Examining The Influence of Instrument, Harmony, and Voice Leading on Listeners' Recognization between Popular and Classical Music

Na Jin
jinn@andrew.cmu.edu
02 December 2019

Abstract

Our study aims to examine the influence of instrument, harmony, and voice Leading on listeners' distinction between popular and classical Music. We examine data collected by Ivan Jimenez and student Vincent Rossi for 70 experimental subjects, using exploratory data analysis, anova and regression analyses and variable selection to analyze our dataset. We found that instrument, harmony and voice leading indeed have statistically significant influence on music genre recognization and among these three experimental factors, instrument has stronggest influence on both classical and popular ratings. Further examination suggests that harmony is the factor that distinguishes the musicians from non-musicians under classical genre. In addtion we found that variable X16.minus.17 has negative influence on classical ratings. Our analyses are limited by the dataset and interaction terms. In the future, regression analyses building on better experimental design, taking all categorical variables or at least three-way interaction terms into account may yield more robust results, if needed.

1. Introduction

Music is very complex but fascinating creature. Different combinations of instrument, harmony and voice leading can compose of different music genres that can touch human being's souls and arouse the listeners' emotional resonance from the depth of heart. However, distingushing between genres is subject to the personal perception of instrument, harmony and voice leading. Therefore, it arouses our interests to explore the influence of these musical components on listerners' distinction between music genres. The main purpose of this study is to examine which factor, instrument, harmony and voice leading exert the greatest influence on listerners' distinction between popular and classical music.

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

- 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?
- Aomng 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?

2. Method

The data for this study is collected by Ivan Jimenez and student Vincent Rossi, a composer and musicologist visiting the University of Pittsburgh, and student Vincent Rossi, in a designed experiment inteneded to measure the influence of instrument, harmonic motion, and voice leading on listeners' identification of music as "classical" or "popular". They presented 36 musical stimuli to 70 listeners, recruited from the population of undergraduates at the University of Pittsburgh, and asked the listeners to rate the music on two different scales:

Table 1: Variables in Original Dataset

variable	Explanation
X	Observation ID
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)
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?

- 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).

Listeners were told that a piece could be rated as both classical and popular, neither classical nor popular, or mostly classical and not popular (or vice versa), so that the scales should have functioned more or less independently. The 36 stimuli were chosen by completely crossing these 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

Hence, this dataset contains 2520 observations with 3 main experimental factors and 25 other variables of interests in addition. The definitions of the variables are given in Table 1.

Our analyses relied on the following approaches: data cleaning and transformation[1], exploratory data analysis by boxplots, ANOVA analysis, regression analysis and automtic backward model selection[2]. After cleaning the initial data, we ended up with 1798 observations and 24 variables in total. By looking at the graphical visualizations, we could examine the relationship between variables pair-by-pair, which also acted as the pre-screening process of variables before the modeling process. The regression analyses combined with analysis of variance(ANOVA), if needed, allows us to quantatively investigate the importance of the relationship between variables of our interests. The automatic variable selection suggests us the best combination of variables with the consideration of model complexity and minimum residual errors. Logarithm transformation, if needed, was applied during our study to correct the skewness of variables without harming the interpretabilty of our model.

[1]Please refer to Appendix Section 1 for full details
[2]Please refer to Appendix Section 4 for full details

3. Results

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

First of all, we looked at the boxplots in Figure 1 and 2 from exploratory data analysis to explore if different levels of these three variables will yield different classical or popular ratings. There appears to be some strong influence of instrument on classical ratings and popular ratings since the medians of instrument vary obviously in Figure 1 and 2. In Figure 1 and 2, we noticed that medians are slightly different across different levels of harmony and voice leading, which indicates that harmony and voice may have some association with classical and popular ratings. Based on our first glancing exploration, we believed that it is worthy for us to further investigate the influence of instrument, harmony and voice leading on classical and popular ratings in a more quantative way.

