`	`Instrument (3 levels)`
`Voice`	`Voice Leading (3 levels)`
`Selfdeclare`	`Are you a musician? (1-6, 1=not at all)`
`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 while listening (0-5, 0=not at all)`
`ConsNotes`	`How much did you concentrate on the notes while listening? (0-5, 0=not at all)`
`Instr.minus.Notes`	`Difference between prev. two variables`
`PachListen`	`How familiar are you with Pachelbel's Canon in D (0-5, 0=not at all)`
`ClsListen`	`How much do you listen to classical music? (0-5, 0=not at all)`
`KnowRob`	`Have you heard Rob Paravonian's Pachelbel Rant (0-5, 0=not at all)`

Table 1: The description of each variable

The data are available in the file `ratings.csv`, in HW 10 file in canvas.

Since the dataset contains many missing values, we decide to firstly clean the data and do some exploratory data analysis before building our own model.

KnowAxis	Have you heard Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music? (0-5, 0=not at all)
X1990s2000s	How much do you listen to pop and rock from the 90's and 2000's? (0-5, 0=not at all)
X1990s2000s.minus.1960s1970s	Difference between prev variable and a similar variable referring to 60's and 70'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 (0-5, 0=not at all)
PianoPlay	Do you play piano (0-5, 0=not at all)
GuitarPlay	Do you play guitar (0-5, 0=not at all)
X1stInstr	How proficient are you at your first musical instrument (0-5, 0=not at all)
X2ndInstr	Same, for second musical instrument
first12	In the experiment, which instrument was presented to the subject in the first 12 stimuli? (IGNORE FOR THIS ASSIGNMENT.)

Table 2: The description of each variable

When we check our dataset, we can see that there are 27 columns(variables) and 2520 rows. As is shown by Figure 1 and Figure 2, we can find that for some of the variables, they contain too many missing values such as X1stInstr(1512 missing values of 2520 in total). So we delete these variables in our dataset. The rule for deleting is to check whether the proportion of missing values is over 10% rows. By applying this rule, we finally decide to delete the following four columns: ConsNotes(360/2520), NoClass(288/2520), X1stInstr(1512/2520), X2ndInstr(2196/2520). And it is worth mentioned that although the missing values for KnowAxis also exceeds the 10% we still keep this variable because it is a very important variable for us to discuss question 1.

In this study, in order to address the first question proposed in the Introduction section, we have to first fit a best model.

Firstly, we use the stepAIC function to fit two conventional linear regression models based on AIC and BIC. Then we utilize the anova test to decide which model we are going to pick.

Secondly, after deciding our conventional model, we add random intercept to our model to see whether this would yield a better result. Anova test is used to determine whether random intercept is needed in our model. After that, we try random slope for each experimental factor and use anova test to determine which variable should include random slope.

Finally, we want to figure out which covariates should be added to the model as fixed effects and whether there should be any changes in random effects after adding other covariates. We start by adding all fixed variables into our model and use AIC criterion to fit a conventional regression model. After that, we use fitLMER function to further select fixed variables as well as determine

```
-- Variable type:factor --------------------------------------------------------
   variable missing complete    n n_unique                            top_counts ordered
    first12       0     2520 2520        3   str: 1080, gui: 720, pia: 720, NA: 0   FALSE
    Harmony       0     2520 2520        4 I-I: 630, I-V: 630, I-V: 630, IV-: 630   FALSE
 Instrument       0     2520 2520        3    gui: 840, pia: 840, str: 840, NA: 0   FALSE
    Subject       0     2520 2520       70        15: 36, 16: 36, 17: 36, 18b: 36   FALSE
      Voice       0     2520 2520        3    con: 840, par: 840, par: 840, NA: 0   FALSE

-- Variable type:integer -------------------------------------------------------
                   variable missing complete    n    mean     sd p0    p25    p50     p75
                   APTheory     216     2304 2520    0.23   0.42  0   0        0       0
                  ClsListen      36     2484 2520    2.16   1.59  0   1        3       3
                CollegeMusic    108     2412 2520    0.79   0.41  0   1        1       1
                  Composing      72     2448 2520    1      1.46  0   0        0       2
                  ConsNotes     360     2160 2520    2.53   1.95  0   0.75     3       5
                 GuitarPlay       0     2520 2520    0.69   1.48  0   0        0       1
                   KnowAxis     288     2232 2520    0.9    1.91  0   0        0       0
                    KnowRob     180     2340 2520    0.77   1.72  0   0        0       0
                    NoClass     288     2232 2520    0.92   1.5   0   0        1       1
                       OMSI       0     2520 2520  225.93 231.32 11  49      145.5   323
                 PachListen      72     2448 2520    4.51   1.1   0   5        5       5
                  PianoPlay       0     2520 2520    1.09   1.72  0   0        0       1
                 Selfdeclare      0     2520 2520    2.44   1.18  1   2        2       3
                          X       0     2520 2520 1260.5  727.61  1 630.75 1260.5 1890.25
                 X1990s2000s    144     2376 2520    4.06   1.56  0   3        5       5
 X1990s2000s.minus.1960s1970s   180     2340 2520    2.02   1.92 -4   0        2       3
                  X1stInstr    1512     1008 2520    2.79   1.59  1   1        3.5     4
                  X2ndInstr    2196      324 2520    1.56   1.17  0   1        1       2
```

Figure 1: Overview of Dataset

```
-- Variable type:numeric -------------------------------------------------------
           variable missing complete    n mean   sd p0  p25  p50  p75  p100     hist
          Classical      27     2493 2520 5.78 2.66  0 4    6    8    19
          ConsInstr       0     2520 2520 2.86 1.58  0 1.67 3    4.33  5
 Instr.minus.Notes        0     2520 2520 0.69 1.69 -4 0    0.34 2     4.33
            Popular      27     2493 2520 5.38 2.5   0 4    5    7    19
        X16.minus.17      0     2520 2520 1.72 2.99 -4 0    1    3     9
```

Figure 2: Overview of Dataset

what else random effects should be included in our model. After all, we get our final best model.

