Analysis of the Influence of Timbre, Harmony, and Voice Leading on Listeners' Distinction between Popular and Classical Music

Laura Zhang*
wenjiaz1@andrew.cmu.edu

06 December 2019

Abstract

This paper studies the influence three design factors (instrument, harmonic motion, and voice leading) have on listeners' ratings on classical and popular music. The data is taken from an experiment designed and conducted by Dr. Ivan Jimenez and Vincent Rossi. We find that, out of all three design factors, instrument has the strongest influence on classical and popular ratings. It is also possible for us to build sensible models to study specific effects each variable in the data set have on the two types of ratings.

1 Introduction

Classical and popular music are two of the fundamental music genres. Composers and musicians around the world are interested in how different factors influence people's perception of classical versus popular music. An experimental task was conducted by Dr. Ivan Jimenez, a composer and musicologist visiting the University of Pittsburgh, and student Vincent Rossi. Within the experiment, 36 musical stimuli was presented to 70 listeners, and listeners were asked to indicate the extent to which the musical stimuli were classical or popular music sounding. Specifically, Dr. Jimenez and Vincent aimed to study three experimental factors (Instrument, Harmonic motion, and Voice leading) and their influences on classical/popular ratings.

This paper will focus on analyzing the three main experimental factors, along with some other experiment-related variables and their effects on listeners' classical and popular ratings. The research questions that we will address are as follows:

- Which experimental factor, or combination of factors, has the strongest influence on ratings?
- Are there differences in the way that musicians and non-musicians identify classical music?
- Are there differences in the things that drive classical versus popular ratings?

The data, and the entire research problem as stated above, are provided by Dr. Ivan Jimenez, who is now working at the Sibelius Institute, University of Arts, Helsinki, Finland.

*Department of Statistics & Data Science, Carnegie Mellon University

2 Methods

2.1 Description of Data

The data for this study are available on the 70 listeners' ratings on 36 musical stimuli. Ratings are given on two different scales:

- How classical does the music sound (1 to 10, 1 = not at all, 10 = very classical sounding)
- How popular does the music sound (1 to 10, 1 = not at all, 10 = very popular sounding)

The 36 stimuli were chosen by completely crossing these three main experimental factors:

- Instrument: String Quartet, Piano, Electric Guitar
- Harmonic Motion: I-V-VI, I-VI-V, I-V-IV, IV-I-V
- Voice Leading: Contrary Motion, Parallel 3rds, Parallel 5ths

Other than the response variables (Classical and Popular ratings) and the three main experimental factors (Instrument, Harmonic motion and Voice leading), there are also data available for other factors. A brief description of all variables available in the data set is as follows:

Y_1 = Classical = How classical does the stimulus sound?
Y_2 = Popular = How popular does the stimulus sound?
x_1 = Subject = Unique subject ID
x_2 = Harmony = Harmonic Motion (4 levels)
x_3 = Instrument = Instrument (3 levels)
x_4 = Voice = Voice Leading (3 levels)
x_5 = Selfdeclare = Are you a musician? (1-6, 1=not at all)
x_6 = OMSI = Score on a test of musical knowledge
x_7 = X16.minus.17 = Auxiliary measure of listener's ability to distinguish classical vs popular music
x_8 = ConsInstr = How much did you concentrate on the instrument while listening? (0-5, 0=not at all)
x_9 = ConsNotes = How much did you concentrate on the notes while listening? (0-5, 0=not at all)
x_{10} = Instr.minus.Notes = Difference between prev. two variables
x_{11} = PachListen = How familiar are you with Pachelbel's Canon in D? (0-5, 0=not at all)
x_{12} = ClsListen = How much do you listen to classical music? (0-5, 0=not at all)
x_{13} = KnowRob = Have you heard Rob Paravonian's Pachelbel Rant? (0-5, 0=not at all)
x_{14} = KnowAxis = Have you heard Axis of Evil's Comedy bit on the 4 Pachelbel chords in popular music? (0-5, 0=not at all)
x_{15} = X1990s2000s = How much do you listen to pop and rock from the 90's and 2000's? (0-5, 0=not at all)

x_{16} = X1990s2000s.minus.1960s1970s = Difference between prev. variable and a similar variable referring to 60's and 70's pop and rock

x_{17} = CollegeMusic = Have you taken music classes in college? (0=no, 1=yes)

x_{18} = NoClass = How many music classes have you taken?

x_{19} = APTheory = Did you take AP Music Theory class in High School? (0=no, 1=yes)

x_{20} = Composing = Have you done any music composing? (0-5, 0=not at all)

x_{21} = PianoPlay = Do you play the piano? (0-5, 0=not at all)

x_{22} = GuitarPlay = Do you play the guitar? (0-5, 0=not at all)

x_{23} = X1stInstr = How proficient are you at your first musical instrument? (0-5, 0=not at all)

x_{24} = X2ndInstr = Same, for second musical instrument

2.2 Data Cleaning

Since the data that we obtained came straight from the experimental study, a thorough cleaning of the data needed to be performed before we could dive into modeling and analysis. The raw data set contained 26 variables and 2,520 rows of observations. As a routine, we first checked for missing values and NA entries within the raw data.

Variable `X1stInstr` had 1,512 NA entries and variable `X2ndInstr` had 2,196 NA entries. Considering that we only had 2,520 observations in total, both variables were missing over half of the entries. In this particular case, imputation of any kind did not seem like a reasonable choice to deal with this level of missing data, since over half of the observations were missing and we had no prior knowledge of the reason behind these missing data. Therefore, we performed column-wise deletion and completely excluded these two variables from the data set.
Similar logic applied when we checked for missing data in other variables. We also eliminated variables `X1990s2000s`, `X1990s2000s.minus.1960s1970s`, `NoClass`, and `APTheory`, because they contained too many missing entries, making it hard for us to justify imputations of any kind.

For the rest of data cleaning, we excluded all remaining missing entries by performing row-wise deletion, and ended up with 20 variables and 1,793 rows of observations.

Taking a closer look at our two response variables, Classical and Popular ratings, we realized that for both scales, some entries are 19, which is outside of the indicated range of rating. This could be because the number 9 and 0 are so close to each other on the computer keyboard that people mis-typed their ratings. To fix this problem, we will change all entries with value 19 to 10.

2.3 Modeling Overview

For our analysis, we will start off with some visualizations and exploratory data analysis to provide a guideline for the rest of the modeling and interpretations. The models that we will build are all hierarchical models with both fixed effects and random effects. Some of the models are justified using residual analysis. Detailed R language code can be found in the Technical Appendix at the end of this paper.