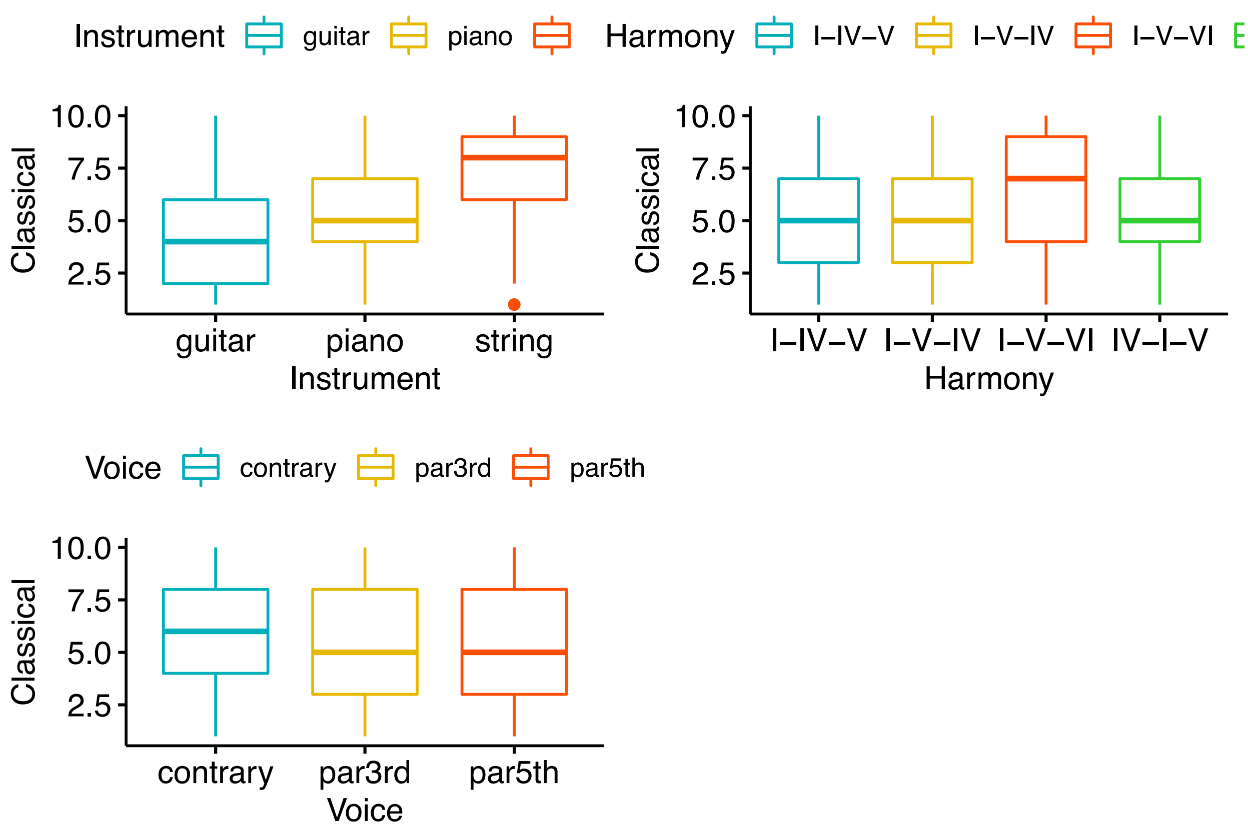

Figure 1: Boxplots for Classical Ratings

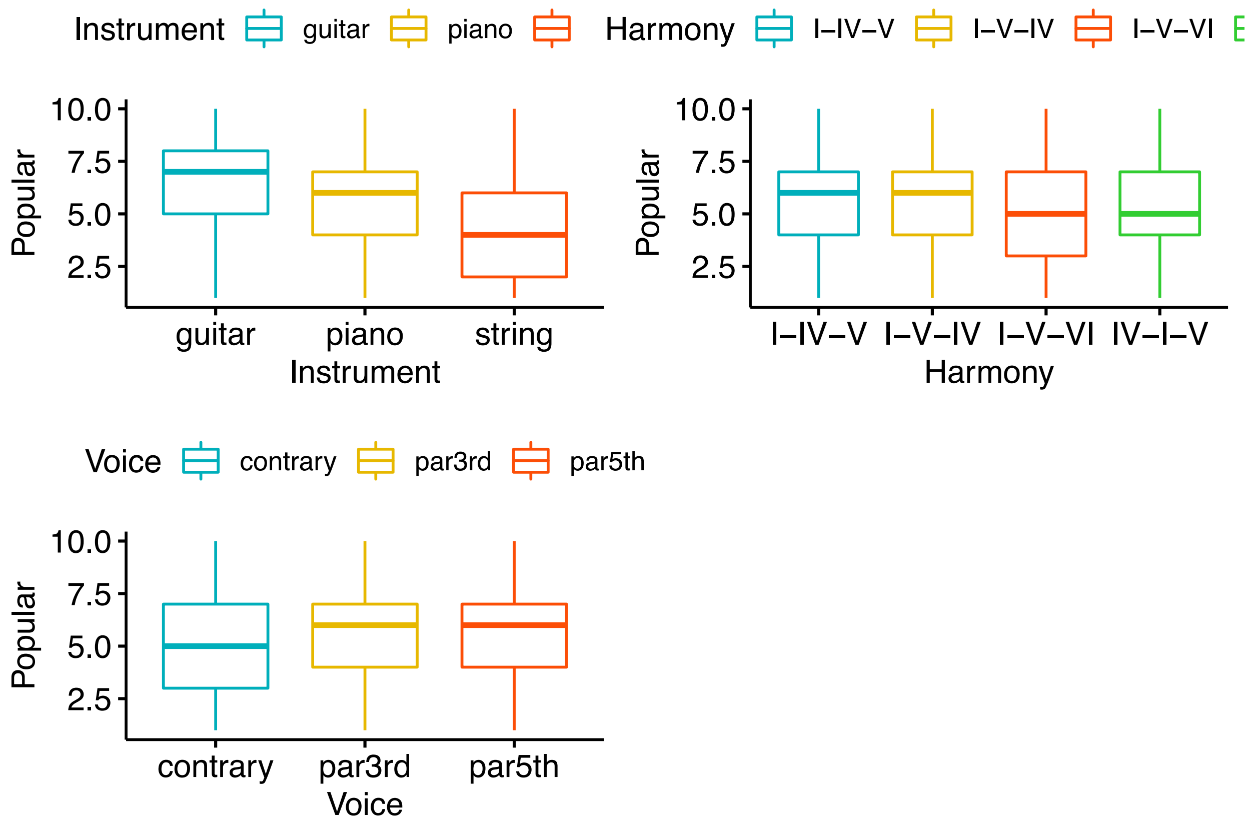

Figure 2: Boxplots for Popular Ratings

To quantatively measure the influence of instrument, harmony and voice leading on classical and popular ratings respectively, two analysis of variance(ANOVA) models, one for classical ratings and the other one for popular ratings were applied. We also included the two-way and three-way interaction terms of these three experimental variables in our ANOVA models because we assumed that different mixes of instrument, harmony and voice leading would influence people's perception of the music and thus may influence how people rate for music genres. The hypothesis to be tested in the ANOVA models is if there is any significant difference between the average ratings of classical/popular in different levels of these three experimental factors. The ANOVA results are given in Table 2 and 3.

Table 2: ANOVA of Multiple Linear Model for Classical

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Instrument	2	2889	1444	291.5	4.16e-110
Harmony	3	281.8	93.95	18.96	4.191e-12
Voice	2	58.8	29.4	5.934	0.002701
Instrument:Harmony	6	21.77	3.628	0.7323	0.6236
Instrument:Voice	4	18.69	4.673	0.9432	0.4379
Harmony:Voice	6	76.11	12.69	2.56	0.01795
Instrument:Harmony:Voice	12	73.11	6.093	1.23	0.2561
Residuals	1762	8729	4.954	NA	NA

Table 4: ANOVA Analysis of Multilevel Linear Model for Classical Ratings

	Df	Sum Sq	Mean Sq	F value
Instrument	2	455.03655	227.51827	93.030552
Harmony	3	48.52873	16.17624	6.614347
Voice	2	58.89211	29.44605	12.040275
Harmony:Voice	6	73.67186	12.27864	5.020647