Since now we have a best model, we can use this model to explain the questions proposed in the Introduction Section. For the first question in question 1, we check the significance and coefficient of each level of Instrument variable to give conclusions about Instrument variable. Similar methods are used to resolve second question in question 1. In addition, as our final best model does not contain Pachelbel rants and Comedy Bits variables, we add these two variables to our final model manually and check the significance and coefficients of these two variables. Anova test is used to decide whether these two variables matter for the response variable.

For the third question in question one, we reduce the intercept in order too show all levels of Voice Leading variable to comprehensively compare their coefficients and significance. By checking their coefficients and significance, we can then figure out whether contrary motion has a strong association with classical ratings.

Regarding the second question, we consider different cutting boundary to define musicians and non-musicians. Then we pick up one boundary to create a new variable musician represents whether this participant considers himself as musician or not. One way anova test is used here to show the relationship between musician variable and classical ratings and the interaction of musician with other variables. By checking the p-values, we can figure out how musician variable affects the relationship between classical ratings and other variables.

About the last question, we compare our final model about classical ratings and popular ratings comprehensively by checking their variables and coefficients. According to the results, we can make convincing conclusions to determine the differences in things that drive classical ratings vs. popular ratings.

All the computing results above are calculated by R language and environment for statistical computing.

3 Results

Firstly, we consider the relationship between classical rating and these experimental factors. As is shown by Figure 3, the conventional linear regression model we fit by AIC is Model 1 and the conventional linear regression model we fit by BIC is Model 2. From the result, we can see that the p-value for the test is 0.003. As the null hypothesis is that Voice and Harmony:Voice should not be included in the final model. So we reject the null hypothesis. Thus we the conventional linear model we choose is

$$\text{Classical} \sim \text{Instrument} + \text{Harmony} + \text{Voice} + \text{Harmony} : \text{Voice}$$

After that, we compare this final conventional regression model with the model adding random intercept by using anova test. Following Figure 4 shows the result of our anova test.

```
Model 1: Classical ~ Instrument + Harmony + Voice + Harmony:Voice
Model 2: Classical ~ Instrument + Harmony
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1   1851 9382.3
2   1859 9500.9 -8   -118.51 2.9225 0.003033 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: Anova Test of AIC and BIC Model

```
Data: rate
Models:
aic_model: Classical ~ Instrument + Harmony + Voice + Harmony:Voice
lmer_line: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1 |
lmer_line:     Subject)
          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
aic_model 15 8335.7 8418.6 -4152.8   8305.7
lmer_line 16 7883.7 7972.2 -3925.8   7851.7 453.99      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4: Anova Test of Conventional Model and Random Intercept Model

The p-value for the anova test is less than 0.05 and the null hypothesis is that random effect is not necessary in this model. So we decide to reject the null hypothesis. The random intercept is needed in the final model.

So after adding random intercept, our model becomes:

$$Classical_i = \alpha_{0j[i]} + \alpha_1 Instrument + \alpha_2 Harmony + \alpha_3 Voice + \alpha_4 Harmony : Voice + \varepsilon_i$$

$$\alpha_0 = \beta_0 + \eta_j, \quad \eta_j \overset{indep}{\sim} N\left(0, \tau_0^2\right)$$