3 Results

3.1 Exploratory Data Analysis

In this section, we will exploratory data analysis (EDA) results on variables within the cleaned data set. First off, we will start by plotting histograms (Figure 1) of the two response variables to detect any patterns and distributions that might be out of the ordinary.

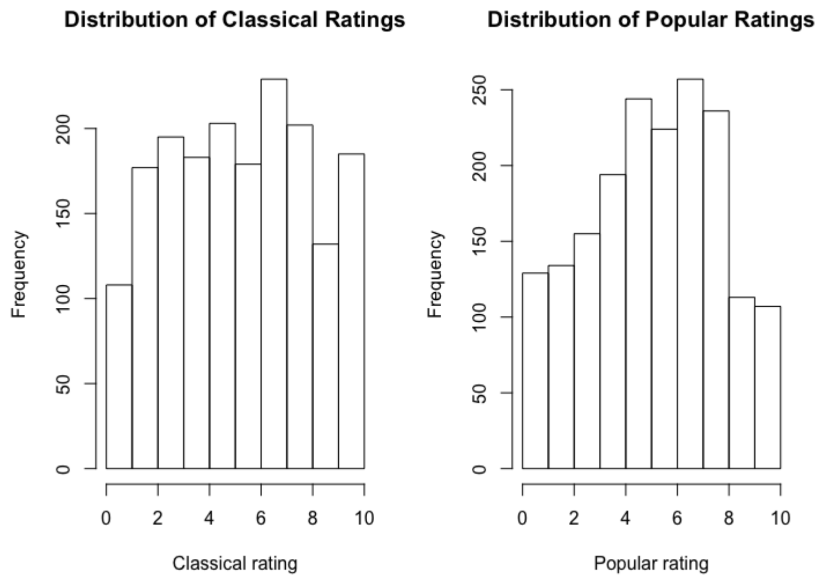

Figure 1: Histograms on Classical and Popular Ratings

Figure 1 shows the above mentioned histograms. Just by looking at the histograms, there is nothing that we should be particularly concerned about. The distributions are neither left-skewed or right-skewed. This means that no transformation is needed, and we can directly make use of these two ratings as response variables in building models and performing analysis.

In order to study the influence experimental factors have on ratings, we first constructed barplots (Figure 2) on average ratings based on all three experimental factors (`Instrument`, `Harmony`, `Voice`).

From the plots, it is not hard to compare the two different types of ratings. Based on the instruments being played in the stimuli, there are differences in both classical and popular ratings associated with different type of instruments. Notice that, interestingly, the differences are in reverse order for classical and popular ratings. Harmony I-V-VI seems to make a difference on classical ratings more than the other three harmonic motions. Other than the above observations, listeners' average ratings seem to not change much with harmonic motions and voice leading.

Although we cannot formally establish the extent of influence different experimental factors have on ratings, these initial EDA results are useful in helping to guide us through model analysis and interpretations in the following sections of this paper.

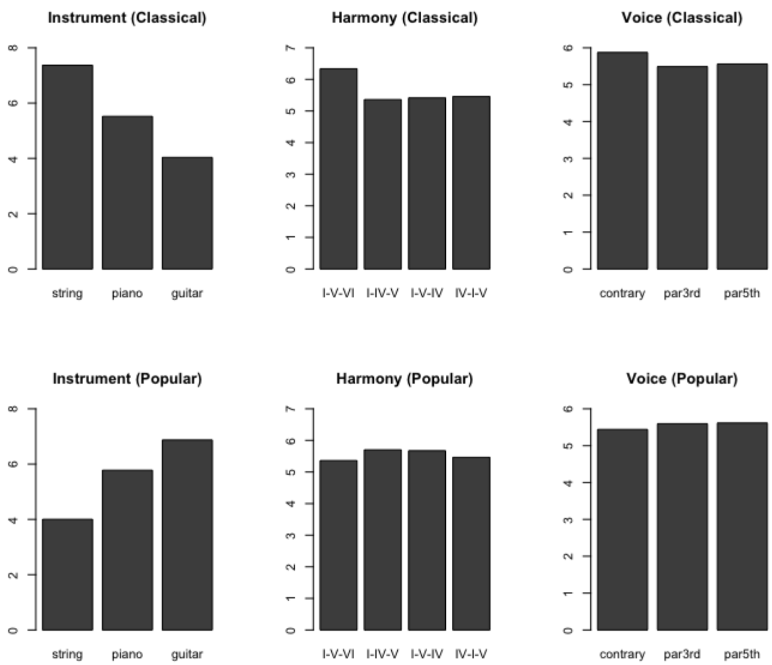

Figure 2: Barplots of Classical and Popular ratings on all three experimental factors

3.2 Which experimental factor, or combination of factors, has the strongest influence on ratings?

Since this data was collected from 70 listeners, we assumed that ratings would be different based on different listeners' perceptions of the three main experimental factors. Therefore, before building our actual model for answering this research question, we decided to fit a random intercept for each listener. In this case, the model that we have will be a hierarchical linear model with both fixed effects and random effects (with random intercept for listener).

In order to examine each experimental factor's influence on ratings and also combination of factors and their influences on ratings, we fitted hierarchical linear models with three experimental factors, their interactions of all orders as fixed effects and experimental factors based on each listener as random effects. In the following two subsections, we will present models and analysis on classical ratings and popular ratings, respectively.

3.2.1 Influences on Classical Ratings

Per Dr. Ivan Jimenez and Vincent Rossi's suggestions, for classical ratings in particular, we will not only study the influences that experimental factors have, but also whether or not the listener is familiar with Rob Paravonian's Pachelbel Rant and Axis of Evil's Comedy bit or not.

Therefore, in this case, the model that we fitted contains `Instrument`, `Harmony`, `Voice`, interactions among these three experimental factors, `KnowRob` and `KnowAxis` as fixed effects, and `1+Instrument+Harmony+Voice|Subject` as random effects.

The following table (Table 1) presents mostly coefficients of factors with significant influence (p-value less than 0.05) on classical ratings. A full summary report of this model is shown in the Technical Appendix (Table 6).