Table 5: ANOVA of Multilevel Linear Model for Popular Ratings

	Df	Sum Sq	Mean Sq	F value
Instrument	2	308.69071	154.345357	63.100738
Harmony	3	28.33269	9.444228	3.861067
Voice	2	29.48096	14.740481	6.026325

Table 3: ANOVA of Multiple Linear Model for Popular

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Instrument	2	2121	1060	224	2.128e-87
Harmony	3	60.62	20.21	4.268	0.005182
Voice	2	29.65	14.82	3.131	0.0439
Instrument:Harmony	6	13.45	2.241	0.4734	0.8285
Instrument:Voice	4	14.4	3.601	0.7606	0.5509
Harmony:Voice	6	36.56	6.093	1.287	0.2598
Instrument:Harmony:Voice	12	60.18	5.015	1.059	0.391
Residuals	1762	8341	4.734	NA	NA

Table 2 and 3 showed that instrument, harmony and voice leading have statistically significant influence on both classical and popular ratings and the effect of harmony on classical ratings is different for different types of voice leading since the p-values of these variables are smaller than 5% significant level. Combined ANOVA outputs(Table 2 and 3) with boxplots(Figure 1 and 2), we could examine that instrument has the strongest influence on both classical and popular ratings.

However, these two ANOVA models for classical and popular ratings violate the heteroscadecity assumption[3]. The heteroscadecity reminds us of the existence of individual variability or personal biases in ratings. Thus we switched to multilevel linear model[4] which takes individual variability into consideration to take care of this heteroscadecity issue in our ANOVA model. From the ANOVA of multilevel linear model(Table 4 and 5), we reached the same conclusion examined from our initial ANOVA output. In Table 4 and 5, we observed that instrument has the largest mean square of residuals among all these three experimental factors. This indicates that instrument can account for the largest portion of variability of classical and popular ratings and thus it has the largest influence on classical and popular ratings.

From boxplots(Figure 1 and 2), we observed that harmony I-V-VI and contrary voice leading may have largest influence on classical ratings while having smallest influence on popular ratings.This finding can be confirmed by comparing coefficients of harmony and voice leading given in Table 6. From our conclusion, this seems that it does matter whether the respondent is familiar with one or the other (or both) of the Pachelbel rants/comedy bits. Harmonic progression, I-V-VI is the beginning progression for Pachelbel's Canon in D, which many people have heard and Table 6 shows that harmonic progression I-V-VI has been rated as more classical-sounding than I-IV-V and IV-I-V. However, Ivan Jimenez and Vincent Rossi found that among 22044 classical master pieces of progression, harmonic progression I-V-VI appear less frequently than I-IV-V and IV-I-V. This indicates that the familarity with Pachelbel's Canon in D played important

[3]Please refer to Appendix Section 2 for ANOVA assumption checking
[4]Please refer to Appendix Section 3 for full details

Table 6: Factor Coefficients of Harmony and Voice of Classical Ratings

Variable	Coefficient
Classical ~ Harmony-1 + Instrument + Voice+Harmony:Voice + (1 + Instrument + Harmony|Subject)	
HarmonyI-IV-V	4.03959
HarmonyI-V-IV	4.28922
HarmonyI-V-VI	5.45488
HarmonyIV-I-V	3.95460
Classical ~ Voice-1 + Instrument + Harmony+Harmony:Voice + (1 + Instrument + Harmony|Subject)	
Voicecontrary	4.03959
Voicepar3rd	3.83028
Voicepar5th	3.87227
Popular ~ Harmony-1 + Instrument + Voice + (1 + Instrument + Harmony|Subject)	
HarmonyI-IV-V	6.63883
HarmonyI-V-IV	6.55534
HarmonyI-V-VI	6.17920
HarmonyIV-I-V	6.32718
Popular ~ Voice-1 + Instrument + Harmony + (1 + Instrument + Harmony|Subject)	
Voicecontrary	6.63878
Voicepar3rd	6.88304
Voicepar5th	6.93181

Note: The reason why variables of interests minus 1 in the model is to give us the means of each levels of the variable

role on people's ratings for classical music.