Then we try to determine which variable should include random slope by checking the AIC and BIC value.

```
Data: rate
Models:
lmer_line: Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1 |
lmer_line:     Subject)
lmer.1: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.1:     (1 + Instrument | Subject)
lmer.2: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.2:     (1 + Voice | Subject)
lmer.3: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.3:     (1 + Harmony | Subject)
lmer.5: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.5:     (1 + Instrument + Voice | Subject)
lmer.4: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.4:     (1 + Instrument + Harmony | Subject)
lmer.6: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.6:     (1 + Voice + Harmony | Subject)
lmer.7: Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
lmer.7:     (1 + Instrument + Harmony + Voice | Subject)
          Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)    
lmer_line 16 7883.7 7972.2 -3925.8   7851.7                              
lmer.1    21 7599.1 7715.3 -3778.6   7557.1 294.547      5     <2e-16 ***
lmer.2    21 7893.6 8009.8 -3925.8   7851.6   0.000      0          1    
lmer.3    25 7810.1 7948.3 -3880.0   7760.1  91.541      4     <2e-16 ***
lmer.5    30 7610.9 7776.9 -3775.5   7550.9 209.128      5     <2e-16 ***
lmer.4    36 7459.3 7658.4 -3693.6   7387.3 163.669      6     <2e-16 ***
lmer.6    36 7822.8 8022.0 -3875.4   7750.8   0.000      0          1    
lmer.7    51 7468.2 7750.3 -3683.1   7366.2 384.642     15     <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: Anova Test of Random Slope

Figure 5 shows that lmer.4 holds the lowest AIC and BIC value. This indicates that the model in can be improved by adding Instrument and Harmony random effect terms. So the best combination would be

$$\begin{gathered}
\text{Classical }_i = \alpha_{0j[i]} + \alpha_{1j[i]} \text{ Instrument }_i + \alpha_{2j[i]} \text{ Harmony }_i + \alpha_3 \text{ Voice }_i + \\
\alpha_4 \text{ Harmony } * \text{ Voice } + \epsilon_i, \quad \epsilon_i \overset{indep}{\sim} N\left(0, \sigma^2\right) \\
\alpha_{0j} = \beta_0 + \eta_{0j}, \quad \eta_{0j} \overset{indep}{\sim} N\left(0, \tau_0^2\right) \\
\alpha_{1j} = \beta_0 + \eta_{1j}, \quad \eta_{1j} \overset{indep}{\sim} N\left(0, \tau_1^2\right) \\
\alpha_{2j} = \beta_0 + \eta_{2j}, \quad \eta_{2j} \overset{indep}{\sim} N\left(0, \tau_2^2\right)
\end{gathered}$$

By using stepAIC function and fitLMER function to further select fixed variables and random effects, we get our best model for classical rating:

$$
\begin{gathered}
\text{Classical}_i = \alpha_{0j[i]} + \alpha_{1j[i]} \text{ Instrument}_i + \alpha_{2j[i]} \text{ Harmony}_i + \alpha_3 \text{ Voice}_i + \\
+ \alpha_4 Selfdeclare + \alpha_5 OMSI + \alpha_6 X16.minus.17 + \alpha_7 ClsListen + \\
\alpha_8 \text{ Harmony} * \text{Voice} + \epsilon_i, \quad \epsilon_i \overset{indep}{\sim} N\left(0, \sigma^2\right) \\
\alpha_{0j} = \beta_0 + \eta_{0j}, \quad \eta_{0j} \overset{indep}{\sim} N\left(0, \tau_0^2\right) \\
\alpha_{1j} = \beta_0 + \eta_{1j}, \quad \eta_{1j} \overset{indep}{\sim} N\left(0, \tau_1^2\right) \\
\alpha_{2j} = \beta_0 + \eta_{2j}, \quad \eta_{2j} \overset{indep}{\sim} N\left(0, \tau_2^2\right)
\end{gathered}
$$

By using similar methods, we can get the final best model for popular rating as well:

$$
\begin{gathered}
\text{Popular}_i = \alpha_{0j[i]} + \alpha_{1j[i]} \text{ Instrument}_i + \alpha_2 \text{ Harmony}_i + \alpha_3 \text{ Voice}_i, \quad \epsilon_i \overset{indep}{\sim} N\left(0, \sigma^2\right) \\
\alpha_{0j} = \beta_0 + \eta_{0j}, \quad \eta_{0j} \overset{indep}{\sim} N\left(0, \tau_0^2\right) \\
\alpha_{1j} = \beta_0 + \eta_{1j}, \quad \eta_{1j} \overset{indep}{\sim} N\left(0, \tau_1^2\right)
\end{gathered}
$$

The coefficients for each model are shown as below in Figure 6 and Figure 7

```
Random effects:
 Groups    Name               Variance  Std.Dev. Corr
 Subject   (Intercept)        0.0001182 0.01087
 Subject.1 Instrumentguitar   0.4415937 0.66453
           Instrumentpiano    1.1904408 1.09107  -0.05
           Instrumentstring   1.5636361 1.25045  -0.97  0.14
 Subject.2 HarmonyI-IV-V      0.9970629 0.99853
           HarmonyI-V-IV      1.5230578 1.23412   1.00
           HarmonyI-V-VI      2.2016499 1.48380   0.43  0.49
           HarmonyIV-I-V      0.8105167 0.90029   0.96  0.98  0.43
 Residual                     2.4704067 1.57175
Number of obs: 1865, groups:  Subject, 52

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       5.12250    0.49804  10.285
Voicepar3rd      -0.38978    0.08917  -4.371
Voicepar5th      -0.31030    0.08917  -3.480
HarmonyI-V-IV     0.02951    0.10885   0.271
HarmonyI-V-VI     0.89632    0.21773   4.117
HarmonyIV-I-V     0.09064    0.10966   0.827
Instrumentpiano   1.52789    0.20199   7.564
Instrumentstring  3.45735    0.27863  12.408
Selfdeclare      -0.69417    0.20682  -3.356
OMSI              0.83822    0.20822   4.026
X16.minus.17     -0.12176    0.04685  -2.599
ClsListen         0.35029    0.09468   3.700
```

Figure 6: Coefficients for Final Classical Rating Model

```
Random effects:
 Groups     Name               Variance Std.Dev. Corr
 Subject    (Intercept)        1.2301   1.1091
 Subject.1  Instrumentguitar   0.2253   0.4747
            Instrumentpiano    1.1113   1.0542   -0.14
            Instrumentstring   1.4864   1.2192   -0.87  0.42
 Residual                      2.9255   1.7104
Number of obs: 1865, groups:  Subject, 52