Guitar instrument, I-IV-V harmonic motion, Contrary Motion voice leading, `KnowRob` rating 0, and `KnowAxis` rating 0 are baseline variables. With the intercept being 3.94,

if the instrument is piano, classical rating is expected to increase by 1.18;

if the instrument is string, classical rating is expected to increase by 3.20;

if the harmony is I-V-IV, classical rating is expected to decrease by 0.24;

if the harmony is I-V-VI, classical rating is expected to increase by 1.01;

if the harmony is IV-I-V, classical rating is expected to decrease by 0.26;

if the voice is par3rd, classical rating is expected to decrease by 0.08;

if the voice is par5th, classical rating is expected to decrease by 0.46;

if `KnowRob` has rating 1, classical rating is expected to decrease by 0.26;

if `KnowRob` has rating 5, classical rating is expected to increase by 0.32;

if `KnowAxis` has rating 1, classical rating is expected to increase by 1.08;

if `KnowAxis` has rating 5, classical rating is expected to increase by 0.40;

if the harmony is fixed to be I-V-VI, when the voice is par3rd, classical rating is expected to decrease by 1.45 ;

if the instrument is fixed to be string, and the harmony is fixed to be I-V-VI, when the voice is par3rd, classical rating is expected to increase by 1.55.

The above interpretation of coefficients of fixed effects is useful in terms of helping us figure out what experimental factor, or combination of factors, has the strongest influence on classical ratings. Notice that both piano and string instrument have significant coefficients, especially for string instrument, it is associated with an increase of 3.20 for classical rating. This is a significant amount of increase in rating considering that the rating system only goes from 1 to 10. Taking a look at the coefficients of other levels of experimental factors, we can see that all, except I-V-VI, are insignificant. This could be a potential indication that `Instrument` exerts the strongest influence among all three design factors in this study. This result corresponds to what we found out in Section 3.1 EDA, where we noticed a change in classical rating based on different instruments.

Other than looking into coefficients of fixed effects, it is also crucial to interpret error terms of random effects to get a better understanding of whether or not the perception of these three design factors vary among listeners.

Figure 3 displays all random effects and their standard deviations. Other than the intercept, piano and string instruments and I-V-VI harmony have standard deviations greater than any of the other random effect factors. This means that the effects that these three predictors have on classical rating can vary more than the other predictors.

Combining the above results, we have evidence to believe that variable `Instrument` exerts the strongest influence on classical rating among the three design factors. Out of the different levels of harmony, only I-V-VI has significant coefficient and it has the strongest association with classical rating. In terms of the different levels of voice leading, only the baseline Contrary Motion within intercept has statistically significant coefficient, therefore Contrary Motion has the strongest influence on classical rating among all three levels of voice leading.

Even though variables indicating whether listeners are familiar with Pachelbel Rant and/or Comedy bit or not are not considered to have significant coefficients, there are differences in their influences on classical rating. For listeners who are not so familiar with Pachelbel Rant (`KnowRob1`), there is a negative influence on classical ratings; while for listeners who are familiar with Pachelbel

	Coefficients of Fixed Effects
(Intercept)	3.94^{***}
	(0.33)
Instrumentpiano	1.18^{**}
	(0.36)
Instrumentstring	3.20^{***}
	(0.41)
HarmonyI-V-IV	-0.24
	(0.31)
HarmonyI-V-VI	1.01^{**}
	(0.36)
HarmonyIV-I-V	-0.26
	(0.31)
Voicepar3rd	-0.08
	(0.31)
Voicepar5th	-0.46
	(0.31)
KnowRob1	-0.26
	(0.62)
KnowRob5	0.32
	(0.53)
KnowAxis1	1.08
	(1.37)
KnowAxis5	0.40
	(0.47)
HarmonyI-V-VI:Voicepar3rd	-1.45^{***}
	(0.43)
Instrumentstring:HarmonyI-V-VI:Voicepar3rd	1.55^{*}
	(0.60)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Mixed model on classical ratings

```
Error terms:
 Groups   Name               Std.Dev.
 Subject  (Intercept)        1.56
          Instrumentpiano    1.43
          Instrumentstring   1.93
          HarmonyI-V-IV      0.39
          HarmonyI-V-VI      1.31
          HarmonyIV-I-V      0.37
          Voicepar3rd        0.44
          Voicepar5th        0.37
 Residual                    1.51
```

Figure 3: Summary info on random effects in classical rating model

Rant (`KnowRob5`), there is a positive influence on classical ratings. For listeners who are not so familiar with Comedy bit (`KnowAxis1`), there is a greater positive influence on classical ratings than those who are familiar with Comedy bit (`KnowAxis5`). Overall, the familiarity of these two pieces does not seem to matter too much in terms of listeners' classical ratings, however, there does exist some level of differences between familiarity.

3.2.2 Influences on Popular Ratings

Similar to Section 3.2.1, we will now be looking into influences on popular ratings. Per experimenters' advice, we only study the influences design factors and their interactions have on popular ratings. The model that we fitted contains `Instrument`, `Harmony`, `Voice`, and interactions among these three experimental factors as fixed effects, and `1+Instrument+Harmony+Voice|Subject` as random effects.

Table 2 presents mostly coefficients of factors with significant influence (p-value less than 0.05) on popular ratings. A full summary report of this model is shown in the Technical Appendix (Table 7).

Same as the mixed model we fitted for classical ratings, in the case of popular ratings, guitar instrument, I-IV-V harmonic motion, and contrary motion voice leading are baseline variables. With the intercept being 6.80,

if the instrument is piano, popular rating is expected to decrease by 1.18;
if the instrument is string, popular rating is expected to decrease by 2.96;
if the harmony is I-V-IV, popular rating is expected to increase by 0.32;
if the harmony is I-V-VI, popular rating is expected to decrease by 0.18;
if the harmony is IV-I-V, popular rating is expected to increase by 0.04;
if the voice is par3rd, popular rating is expected to increase by 0.08;

	Coefficients of Fixed Effects
(Intercept)	6.80^{***}
	(0.28)
Instrumentpiano	-1.18^{**}
	(0.36)
Instrumentstring	-2.96^{***}
	(0.39)
HarmonyI-V-IV	0.32
	(0.32)
HarmonyI-V-VI	-0.18
	(0.34)
HarmonyIV-I-V	0.04
	(0.33)
Voicepar3rd	0.08
	(0.32)
Voicepar5th	0.30
	(0.32)
Instrumentpiano:HarmonyI-V-VI:Voicepar3rd	-1.34^{*}
	(0.63)
Instrumentstring:HarmonyI-V-VI:Voicepar3rd	-1.85^{**}
	(0.63)
Instrumentstring:HarmonyIV-I-V:Voicepar3rd	-1.74^{**}
	(0.63)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Mixed model on popular ratings

if the voice is par5th, popular rating is expected to increase by 0.30;

if the instrument is fixed to be piano, and the harmony is fixed to be I-V-VI, when the voice is par3rd, popular rating is expected to decrease by 1.34;

if the instrument is fixed to be string, and the harmony is fixed to be I-V-VI, when the voice is par3rd, popular rating is expected to decrease by 1.85;

if the instrument is fixed to be string, and the harmony is fixed to be IV-I-V, when the voice is par3rd, popular rating is expected to decrease by 1.74.