In summary, from the outputs of multilevel linear models, we found that instrument played the most important role among all three experimental factors on people's recognition of classical and popular music. Harmony I-V-VI and contrary voice leading have largest influence on classical ratings while having smallest influence on popular ratings. Besides, the reason why harmonic regression I-V-VI has been frequently recognized as classical music maybe due to people's familarity of Pachelbel's Canon in D since I-V-VI is the beginning progression of Pachelbel's Canon in D.

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

To examine whether there is any difference in factors that influence classical and popular ratings, we included other individual covariates in the multilevel linear mdoels and conducted the backward variable selections for fixed effects and random effects. We obeyed the following rules to select our final models for both classical[5] and popular ratings[6]:

I. Best reflects the knowledge of music and the meaning of the variables

II.Best satisfies modeling assumptions

III.Is most clearly indicated by the data

IV.Can be explained to someone who is more interested in musical factors than in mathematics and statistics

We ended up with two different models accounting for classical and popular ratings respectively. Both models are not standard repeated measures model because our final models for these two genres contain other random effects, Instrument and Harmony. The process of how we obtained these final models are presented in Appendix Section 4. The random intercepts in both repeated measures models for popular and classical genera can account for "personal biases" in ratings, that is: what is personal inclination of recognizing one stimulus as classical against popular and vice versa. By including the other random effects,

[5]R code: lmer(Classical ~ Harmony + Instrument + Voice + X16.minus.17 + (1 + Harmony + Instrument | Subject) + Harmony:Voice, data, REML = FALSE, control = lmerControl(optimizer = "bobyqa"))

[6]R code: lmer(Popular ~ Harmony + Instrument + Voice + (1 + Harmony + Instrument| Subject), data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

these two models give us hints about how the perception of harmony and instrument for each genre differ in the individual levels. Larger variance components of random effects account for the greater individual difference and smaller one implies that there is more homogeneity in the individual levels for genre recognition under specific explanatory variables.

The outputs of final models are shown in Table 7 and 8. As we examined before, instrument, harmony and voice played an important role on both classical and popular ratings.On the other hand, from our final models, we found that the interaction between harmony and voice and X16.minus.17 have statistically significant influence on classical ratings but they are not important factors of influencing popular ratings. As expected, Table 7 shows that compared with harmony I-IV-V, stimuli with Harmony I-V-IV or Harmony I-V-VI would increase classical ratings but decrease popular ratings; compared with guitar, stimuli played by piano or string would be more likely to be perceived as classical than popular; parallel 3rd and parallel 5th would be more popular-sounding to the audience than contrary voice leading.

Interestingly, Table 7 indicates that other than three main experimental factors, X16.minus.17 played an important role on influencing classical ratings. X16.minus.17 is the auxiliary measure of listener's ability to distinguish classical vs popular music. Our results represent that the larger difference between X16 and X17, the less classical ratings the experimental subject would assign to the stimuli they heard. This interesting finding may deserve further investigation.

To sum up, different instrument, harmony and voice have different effects on classical and popular ratings. The audience tended to be more likely to recognize stimuli with Harmony I-V-IV or Harmony I-V-VI as classical than popular music. Piano and string are more classical-sounding. Parallel 3rd and parallel 5th sound more popular to the audience compared with contrary voice leading on average. One interesting finding is that the difference between X16.minus.17 would have negative influence on classical ratings, which may arouse the interest of future investigation.

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

To quantatively measure whether there are differences in the way that muscians and non-musicians identify classical music, we dichotomized the Selfdeclare rather than the OMSI score. The reason is that there are less than 18% of subjects having OMSI scores greater than 500[7]. Although OMSI scores are more objective measure of classifying musician and non-musician, due to its unbalanced sizes of musician and non-musician, we decided to dichotomize on the Selfdeclare, the self-report measure. We applied two ways to dichotomize selfdeclare, cutting off selfdeclare on 2 and 3 seperately in order to test the sensitivity of our results.We added the dichotomized selfdeclare to the final models mentioned above and the interaction terms between the dichotomized selfdeclare and other important predictors in the final models as well.The ANOVA outputs are given in Table 9 and 10. From Table 9 and 10, it can be observed that the interaction between harmony and selfdeclare is statistically significant under two models but with different significance level, one with 10% significance level and the other one with 5% significance level. We shall ignore such trivial significance difference since the interaction between harmony and selfdeclare has decent significance in both ANOVA outputs. Thus, we can conclude that no matter how we dichotomized the selfdeclare, different perception of harmony can distinguish a musician and non-musician when rating the classical music.