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       6.73825    0.20130  33.474
HarmonyI-V-IV    -0.00456    0.11200  -0.041
HarmonyI-V-VI    -0.33544    0.11206  -2.993
HarmonyIV-I-V    -0.25700    0.11193  -2.296
Instrumentpiano  -1.11710    0.19468  -5.738
Instrumentstring -2.99010    0.24851 -12.032
Voicepar3rd       0.22656    0.09703   2.335
Voicepar5th       0.24359    0.09703   2.510
```

Figure 7: Coefficients for Final Popular Rating Model

Next we are going to check the influence of Instrument Variable. When the response variable is Classical Rating, we can see from Figure 6 that both Instrumentpiano and Instrumentstring are significant in our final model because both of their t-values exceed 1.96. In addition, when we check the random effect of Instrumentpiano and Instrumentstring, we can see that their Standard Deviation are 1.09 and 1.25. So there is approximately 95% chance that the interval [1.53 - 2 * 1.09, 1.53 + 2 * 1.09] = [-0.65,3.71] and [3.45 - 2 * 1.25, 3.45 + 2 * 1.25] = [0.95,5.95] contain the true coefficients for Instrumentpiano and Instrumentstring. We can find that the interval of coefficient of Instrumentpiano contains zero, however most of the data would exceed zero. In addition, the interval of coefficient of Instrumentstring exceeds zero, which mean that keep other variables fixed, if we change the variable Instrumentguitar(baseline) to Instrumentpiano or Instrumentstring, there would be a strong positive increase in classical ratings. Compared the coefficients of fixed effects, we can find that the Instrument variable has the largest coefficient. This proves that Instrument exerts the strongest influence among three design factors.

When the response variable is Popular Rating, we can see from Figure 7 that both Instrumentpiano and Instrumentstring exceed -1.96, which means that these two variables are significant in our final model. Additionally, when we check the random effects of Instrument variable, we can find that the Standard Deviations for Instrumentpiano and Instrumentstring are accordingly 1.05 and 1.22. So there is approximately 95% chance that the interval [-1.12 - 2 * 1.05, -1.12 + 2 * 1.05] = [-3.22,0.98] and [-2.99 - 2 * 1.22, -2.99 + 2 * 1.22] = [-5.43,-0.55] contain the true coefficients for Instrumentpiano and Instrumentstring. We can find that the interval of coefficient of Instrumentpiano contains zero, however most of the data would be below zero. In addition, the interval of coefficient of Instrumentstring is below zero, which means that keep other variables fixed, if we change the variable Instrumentguitar(baseline) to Instrumentpiano or Instrumentstring, there would be a relatively strong negative increase in popular ratings. Compared the coefficients of fixed effects, we can find that the Instrument variable has the largest coefficient(absolute value). This proves that Instrument exerts the strongest influence among three design factors.

So the conclusion for the first question in question 1 would be yes. Instrument does exert the strongest influence among the three design factors as the researchers suspect.

After that, we want to figure out whether I-V-VI has a strong association with classical ratings among all levels of Harmonic Motion. At first, we check the t-value for this level and find that the t-value for HarmonyI-V-VI is $4.12 > 1.96$. This indicates that HarmonyI-V-VI is significant in our model. Besides, HarmonyI-V-VI holds the largest coefficient(0.90) compared to other two levels:HarmonyI-V-IV(0.03) and HarmonyIV-I-V(0.09), which means if we change HarmonyI-IV-V into HarmonyI-V-VI, there will be 0.90 increase in classical ratings. In conclusion, I-V-VI has the strongest association with classical ratings among all levels in Harmonic Motion.

```
Data: NULL
Models:
lmer_classical: Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
lmer_classical:     X16.minus.17 + ClsListen + (1 | Subject) + (0 + Instrument |
lmer_classical:     Subject) + (0 + Harmony | Subject)
lmer_KnowRob: Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
lmer_KnowRob:     X16.minus.17 + ClsListen + KnowRob + (1 | Subject) + (0 +
lmer_KnowRob:     Instrument | Subject) + (0 + Harmony | Subject)
               Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
lmer_classical 30 7470.8 7636.7 -3705.4   7410.8
lmer_KnowRob   31 7473.2 7644.7 -3705.6   7411.2     0      1          1
refitting model(s) with ML (instead of REML)
Data: NULL
Models:
lmer_classical: Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
lmer_classical:     X16.minus.17 + ClsListen + (1 | Subject) + (0 + Instrument |
lmer_classical:     Subject) + (0 + Harmony | Subject)
lmer_KnowAxis: Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
lmer_KnowAxis:     X16.minus.17 + ClsListen + KnowAxis + (1 | Subject) + (0 +
lmer_KnowAxis:     Instrument | Subject) + (0 + Harmony | Subject)
               Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
lmer_classical 30 7470.8 7636.7 -3705.4   7410.8
lmer_KnowAxis  31 7481.8 7653.3 -3709.9   7419.8     0      1          1
refitting model(s) with ML (instead of REML)
Data: NULL
Models:
lmer_classical: Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
lmer_classical:     X16.minus.17 + ClsListen + (1 | Subject) + (0 + Instrument |
lmer_classical:     Subject) + (0 + Harmony | Subject)
lmer_both: Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
lmer_both:     X16.minus.17 + ClsListen + KnowRob + KnowAxis + (1 | Subject) +
lmer_both:     (0 + Instrument | Subject) + (0 + Harmony | Subject)
               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer_classical 30 7470.8 7636.7 -3705.4   7410.8
lmer_both      32 7473.4 7650.4 -3704.7   7409.4 1.3992      2     0.4968
```

Figure 8: Anova Test for Including KnowAxis and KnowRob

Next we want to know whether Pachelbel and Comedy Bits matter with classical ratings. From Figure 8 we can see that none of the anova tests has a significant result. And after adding KnowAxis and KnowRob into our final model, the AIC increases from 7470 to 7473. So we decide not to include these two variables in our final model and it does not matter whether the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits.

For the final part of question one, the coefficient each variable is as shown in Figure 9. This Figure is based on eliminating the baseline of Voice to comprehensively show the impact of each level. From Figure 9 we can figure out that all of the coefficients of these three levels in Voice are significant. The t-values for contrary, par3rd, and par5th are 10.3, 9.5 and 9.7. Their coefficients are accordingly 5.12, 4.73, 4.81. As the coefficient for Voicecontrary holds the largest value, we can conclude that contrary motion has the strongest association with classical ratings.