Comparing these coefficients interpretations with our previous ones on classical ratings, we found that once again both piano and string instrument have significant coefficients. Only this time, both of these coefficients are negative. Out of all three experimental factors, `Instrument` is the only one with statistically significant coefficients. This might indicate that variable `Instrument` has the strongest influence on popular rating than the other two design factors.

```
Error terms:
 Groups   Name            Std.Dev.
 Subject  (Intercept)     1.19
          Instrumentpiano 1.24
          Instrumentstring 1.66
          HarmonyI-V-IV   0.42
          HarmonyI-V-VI   0.96
          HarmonyIV-I-V   0.64
          Voicepar3rd     0.29
          Voicepar5th     0.33
 Residual                 1.57
```

Figure 4: Summary info on random effects in popular rating model

Figure 4 shows all random effects and their standard deviations for the mixed model on popular ratings. Notice that the intercept, piano and string have greater standard deviations than all other design factors, which means the influences different instruments have on popular ratings vary more than harmony and voice.

Combining the above analysis, we have reasons to believe that `Instrument` has the strongest influence on popular ratings. It also makes sense for both piano and string to have negative coefficients, rather than positive ones for popular ratings. This distinguishes classical music from popular music, at the same time, this kind of contradiction aligns with the barplots we presented in Section 3.1.

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

In addition to studying just the three main experimental factors, we are also interested in learning about whether listeners' musical talent plays a role in affecting their classical ratings. Therefore, we will introduce a new variables `dichotomize` to split listeners into musicians and non-musicians according to their `Selfdeclare` information. If a listener has a `Selfdeclare` value of 1 or 2, he or she is identified as a non-musician, and if his or her `Selfdeclare` value is 3 or 4 or 5 or 6, he or she is identified as a musician.

Before building a final model, we need to decide which variables to include in our model. Since `Instrument`, `Harmony`, `Voice` are design factors, we will always keep design factors in our model. Due to the fact that this research question is exploring musicians' vs. non-musicians' influence on classical ratings, we will include variable `dichotomize` into the final model. All other variables and interactions included in the model are decided by hand, based on how much adding each variable will reduce AIC value of the entire model. As a general rule of thumb, we only included variables that reduced AIC value by more than 3. Similar to previous models that we built in Section 3.2.1 and 3.2.2, this model will also be a hierarchical model with both fixed effects and random effects.

In the end, our model contains `Instrument`, `Harmony`, `Voice`, `dichotomize`, `X16.minus.17`, `ConsInstr`, `PianoPlay`, `dichotomize:X16.minus.17`, `dichotomize:ConsInstr`, and `dichotomize:PianoPlay` as fixed effects. The model also contains `Instrument-1+Harmony+Voice|Subject` as random effects.

The following table (Table 3) has coefficients of fixed effects of selected variables and interactions (mostly the ones with p-values less than 0.05). A full summary of this model is included in the Technical Appendix (Table 8).

If the instrument is guitar, classical rating is expected to increase by 4.50;

if the instrument is piano, classical rating is expected to increase by 5.98;

if the instrument is string, classical rating is expected to increase by 7.83;

if the harmony is I-V-IV, classical rating is expected to increase by 0.05;

if the harmony is I-V-VI, classical rating is expected to increase by 0.98;

if the harmony is IV-I-V, classical rating is expected to increase by 0.10;

if the voice is par3rd, classical rating is expected to decrease by 0.38;

if the voice is par5th, classical rating is expected to decrease by 0.31;

if the listener is a musician, classical rating is expected to decrease by 0.07;

if the listener plays the piano (rating 4), classical rating is expected to increase by 2.71;

if the listener is fixed to be a musician, when the listener can play the piano (rating 4), classical rating is expected to decrease by 1.04.

Looking at the coefficient for `dichotomize1` alone, we can say that being a musician does not have a lot of influence on his or her classical rating, since the rating is only expected to decrease by 0.07. However, if we look at the coefficient of the interaction term `dichotomize1:PianoPlay4`, 1.04 indicates a much stronger influence on classical rating. Therefore, regarding the research question whether there are any differences between how musicians and non-musicians identify classical music, the answer does not simply depend only on variable `dichotomize`. Being a musician alone might not have much influence on the listener's classical rating, but combining the musician identity (`dichotomize`) with other factors, such as whether the listener plays the piano or not, we might end up having much a greater influence.

	Coefficients of Fixed Effects
Instrumentguitar	4.50^{***}
	(0.99)
Instrumentpiano	5.98^{***}
	(1.00)
Instrumentstring	7.83^{***}
	(1.00)
HarmonyI-V-IV	0.05
	(0.12)
HarmonyI-V-VI	0.98^{***}
	(0.21)
HarmonyIV-I-V	0.10
	(0.11)
Voicepar3rd	-0.38^{***}
	(0.11)
Voicepar5th	-0.31^{**}
	(0.11)
dichotomize1	-0.07
	(1.68)
PianoPlay4	2.71^{**}
	(1.00)
dichotomize1:PianoPlay4	-1.04
	(1.49)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Mixed model on musician vs non-musician

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

3.4.1 Modeling and Interpretations

The last research question that we will focus on in this paper is related to the differences in variables that influence classical and popular ratings. Specifically, following our previous discussion in Section 3.2.1 and 3.2.2, we are not only interested in the different effects design factors have on classical and popular ratings, but also interested in other variables available to us and uncovering different driving factors in influencing the two types of ratings.

In order to answer this research question, we will fit two mixed models, one taking classical ratings as the response variable and the other one taking popular ratings as the response variable. Both models are mixed models with fixed effects and random effects. The methods that we use to fit these two models are the same. Fixed effects are selected first based on AIC values of the model. More in detail, all variables are included in the model in the beginning, and then backward selection was performed based on AIC values to determine the best variables to be used for estimating the two types of ratings. After we have chosen the fixed effects, we pick random effects with the aid of `fitLMER.fnc` function in R. This function is an automatic method to select random effects based on existing fixed effects. Again, AIC value is used here as a baseline for which random effects to include.