4. Discussion

Our study examines the influence of three main experiment factors, Instrument, Voice, and Harmony on classical and popular ratings separately. From our initial examination, we found that as we expected, instrument was more influential than harmonic motion and voice leading for listeners' identification of musical genre regardless of the extent of listeners' musical training. According to our results, listeners' identification of genre was not associated with the presence or absence of harmonic retrogressions or parallel fifths but to other harmonic and voice leading features. We found that different combination of harmony and voice leading of a stimulus might influence listeners' identification of classical genre.

For classical genre identification, we found that there is another individual covariate, X16.minus.17, has great impact. X16.minus.17 is defined as auxiliary measure of listener's ability to distinguish classical vs popular music. However, for our current information it is difficult to know what X16 and 17 represent. Back

[7]OMSI greater than 500 will be classified as musician, source: http://marcs-survey.uws.edu.au/OMSI/omsi.php

Table 7: Model Output of Fixed Effects of Final Models For Classical and Popular Ratings

	Dependent variable:	
	Classical	Popular
	(1)	(2)
HarmonyI-V-IV	0.251	−0.083
	(0.185)	(0.118)
HarmonyI-V-VI	1.416***	−0.460***
	(0.263)	(0.173)
HarmonyIV-I-V	−0.085	−0.312**
	(0.183)	(0.131)
Instrumentpiano	1.367***	−0.979***
	(0.194)	(0.185)
Instrumentstring	3.127***	−2.677***
	(0.279)	(0.264)
Voicepar3rd	−0.209	0.244***
	(0.180)	(0.090)
Voicepar5th	−0.167	0.293***
	(0.181)	(0.090)
X16.minus.17	−0.105**	
	(0.051)	
HarmonyI-V-IV:Voicepar3rd	−0.459*	
	(0.256)	
HarmonyI-V-VI:Voicepar3rd	−0.824***	
	(0.256)	
HarmonyIV-I-V:Voicepar3rd	0.475*	
	(0.256)	
HarmonyI-V-IV:Voicepar5th	−0.240	
	(0.256)	
HarmonyI-V-VI:Voicepar5th	−0.598**	
	(0.256)	
HarmonyIV-I-V:Voicepar5th	0.092	
	(0.255)	
Constant	4.221***	6.639***
	(0.247)	(0.196)
Observations	1,798	1,798
Log Likelihood	−3,561.689	−3,561.178
Akaike Inf. Crit.	7,197.379	7,182.357
Bayesian Inf. Crit.	7,400.673	7,347.190
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 8: Random Effects of Final Models of Classical and Popular Ratings

variable	standard_deviation
Classical Ratings	
Intercept	1.31508
HarmonyI-V-IV	0.27561
HarmonyI-V-VI	1.35747
HarmonyIV-I-V	0.23787
Instrumentpiano	1.22513
Instrumentstring	1.88264
Popular Ratings	
Intercept	1.18471
HarmonyI-V-IV	0.38897
HarmonyI-V-VI	0.97974
HarmonyIV-I-V	0.56460
Instrumentpiano	1.15370
Instrumentstring	1.76861

Table 9: ANOVA of Multilevel Model Dichotomizing Selfdeclare on 2

	Chisq	Df	Pr(>Chisq)
Harmony	22.5310068	3	0.0000506
Instrument	126.0017843	2	0.0000000
Voice	24.0213043	2	0.0000061
Selfdeclare	1.1836520	1	0.2766138
X16.minus.17	5.1942130	1	0.0226622
Harmony:Selfdeclare	6.8590288	3	0.0765294
Instrument:Selfdeclare	0.3401148	2	0.8436164
Voice:Selfdeclare	0.4693409	2	0.7908314
Selfdeclare:X16.minus.17	0.8202511	1	0.3651067
Harmony:Voice	30.0286759	6	0.0000388