```
Random effects:
 Groups    Name              Variance  Std.Dev. Corr
 Subject   (Intercept)       0.0002068 0.01438
 Subject.1 Instrumentguitar  0.4676562 0.68385
           Instrumentpiano   1.2215422 1.10523  -0.01
           Instrumentstring  1.5901873 1.26103  -0.91  0.16
 Subject.2 HarmonyI-IV-V     0.9709297 0.98536
           HarmonyI-V-IV     1.4967678 1.22342  1.00
           HarmonyI-V-VI     2.1753125 1.47489  0.42  0.48
           HarmonyIV-I-V     0.7828967 0.88481  0.96  0.98  0.42
 Residual                    2.4701819 1.57168
Number of obs: 1865, groups:  Subject, 52

Fixed effects:
                 Estimate Std. Error t value
Voicecontrary     5.12322    0.49812  10.285
Voicepar3rd       4.73345    0.49814   9.502
Voicepar5th       4.81292    0.49813   9.662
HarmonyI-V-IV     0.02949    0.10901   0.270
HarmonyI-V-VI     0.89632    0.21766   4.118
HarmonyIV-I-V     0.09064    0.10967   0.826
Instrumentpiano   1.52787    0.20203   7.563
Instrumentstring  3.45734    0.27862  12.409
Selfdeclare      -0.69430    0.20686  -3.356
OMSI              0.83854    0.20825   4.027
X16.minus.17     -0.12174    0.04686  -2.598
ClsListen         0.35009    0.09470   3.697
```

Figure 9: Summary of Classical Rating Model by Eliminating Baseline

To solve the second question, we use the boundary 2.5 to split the data to make musician and non-musician approximately the same and create a new variable musician. If the Selfdecare exceeds 2.5, then the musician variable would be 1, else it would be 0. Then we use one way anova analysis to see which interaction variable is significant in our test. As as shown by the Figure 10, musician:Harmony, musician:Instrument, musician:X16.minus.17 are three significant interaction variables in the model. This means that whether the participant is a musician will affect the relationship between Harmony, Instrument, ClsListen and classical ratings.

Regarding the last question, from Figure 6 and Figure 7 we can compare the coefficients of these two models for classical ratings and popular ratings. For the final model of classical ratings, there are 7 variables included, which contains 3 design variables and other 4 independent variables. In contrast, for the final model of popular ratings, there are only three variables in the final model,

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
51 observations deleted due to missingness
                       Df Sum Sq Mean Sq F value   Pr(>F)    
musician                1    184   184.1  37.189 1.24e-09 ***
Harmony                 3     35    11.7   2.357  0.06993 .  
Instrument              2   2914  1456.8 294.220  < 2e-16 ***
Voice                   2     17     8.3   1.680  0.18653    
OMSI                    1     16    15.5   3.136  0.07671 .  
X16.minus.17            1     97    97.0  19.586 1.00e-05 ***
ClsListen               1      0     0.1   0.017  0.89618    
musician:Harmony        3     69    23.1   4.671  0.00294 ** 
musician:Instrument     2     52    26.1   5.274  0.00518 ** 
musician:Voice          2      3     1.5   0.300  0.74064    
musician:OMSI           1      1     0.5   0.106  0.74440    
musician:X16.minus.17   1      1     0.6   0.125  0.72323    
musician:ClsListen      1     35    35.1   7.079  0.00785 ** 
Residuals            2447  12116     5.0                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
51 observations deleted due to missingness
```

Figure 10: One Way Anova Test for Including Musician Variable

which are three design variables and the model's random effects.

In addition, besides the number of variables, the coefficients of variables are also different. The similar thing is that Instrument variable exerts the strongest influence among all three design factors in both models of popular ratings and classical ratings. But the difference is that when we look at Harmony variable, we can find that HarmonyI-V-VI has a relatively strong positive relationship with classical ratings and the coefficient for HarmonyI-V-VI is 0.90. However, in popular ratings model, HarmonyI-V-VI only shows weak negative relationship with popular ratings(coefficient:-0.33).