The two mixed models that we end up having are presented as follows:

- **Classical rating fixed effects:** `Instrument`, `Harmony`, `Voice`, `ConsInstr`, `ClsListen`, `Composing`, `GuitarPlay`
- **Classical rating random effects:** `1+Instrument|Subject`
- **Popular rating fixed effects:** `Instrument`, `Harmony`, `Voice`, `ConsInstr`, `PachListen`, `ClsListen`, `KnowRob`, `KnowAxis`, `CollegeMusic`, `Composing`, `GuitarPlay`
- **Popular rating random effects:** `1+Instrument|Subject`

Before diving into coefficient interpretations for both models, let us first take a look at the different variables selected. Both models contain the same random-effect variables, i.e. random intercept and instrument based on listeners (`Subject`). Other than the three main experimental factors which are included in both models, the popular rating model contains all variables included in the classical rating model and a few more of its own. `PachListen`, `KnowRob`, `KnowAxis`, and `CollegeMusic` are variables that are exclusively included in the popular rating model. It is interesting that there are more driven factors for popular ratings than for classical ratings. A further analysis of coefficients will help determine how big of an influence these extra driven factors have on popular ratings.

Table 4 displays coefficients of fixed effects of classical ratings model with p-values less than 0.05 and coefficients of all levels for `ConsInstr`, whereas Table 5 displays coefficients of fixed effects of popular ratings with p-value less than 0.05 and coefficients of variables that are exclusive to this model. Interpretations of these selected fixed effects are similar to what we presented in previous sections, such as Section 3.2 and 3.3.

For classical ratings, the instruments used in the stimuli and whether the listener plays the guitar or not seem to play bigger roles than other variables in the model. However, for popular ratings, aside from instrument having strong influence, variable `ConsInstr` also has strong influence, depending on its level. It seems like, in general, the more listeners concentrate on the instrument

	Coefficients of Fixed Effects
(Intercept)	2.84^{*}
	(1.28)
Instrumentpiano	1.47^{***}
	(0.22)
Instrumentstring	3.33^{***}
	(0.29)
HarmonyI-V-VI	0.97^{***}
	(0.11)
Voicepar3rd	-0.38^{***}
	(0.10)
Voicepar5th	-0.31^{**}
	(0.10)
ConsInstr0.67	2.70
	(1.77)
ConsInstr1	-1.97
	(1.11)
ConsInstr1.67	-0.63
	(1.26)
ConsInstr2	-0.82
	(1.59)
ConsInstr2.33	-0.76
	(1.33)
ConsInstr2.67	0.99
	(1.59)
ConsInstr3	-0.13
	(1.10)
ConsInstr3.33	-2.71
	(1.84)
ConsInstr3.67	-1.12
	(1.20)
ConsInstr4	0.64
	(1.77)
ConsInstr4.33	-0.73
	(1.12)
ConsInstr5	-1.00
	(1.09)
GuitarPlay2	2.92^{*}
	(1.42)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Mixed model on classical ratings

	Coefficients of Fixed Effects
(Intercept)	8.72^{***}
	(1.90)
Instrumentpiano	-1.10^{***}
	(0.20)
Instrumentstring	-2.87^{***}
	(0.25)
HarmonyI-V-VI	-0.34^{**}
	(0.11)
HarmonyIV-I-V	-0.24^{*}
	(0.11)
ConsInstr0.67	-5.46^{**}
	(1.86)
ConsInstr2	-4.20^{**}
	(1.63)
ConsInstr2.67	-2.70^{*}
	(1.25)
ConsInstr3	-2.53^{*}
	(1.09)
ConsInstr4	-3.04^{*}
	(1.51)
PachListen3	0.41
	(0.99)
PachListen5	−0.79
	(0.90)
	(0.86)
KnowRob1	1.11
	(0.81)
KnowRob5	0.30
	(0.71)
KnowAxis1	0.07
	(1.39)
KnowAxis5	−0.35
	(0.50)
CollegeMusic1	0.10
	(0.54)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Mixed model on popular ratings

while listening to the stimuli, the less they are likely to reduce their popular ratings. This particular variable is present in both classical and popular rating model, but it has quite different influences. For classical rating, whether listeners concentrate on the instrument or not mostly affect rating by less than 1, however, in the case of popular rating, the effect is at least larger than 2, and sometimes even up to 5.

In terms of coefficients for variables that are exclusive to the popular rating model, most of them have absolute values less than 1, so the effects are not really significant.

To summarize our above analysis, there definitely are differences in the factors that drive classical vs. popular ratings. `Instrument` seems to be a driving factor for both types of ratings and there appear to be more driving factors for popular ratings, such as whether listeners concentrate on the instrument while listening or not.

3.4.2 Model Justification: Residual Analysis

As a validation of the previous two models that we built, Figure 5 and 6 are Normal Q-Q plots of the conditional residuals.

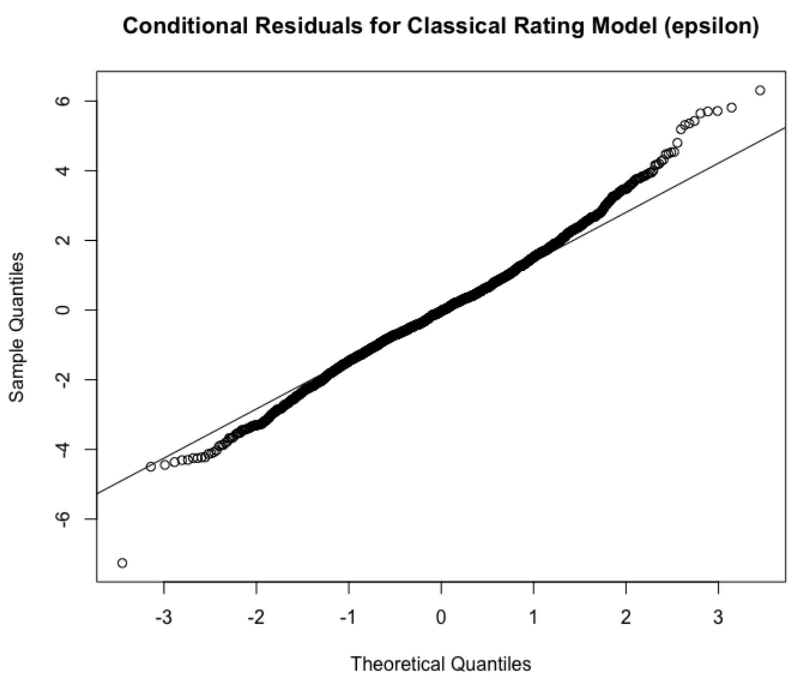

Figure 5: Normal Q-Q plot of conditional residuals for classical ratings model

Observe from both plots that residual points line up pretty well along the normal Q-Q line. Minor deviations are present at lower and upper tails, but there is nothing significant that we should be worried about. These two normal Q-Q plots provide justification that our models are sensible and our analysis was based on reasonable representation of the data.