What's more, from Figure 6 we can see that OMSI has a relatively strong relationship with classical ratings as the coefficient for OMSI is 0.83 and the t-value for it is 4.02(>1.96). So this is a significant and important variable to determine the final classical ratings model. In contrast, the popular ratings model does not even contain this variable.

Finally, we can see that in classical ratings model, there are random effects of Instrument and Harmony. However in popular ratings model, there is only random effect on Instrument variable. All of these things above indicate that there are differences in things that drive classical vs. popular ratings.

4 Discussion

In the Results section, we have shown that the best model for classical ratings is

$$
\begin{gathered}
\text{Classical }_i=\alpha_{0j[i]}+\alpha_{1j[i]} \text{ Instrument }_i+\alpha_{2j[i]} \text{ Harmony }_i+\alpha_3 \text{ Voice }_i+ \\
+\alpha_4 Selfdeclare+\alpha_5 OMSI+\alpha_6 X16.minus.17+\alpha_7 ClsListen+ \\
\alpha_8 \text{ Harmony } * \text{ Voice }+\epsilon_i, \quad \epsilon_i \overset{indep}{\sim} N\left(0, \sigma^2\right) \\
\alpha_{0j}=\beta_0+\eta_{0j}, \quad \eta_{0j} \overset{indep}{\sim} N\left(0, \tau_0^2\right)
\end{gathered}
$$

$$\alpha_{1j} = \beta_0 + \eta_{1j}, \quad \eta_{1j} \overset{indep}{\sim} N\left(0, \tau_1^2\right)$$
$$\alpha_{2j} = \beta_0 + \eta_{2j}, \quad \eta_{2j} \overset{indep}{\sim} N\left(0, \tau_2^2\right)$$

And the best model for popular ratings is

$$\text{Popular}_{\ i} = \alpha_{0j[i]} + \alpha_{1j[i]} \text{ Instrument}_{\ i} + \alpha_2 \text{ Harmony}_{\ i} + \alpha_3 \text{ Voice}_{\ i}, \quad \epsilon_i \overset{indep}{\sim} N\left(0, \sigma^2\right)$$
$$\alpha_{0j} = \beta_0 + \eta_{0j}, \quad \eta_{0j} \overset{indep}{\sim} N\left(0, \tau_0^2\right)$$
$$\alpha_{1j} = \beta_0 + \eta_{1j}, \quad \eta_{1j} \overset{indep}{\sim} N\left(0, \tau_1^2\right)$$

As we discussed in Results section, the Instrument variable exerts the strongest influence among three design factors in both popular ratings model and classical ratings model. This conclusion coincides with our intuitive because using instrument to distinguish the music type is a very direct way. Usually if the music belongs to classical category, then the musician would tend to use piano or string(violin,cello.etc) to play the classical music. On the other hand, if the music type is popular, musician will be more likely to use guitar to play this kind of music.

Using Voice and Harmony to distinguish popular music and classical music may not be very intuitive because for those ordinary people, they have not accepted systematic training for the music theory, it's difficult for them to tell the difference between Voice and Harmony in popular and classical music. For those musicians, Voice and Harmony could be a good way to distinguish because they have more domain knowledge. But for ordinary people, using instrument would be a more intuitive and direct way.

For the second question in question 1, we can see that HarmonyI-V-VI has the strongest relationship with classical ratings. And it does matter whether the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits. This conclusion is reasonable because these two people are very famous musicians. Having heard of their music represents that the participant has a passion for music. Since the participant has a passion for it, he would tend to learn more musical domain knowledge. This would make him easier to distinguish classical and popular music.

For the last question in question 1, we can see from the result that contrary motion has the strongest association with classical ratings. Contrary motion is motion in opposite directions. That is, when one of the lines moves up, the other line moves down. Parallel motion at an interval of a perfect fifth is known as parallel or consecutive fifths, and at an interval of an octave is known as parallel or consecutive octaves. Compared with parallel motion, contrary motion is easier for listener to distinguish and for performer to play because the fingering is the same in both hands. People usually like the sound of a contrary motion scale, and it is so easily done before they can put hands together in a tune. That may be the reason why contrary motion has a more influential association with classical ratings compared with parallel motion.

Then when we turn our eye on the result in question 2, we can find that there is a difference in the way that musicians and non-musicians identify classical music. From Figure 10 we can see that whether the participant is a musician or not will affect the relationship between Harmony, Instrument, and ClsListen and classical ratings. This is probably because if the participant deem himself as a musician, he may consider himself having more domain knowledge in music field. And distinguish Harmony Motion and Instrument correctly needs strong domain knowledge in music field. Especially for classical music, the rules are much more complicated than popular music and it needs professional knowledge about music. In addition, as the ClsListen variable represents for how much does the participant listen to classical music, the relationship between ClsListen and classical ratings will be definitely affected whether the participant is a musician or not. For those people who deem themselves as musicians, they usually listen to much classical music. This would cause a stronger connection between classical ratings and CLsListen. So to conclude, there is a difference in the way that musicians and non-musicians identify classical music.

Finally, when we check the result of question 3 in Figure 6 and Figure 7, we can see that four more variables are included in the final model of classical ratings compared to popular ratings. They are Selfdeclare, OMSI, X16.minus.17, and ClsListen. The reason why the final classical ratings model contains more variables may be that classical music is usually considered more complicated than popular music. In order to give ratings for classical music, people have to consider in multiple perspectives.