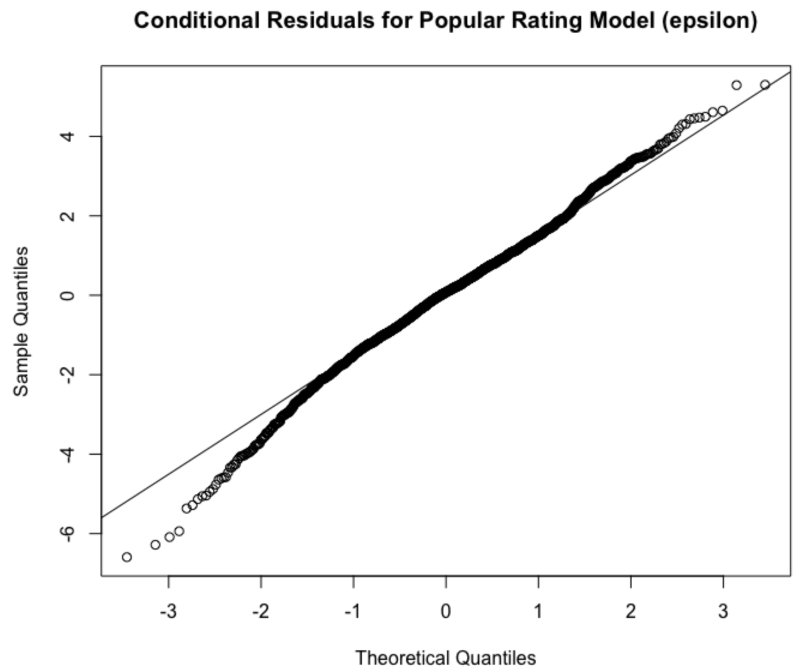

Figure 6: Normal Q-Q plot of conditional residuals for popular ratings model

4 Discussion

Over the course of this paper, we addressed three different research questions:

- Which experimental factor, or combination of factors, has the strongest influence on ratings?
- Are there differences in the way that musicians and non-musicians identify classical music?
- Are there differences in the things that drive classical versus popular ratings?

In Section 3.2, we analyzed influences of all three experimental factors and their interactions have on two types of ratings. For classical ratings specifically, we also looked at whether listeners' familiarity with Pachelbel Rants and Comedy Bits makes any difference on their ratings.

In Section 3.3, we introduced a new dichotomization variable in order to identify musicians and non-musicians. Once this new variable was introduced, we used it to build a mixed model in helping us uncover the differences between how musicians and non-musicians identify classical music.

In Section 3.4, we built two separate models, one for classical ratings and another for popular ratings. We were interested in finding out different driving factors and their influences on the two types of ratings.

Even though all of our models were able to support the analysis and answer each individual research question, there still exist limitations. For example, around 800 rows of observations were removed due to missing data. This is about 30% of the raw data set. Removing these observations may lead to uneven distribution of data and under-representation of some listeners' information.

Future research might focus on collecting a more complete data set prior to any modeling and analysis.

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/.

Technical Appendix

```
## Load the dataset
ratings <- read.csv("~/Desktop/36617/hw10/ratings.csv")

## Data Cleaning

# Initial EDA
sum(is.na(ratings$X1stInstr))
sum(is.na(ratings$X2ndInstr))
unique(ratings$Classical)
unique(ratings$Popular)

# Imputation for Classical and Popular ratings (19.0)
ratings[which(ratings$Classical == 19.0),]$Classical <- 10.0
ratings[which(ratings$Popular == 19.0),]$Popular <- 10.0

# Column-wise deletion
ratings <- subset(ratings, select = -c(X1stInstr, X2ndInstr, X1990s2000s,
X1990s2000s.minus.1960s1970s, NoClass, APTheory, first12))
ratings <- na.omit(ratings) ## Remove all NAs

ratings <- data.frame(lapply(ratings, as.character), stringsAsFactors = FALSE)
ratings$Classical <- as.numeric(ratings$Classical)
ratings$Popular <- as.numeric(ratings$Popular)
ratings$OMSI <- as.numeric(ratings$OMSI)

## EDA

# Histograms of ratings
par(mfrow = c(1,2))
hist(ratings$Classical, xlab = "Classical rating", main = "Distribution of Classical
Ratings")
hist(ratings$Popular, xlab = "Popular rating", main = "Distribution of Popular Ratings")

# Barplot of classical ratings on exp factors
par(mfrow = c(1,3))
```

```
string_cla <- mean(ratings[which(ratings$Instrument == 'string'),]$Classical)
piano_cla <- mean(ratings[which(ratings$Instrument == 'piano'),]$Classical)
guitar_cla <- mean(ratings[which(ratings$Instrument == 'guitar'),]$Classical)
instr_cla <- cbind(string_cla, piano_cla, guitar_cla)
barplot(instr_cla, ylim = c(0,8), main = "Instrument (Classical)", names.arg =
c("string", "piano", "guitar"))

IVVI_cla <- mean(ratings[which(ratings$Harmony == 'I-V-VI'),]$Classical)
IIVV_cla <- mean(ratings[which(ratings$Harmony == 'I-IV-V'),]$Classical)
IVIV_cla <- mean(ratings[which(ratings$Harmony == 'I-V-IV'),]$Classical)
IVI_V_cla <- mean(ratings[which(ratings$Harmony == 'IV-I-V'),]$Classical)
har_cla <- cbind(IVVI_cla, IIVV_cla, IVIV_cla, IVI_V_cla)
barplot(har_cla, ylim = c(0,7), main = "Harmony (Classical)", names.arg =
c("I-V-VI", "I-IV-V", "I-V-IV", "IV-I-V"))

contrary_cla <- mean(ratings[which(ratings$Voice == 'contrary'),]$Classical)
par3rd_cla <- mean(ratings[which(ratings$Voice == 'par3rd'),]$Classical)
par5th_cla <- mean(ratings[which(ratings$Voice == 'par5th'),]$Classical)
voice_cla <- cbind(contrary_cla, par3rd_cla, par5th_cla)
barplot(voice_cla, ylim = c(0,6), main = "Voice (Classical)", names.arg =
c("contrary", "par3rd", "par5th"))

# Barplot of popular ratings on exp factors
par(mfrow = c(1,3))
string_pop <- mean(ratings[which(ratings$Instrument == 'string'),]$Popular)
piano_pop <- mean(ratings[which(ratings$Instrument == 'piano'),]$Popular)
guitar_pop <- mean(ratings[which(ratings$Instrument == 'guitar'),]$Popular)
instr_pop <- cbind(string_pop, piano_pop, guitar_pop)
barplot(instr_pop, ylim = c(0,8), main = "Instrument (Popular)", names.arg =
c("string", "piano", "guitar"))

IVVI_pop <- mean(ratings[which(ratings$Harmony == 'I-V-VI'),]$Popular)
IIVV_pop <- mean(ratings[which(ratings$Harmony == 'I-IV-V'),]$Popular)
IVIV_pop <- mean(ratings[which(ratings$Harmony == 'I-V-IV'),]$Popular)
IVI_V_pop <- mean(ratings[which(ratings$Harmony == 'IV-I-V'),]$Popular)
har_pop <- cbind(IVVI_pop, IIVV_pop, IVIV_pop, IVI_V_pop)
barplot(har_pop, ylim = c(0,7), main = "Harmony (Popular)", names.arg =
c("I-V-VI", "I-IV-V", "I-V-IV", "IV-I-V"))

contrary_pop <- mean(ratings[which(ratings$Voice == 'contrary'),]$Popular)
par3rd_pop <- mean(ratings[which(ratings$Voice == 'par3rd'),]$Popular)
par5th_pop <- mean(ratings[which(ratings$Voice == 'par5th'),]$Popular)
voice_pop <- cbind(contrary_pop, par3rd_pop, par5th_pop)
barplot(voice_pop, ylim = c(0,6), main = "Voice (Popular)", names.arg =
c("contrary", "par3rd", "par5th"))
```

```
## Exp factors and their influences on ratings
library(lme4)
library(arm)
library(texreg)

# Classical rating
lmer.random.intercept.cla <- lmer(Classical ~ Instrument * Harmony * Voice
- 1 + KnowRob + KnowAxis + (Instrument-1+Harmony+Voice|Subject),
data = ratings, REML = F)
display(lmer.random.intercept.cla)
texreg(lmer.random.intercept.cla)

# Popular rating
lmer.random.intercept.pop <- lmer(Popular ~ Instrument * Harmony * Voice - 1 +
(Instrument-1+Harmony+Voice|Subject), data = ratings, REML = F)
display(lmer.random.intercept.pop)
texreg(lmer.random.intercept.pop)

## Musicians vs non-musicians (dichotomization)
ratings$dichotomize <- NA
half <- length(ratings$Selfdeclare) / 2
ratings <- ratings[order(ratings$Selfdeclare),]
no.1 <- length(ratings[which(ratings$Selfdeclare == 1),]$Selfdeclare)
no.2 <- length(ratings[which(ratings$Selfdeclare == 2),]$Selfdeclare)
no.3 <- length(ratings[which(ratings$Selfdeclare == 3),]$Selfdeclare)
no.4 <- length(ratings[which(ratings$Selfdeclare == 4),]$Selfdeclare)
no.5 <- length(ratings[which(ratings$Selfdeclare == 5),]$Selfdeclare)
no.6 <- length(ratings[which(ratings$Selfdeclare == 6),]$Selfdeclare)
ratings$dichotomize[1:(no.1 + no.2)] <- 0
ratings$dichotomize[(no.1+no.2+1):1793] <- 1
ratings$dichotomize <- as.factor(ratings$dichotomize)

mdl.dichotomize <- lmer(Classical ~ Instrument - 1 + Harmony + Voice +
dichotomize + as.numeric(X16.minus.17) + ConsInstr + PianoPlay +
dichotomize:as.numeric(X16.minus.17) + dichotomize:ConsInstr +
dichotomize:PianoPlay + (Instrument-1+Harmony+Voice|Subject), data = ratings)
display(mdl.dichotomize)
texreg(mdl.dichotomize)

## Variable selection classical vs popular rating

# Classical rating
library(LMERConvenienceFunctions)
lmer.classical <- lmer(Classical ~ Instrument + Harmony + Voice + ConsInstr +
```

```
ClsListen + Composing + GuitarPlay + (1+Instrument|Subject), data = ratings)
fitLMER.fnc(lmer.classical, ran.effects=c("(Instrument|Subject)",
"(Harmony|Subject)", "(Voice|Subject)"), method="AIC")
display(lmer.classical)
texreg(lmer.classical)


# Conditional residual analysis for classical rating
r.cond <- function(m) {residuals(m)}
resid.cond.cla <- r.cond(lmer.classical)

par(mfrow = c(1,1))
qqnorm(resid.cond.cla, main="Conditional Residuals for Classical
Rating Model (epsilon)")
qqline(resid.cond.cla)

# Popular rating
lmer.popular <- lmer(Popular ~ Instrument + Harmony + Voice + ConsInstr +
PachListen + ClsListen + KnowRob + KnowAxis + CollegeMusic + Composing +
GuitarPlay + (1+Instrument|Subject), data = ratings)
fitLMER.fnc(lmer.popular, ran.effects=c("(Instrument|Subject)", "(Harmony|Subject)",
"(Voice|Subject)"), method="AIC")
display(lmer.popular)
texreg(lmer.popular)


# Conditional residual analysis for popular rating
resid.cond.pop <- r.cond(lmer.popular)

par(mfrow = c(1,1))
qqnorm(resid.cond.pop, main="Conditional Residuals for Popular
Rating Model (epsilon)")
qqline(resid.cond.pop)
```

	Coefficients of Fixed Effects
(Intercept)	3.94***
	(0.33)
Instrumentpiano	1.18**
	(0.36)
Instrumentstring	3.20***
	(0.41)
HarmonyI-V-IV	−0.24
	(0.31)
HarmonyI-V-VI	1.01**
	(0.36)
HarmonyIV-I-V	−0.26
	(0.31)
Voicepar3rd	−0.08
	(0.31)
Voicepar5th	−0.46
	(0.31)
KnowRob1	−0.26
	(0.62)
KnowRob5	0.32
	(0.53)
KnowAxis1	1.08
	(1.37)
KnowAxis5	0.40
	(0.47)
Instrumentpiano:HarmonyI-V-IV	0.65
	(0.43)
Instrumentstring:HarmonyI-V-IV	0.74
	(0.43)
Instrumentpiano:HarmonyI-V-VI	0.63
	(0.43)
Instrumentstring:HarmonyI-V-VI	0.27
	(0.43)
Instrumentpiano:HarmonyIV-I-V	0.22
	(0.43)
Instrumentstring:HarmonyIV-I-V	−0.06
	(0.43)
Instrumentpiano:Voicepar3rd	−0.30
	(0.43)
Instrumentstring:Voicepar3rd	−0.46
	(0.43)
Instrumentpiano:Voicepar5th	0.42
	(0.43)