If we dig into these four variables, the result seems coincide with our intuitive. First, Selfdeclare represents for whether the participant consider himself as a musician or not. Usually, participant who really knows the expertise of classical music would be very humble and courteous, so generally they may give themselves lower ratings for Selfdeclare compared with their real ratings. That's why there is a negative relation between Selfdeclare and classical ratings. Secondly, OMSI represents that Score on a test of musical knowledge. If people has more domain knowledge in classical music, he will definitely get a higher score on OMSI. And people who have more domain knowledge in classical music will tend to give higher classical ratings. That's why there is a relatively strong positive relationship between classical ratings and OMSI.

After that, X16.minus.17 seems also have an impact on classical ratings according to Figure 6 but the relationship may be mot very strong(coefficient:-0.12). This may because the score of X16.minus.17 is based on auxiliary measure instead of direct measure. The result of this variable may not be very convincing. Additionally, ClsListen has a relatively strong impact on classical ratings(coefficient:0.35). This result is very intuitive because if the participants usually listen to classical music quite often, then they are more likely to give higher classical ratings since they get more exposed to classical music and have a deeper understanding of classical music.

On the other hand, popular music is simpler compared with classical music. So in the final model there are only three design factors in the final model. And there are only random effects on intercept and Instrument variables rather than intercept, Instrument, and Harmony in classical ratings model. This represents that the coefficient for Harmony variable varies in group in classical ratings model but fixed in popular ratings model. This is because for classical ratings, Harmony is very complicated to distinguish and usually Harmony in classical music is more changeable than Harmony in popular music.

Overall, our final models regarding classical ratings and popular ratings are based on many variable selection process. The strength for this is that our final model only include a few variables in our model, which makes our model more interpretable and less complex.

However, our model also has many drawbacks. One is that when choosing variables in our final model, we use stepwise a lot of times. Although this could delete those variables which may be not very correlated with our model mathematically. This may also delete many variables which are very important in real life because some of the variables have their own meanings in real life and it may not be an ideal way to delete them all.

Furthermore, as our data we use in this study is taken from Ivan Jimenez et al. (2012), the information generally pertains to the years 2012. The information could be changing after almost 8 year passed. So the conclusion we get in this report may be convincing in 8 years ago, but may not be very reasonable today.

In our future research, we will consider not only choosing each variable mathematically but also choosing it meaningfully. This would better help us understand the true relationship between response variables and independent variables.

In conclusion, in this study we find out the relationship between classical ratings, popular ratings and Harmony, Instrument, and Voice. We finally build two robust models to comprehensively show the relationship between them. These two models can help us figure out how do these three design factors affect classical ratings and popular ratings. In the future, we will try to learn more musical knowledge to analyze our models in more musical ways to optimize our models.

References

Sheather, S.J. (2009), *A Modern Approach to Regression with R.* New York: Springer Science + Business Media LLC.