	Coefficients of Fixed Effects
Instrumentstring:Voicepar5th	0.32
	(0.43)
HarmonyI-V-IV:Voicepar3rd	-0.02
	(0.43)
HarmonyI-V-VI:Voicepar3rd	-1.45^{***}
	(0.43)
HarmonyIV-I-V:Voicepar3rd	0.34
	(0.43)
HarmonyI-V-IV:Voicepar5th	0.36
	(0.43)
HarmonyI-V-VI:Voicepar5th	0.23
	(0.43)
HarmonyIV-I-V:Voicepar5th	0.62
	(0.43)
Instrumentpiano:HarmonyI-V-IV:Voicepar3rd	-0.57
	(0.60)
Instrumentstring:HarmonyI-V-IV:Voicepar3rd	-0.22
	(0.60)
Instrumentpiano:HarmonyI-V-VI:Voicepar3rd	1.02
	(0.61)
Instrumentstring:HarmonyI-V-VI:Voicepar3rd	1.55^{*}
	(0.60)
Instrumentpiano:HarmonyIV-I-V:Voicepar3rd	0.19
	(0.60)
Instrumentstring:HarmonyIV-I-V:Voicepar3rd	0.88
	(0.60)
Instrumentpiano:HarmonyI-V-IV:Voicepar5th	-0.58
	(0.61)
Instrumentstring:HarmonyI-V-IV:Voicepar5th	-1.18
	(0.60)
Instrumentpiano:HarmonyI-V-VI:Voicepar5th	-1.01
	(0.61)
Instrumentstring:HarmonyI-V-VI:Voicepar5th	-0.88
	(0.60)
Instrumentpiano:HarmonyIV-I-V:Voicepar5th	-0.38
	(0.60)
Instrumentstring:HarmonyIV-I-V:Voicepar5th	-0.82
	(0.60)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Mixed model on classical ratings

	Coefficients of Fixed Effects
(Intercept)	6.80***
	(0.28)
Instrumentpiano	−1.18**
	(0.36)
Instrumentstring	−2.96***
	(0.39)
HarmonyI-V-IV	0.32
	(0.32)
HarmonyI-V-VI	−0.18
	(0.34)
HarmonyIV-I-V	0.04
	(0.33)
Voicepar3rd	0.08
	(0.32)
Voicepar5th	0.30
	(0.32)
Instrumentpiano:HarmonyI-V-IV	−0.32
	(0.44)
Instrumentstring:HarmonyI-V-IV	−0.66
	(0.44)
Instrumentpiano:HarmonyI-V-VI	0.15
	(0.45)
Instrumentstring:HarmonyI-V-VI	−0.09
	(0.44)
Instrumentpiano:HarmonyIV-I-V	0.28
	(0.44)
Instrumentstring:HarmonyIV-I-V	0.38
	(0.44)
Instrumentpiano:Voicepar3rd	0.34
	(0.44)
Instrumentstring:Voicepar3rd	0.86
	(0.44)
Instrumentpiano:Voicepar5th	0.03
	(0.44)
Instrumentstring:Voicepar5th	0.18
	(0.44)
HarmonyI-V-IV:Voicepar3rd	−0.18
	(0.44)
HarmonyI-V-VI:Voicepar3rd	0.78
	(0.44)
HarmonyIV-I-V:Voicepar3rd	−0.16
	(0.44)
HarmonyI-V-IV:Voicepar5th	−0.50
	(0.44)

	Coefficients of Fixed Effects
HarmonyI-V-VI:Voicepar5th	-0.50
	(0.44)
HarmonyIV-I-V:Voicepar5th	-0.60
	(0.44)
Instrumentpiano:HarmonyI-V-IV:Voicepar3rd	0.18
	(0.63)
Instrumentstring:HarmonyI-V-IV:Voicepar3rd	0.00
	(0.63)
Instrumentpiano:HarmonyI-V-VI:Voicepar3rd	-1.34^{*}
	(0.63)
Instrumentstring:HarmonyI-V-VI:Voicepar3rd	-1.85^{**}
	(0.63)
Instrumentpiano:HarmonyIV-I-V:Voicepar3rd	-0.53
	(0.63)
Instrumentstring:HarmonyIV-I-V:Voicepar3rd	-1.74^{**}
	(0.63)
Instrumentpiano:HarmonyI-V-IV:Voicepar5th	0.59
	(0.63)
Instrumentstring:HarmonyI-V-IV:Voicepar5th	1.12
	(0.63)
Instrumentpiano:HarmonyI-V-VI:Voicepar5th	0.08
	(0.63)
Instrumentstring:HarmonyI-V-VI:Voicepar5th	0.63
	(0.63)
Instrumentpiano:HarmonyIV-I-V:Voicepar5th	0.21
	(0.63)
Instrumentstring:HarmonyIV-I-V:Voicepar5th	-0.16
	(0.63)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: Mixed model on popular ratings

	Coefficients of Fixed Effects
Instrumentguitar	4.50***
	(0.99)
Instrumentpiano	5.98***
	(1.00)
Instrumentstring	7.83***
	(1.00)
HarmonyI-V-IV	0.05
	(0.12)
HarmonyI-V-VI	0.98***
	(0.21)
HarmonyIV-I-V	0.10
	(0.11)
Voicepar3rd	−0.38***
	(0.11)
Voicepar5th	−0.31**
	(0.11)
dichotomize1	−0.07
	(1.68)
as.numeric(X16.minus.17)	−0.11
	(0.10)
ConsInstr0.67	0.38
	(1.55)
ConsInstr1	−1.90
	(1.19)
ConsInstr1.67	−0.59
	(1.41)
ConsInstr2	−1.06
	(1.38)
ConsInstr2.33	−0.47
	(1.21)
ConsInstr2.67	1.45
	(1.38)
ConsInstr3	−1.73
	(1.09)
ConsInstr3.33	0.51
	(1.89)
ConsInstr3.67	−2.04
	(1.15)
ConsInstr4	−1.34
	(1.66)
ConsInstr4.33	−0.89
	(1.20)
ConsInstr5	−1.54
	(1.14)

	Coefficients of Fixed Effects
PianoPlay1	1.11
	(0.66)
PianoPlay4	2.71**
	(1.00)
PianoPlay5	0.77
	(0.65)
dichotomize1:as.numeric(X16.minus.17)	−0.08
	(0.20)
dichotomize1:ConsInstr1	−0.04
	(1.67)
dichotomize1:ConsInstr1.67	−0.49
	(1.93)
dichotomize1:ConsInstr2.33	0.53
	(2.08)
dichotomize1:ConsInstr3	1.56
	(1.71)
dichotomize1:ConsInstr3.67	−0.30
	(2.07)
dichotomize1:ConsInstr4.33	−2.40
	(1.66)
dichotomize1:ConsInstr5	1.21
	(1.83)
dichotomize1:PianoPlay1	1.88
	(1.10)
dichotomize1:PianoPlay4	−1.04
	(1.49)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: Mixed model on musician vs non-musician