Code Appendix

```
#packages
library(skimr)
library(MASS)
library(lme4)
library(arm)
library(RLRsim)
library(LMERConvenienceFunctions)
library(dplyr)
library(ggplot2)

#data overview and cleaning
rate <- read.csv("ratings.csv")
rate <- rate[,-1]
rate <- subset(rate,select = -c(X1stInstr,X2ndInstr,first12,ConsNotes,NoClass))
rate$OMSI <- scale(rate$OMSI,center = T,scale = T)
rate <- na.omit(rate)
skim(rate)
attach(rate)

#eda
boxplot(rate$Classical ~ rate$Harmony)
boxplot(rate$Classical ~ rate$Instrument)
boxplot(rate$Classical ~ rate$Voice)
boxplot(rate$Popular ~ rate$Harmony)
boxplot(rate$Popular ~ rate$Instrument)
boxplot(rate$Popular ~ rate$Voice)

#fit conventional line and using AIC and BIC to choose final model
conv_line <- lm(Classical ~ Instrument*Harmony*Voice, data = rate )
```

```
aic_model <- stepAIC(conv_line,direction = "both")
bic_model <- stepAIC(conv_line,direction = "both",k=log(nrow(rate)))
anova(aic_model,bic_model)
summary(aic_model)
# the result shows that aic_model is better

#fit lmer line with random intercept
lmer_line <- lmer(Classical ~ Instrument + Harmony + Voice + Harmony:Voice + (1|Subject),
data = rate, control = lmerControl("bobyqa"),REML = FALSE)
summary(lmer_line)

# anova test
anova(lmer_line,aic_model)

lmer.1 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Instrument|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.2 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Voice|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.3 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Harmony|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.4 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Instrument + Harmony|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.5 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Instrument + Voice|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.6 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Voice + Harmony|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.7 <- lmer(Classical ~ 1 + Instrument + Harmony + Voice + Harmony:Voice +
(1 + Instrument + Harmony + Voice|Subject),REML = F, control = lmerControl(optimizer = "bobyqa
data = rate)

# implement the anova test
anova(lmer_line, lmer.1, lmer.2, lmer.3, lmer.4, lmer.5, lmer.6, lmer.7)

# translate the numerical column into categorical
rate$CollegeMusic <- as.factor(rate$CollegeMusic)
rate$APTheory <- as.factor(rate$APTheory)
str(rate)

fixed_line <- lm(Classical ~ . - Popular - Subject, data = rate)
```

```
# choose fixed variables by using stepAIC
stepAIC(fixed_line, direction = "both")

# fitlmer for classical rating
lmer_classical <- lmer(formula = Classical ~ Harmony + Instrument + Voice + Selfdeclare +
OMSI + X16.minus.17 + ConsInstr + Instr.minus.Notes + PachListen + ClsListen + KnowRob + X1990s

fitLMER.fnc(lmer_classical,ran.effects = c("(Selfdeclare|Subject)","(OMSI|Subject)","(X16.minus


#summary of classical rating model
lmer_classical <- lmer(Classical ~ -1 + Voice + Harmony + Instrument + Selfdeclare + OMSI +
X16.minus.17 + ClsListen + (1 | Subject) + (0 + Instrument |
Subject) + (0 + Harmony | Subject),optimizer = "bobyqa")
summary(lmer_classical)


# anova test for including KnowAxis and KnowRob
lmer_KnowRob <- lmer(Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
X16.minus.17 + ClsListen + KnowRob + (1 | Subject) + (0 + Instrument |
Subject) + (0 + Harmony | Subject),optimizer = "bobyqa")

lmer_KnowAxis <- lmer(Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
X16.minus.17 + ClsListen + KnowAxis + (1 | Subject) + (0 + Instrument |
Subject) + (0 + Harmony | Subject),optimizer = "bobyqa")

lmer_both <- lmer(Classical ~ Voice + Harmony + Instrument + Selfdeclare + OMSI +
X16.minus.17 + ClsListen + KnowRob + KnowAxis + (1 | Subject) + (0 + Instrument |
Subject) + (0 + Harmony | Subject),optimizer = "bobyqa")

anova(lmer_classical,lmer_KnowRob)
anova(lmer_classical,lmer_KnowAxis)
anova(lmer_classical,lmer_both)


#dicotomization
rate_new <- rate
rate_new$musician <- 0
rate_new$musician[which(rate_new$Selfdeclare >= 2)] <- 1
rate_new$musician <- as.factor(rate_new$musician)
summary(aov(Classical ~ musician*(Harmony + Instrument + Voice + OMSI +
X16.minus.17 + ClsListen),data = rate_new))

rate_new <- rate
```

```
rate_new$musician <- 0
rate_new$musician[which(rate_new$Selfdeclare >= 3)] <- 1
rate_new$musician <- as.factor(rate_new$musician)
summary(aov(Classical ~ musician*(Harmony + Instrument + Voice + OMSI +
X16.minus.17 + ClsListen),data = rate_new))

rate_new <- rate
rate_new$musician <- 0
rate_new$musician[which(rate_new$Selfdeclare >= 4)] <- 1
rate_new$musician <- as.factor(rate_new$musician)
summary(aov(Classical ~ musician*(Harmony + Instrument + Voice + OMSI +
X16.minus.17 + ClsListen),data = rate_new))

rate_new <- rate
rate_new$musician <- 0
rate_new$musician[which(rate_new$Selfdeclare >= 5)] <- 1
rate_new$musician <- as.factor(rate_new$musician)
summary(aov(Classical ~ musician*(Harmony + Instrument + Voice + OMSI +
X16.minus.17 + ClsListen),data = rate_new))


# follows is the way we use similar method to deal with popular ratings
conv_line <- lm(Popular ~ Instrument*Harmony*Voice, data = rate)

aic_model <- stepAIC(conv_line,direction = "both")
bic_model <- stepAIC(conv_line,direction = "both",k=log(nrow(rate)))

anova(aic_model,bic_model)
summary(aic_model)

#fit lmer with random intercept
lmer_line <- lmer(Popular ~ Instrument + Harmony + Voice + (1|Subject),
data = rate, control = lmerControl("bobyqa"),REML = FALSE)
anova(lmer_line,aic_model)

lmer.1 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
(1 + Instrument|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.2 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
(1 + Voice|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.3 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
(1 + Harmony|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.4 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
```

```
(1 + Instrument + Harmony|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.5 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
(1 + Instrument + Voice|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.6 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
(1 + Voice + Harmony|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"),
data = rate)
lmer.7 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice +
(1 + Instrument + Harmony + Voice|Subject),REML = F, control = lmerControl(optimizer = "bobyqa"
data = rate)

# implement the anova test
anova(lmer_line, lmer.1, lmer.2, lmer.3, lmer.4, lmer.5, lmer.6, lmer.7)

fixed_line <- lm(Popular ~ . - Classical - Subject, data = rate)

# choose fixed variables by using stepAIC
stepAIC(fixed_line, direction = "both")

lmer_Popular <- lmer(formula = Popular ~ Harmony + Instrument + Voice + Selfdeclare + X16.minus
ConsInstr + Instr.minus.Notes + PachListen + ClsListen +
KnowRob + X1990s2000s + X1990s2000s.minus.1960s1970s + CollegeMusic +
APTheory + Composing + GuitarPlay + (1|Subject) + (0 + Instrument|Subject) + (0 + Harmony), REM

fitLMER.fnc(lmer_Popular,ran.effects = c("(X16.minus.17|Subject)", "(ConsInstr)","(Instr.minus
"(APTheory|Subject)","(Composing|Subject)"),
method = "llrt")

summary(lmer(Popular ~ Harmony + Instrument + Voice + (1 | Subject) + (0 +  Instrument | Subje
```