Table 10: ANOVA of Multilevel Model Dichotomizing Selfdeclare on 3

	Chisq	Df	Pr(>Chisq)
Harmony	23.4887704	3	0.0000319
Instrument	128.3811742	2	0.0000000
Voice	24.0387579	2	0.0000060
Selfdeclare	1.1838887	1	0.2765657
X16.minus.17	4.6360955	1	0.0313062
Harmony:Selfdeclare	10.6192253	3	0.0139736
Instrument:Selfdeclare	1.3391646	2	0.5119224
Voice:Selfdeclare	0.1482353	2	0.9285624
Selfdeclare:X16.minus.17	0.8722364	1	0.3503369
Harmony:Voice	30.1024068	6	0.0000376

to our model, we found that this individual covariate has negative relationship with the classical ratings. This means that the greater the difference between X16 and 17, the stimuli will be less likely to be recognized as classical. This finding may arouse someone's interests and maybe worthy to do further investigation on this individual covariate.

Besides, we also examined how the main effects influence the musician and non-musician. Through our study, we found that harmony was more influential for classical ratings of musicians than ratings of non-musicians on genre identification. This finding is not sensitive to how we dichotomized the data for classical genre.

This analyses was based on the multivariate analyses of the main effects of variables, and as such could not readily consider the interaction of three or more variables influencing classical and popular ratings. In addition, to simplify our analyses, in addition to instrument, harmony and voice leading, we only took binary variables, CollegeMusic and APTheory, as categorical variables and all other variables were treated as continuous variables. Thus our models lack of considering all possible categorical effects. Future analyses should at least consider examining the influence of all categorical variables on classical and popular ratings and/or including three-way interactions in the models to see if there is any improvement.

Our study was also limited by the use of dataset collected by Ivan Jimenez and student Vincent Rossi. There are many self-reported variables, such as APTheory and Selfdeclare, which makes this dataset lack of objectivity. Besides, this dataset also lack of representativeness. As we mentioned before, less than 18% of subjects have OMSI scores greater than 500, which means that our model building off this dataset cannot generalize how musicians and non-musicians differ in recognizing the music genre. Thus, for further investigation, a better experimental design is needed and the experimental design should be at least objective and representative.

In summary, keeping the caveats of the last two paragraphs in mind, our results are limited to this specific dataset and cannot be generalized to other conditions. For this specific experiment, our models and analyses may provide valuable insights about how main experimental factors influenced the recognization of music genre. In general, instrument, harmony and voice leading indeed had influence on the recognization of music genre and musicians and harmonic progression is the factor that distinguishes musicians from non-musicians under classical genre.

Appendix

1. Data Cleaning and Transformation

1.1 Data Cleaning

When we looked at the original data, we found that two variables, “X” and “first12” are not of our research interests so we deleted those two variables from our dataset. Besides, when we checked the value of each variable of our data, we found that some of the variables have missing values and those variables are shown in **Table 1**.

Table 1: Variables with Missing Values

variables	types	missing_count	missing_percent
X2ndInstr	integer	2196	87.142857
X1stInstr	integer	1512	60.000000
ConsNotes	integer	360	14.285714
KnowAxis	integer	288	11.428571
NoClass	integer	288	11.428571
APTheory	integer	216	8.571429
KnowRob	integer	180	7.142857
X1990s2000s.minus.1960s1970s	integer	180	7.142857
X1990s2000s	integer	144	5.714286
CollegeMusic	integer	108	4.285714
PachListen	integer	72	2.857143
Composing	integer	72	2.857143
ClsListen	integer	36	1.428571
Classical	numeric	27	1.071429
Popular	numeric	27	1.071429

Corlors in Table 1 represent the severity of missingness issue and the severity is defined as below along with the strategy to deal with each missingness issue:

- Severe missingness(Red, top 2 rows in Table 1): Missing_percent > 59.0%
  - Strategy: Deleting the variables from the dataset
- Moderate missingness(Orange, middle 3 rows in Table 1): 10.0% < Missing_percent < 20.0%
  - Strategy: Replacing the missing values with the variable mode(The shape of these variables before and after imputation is shown in Figure 1)

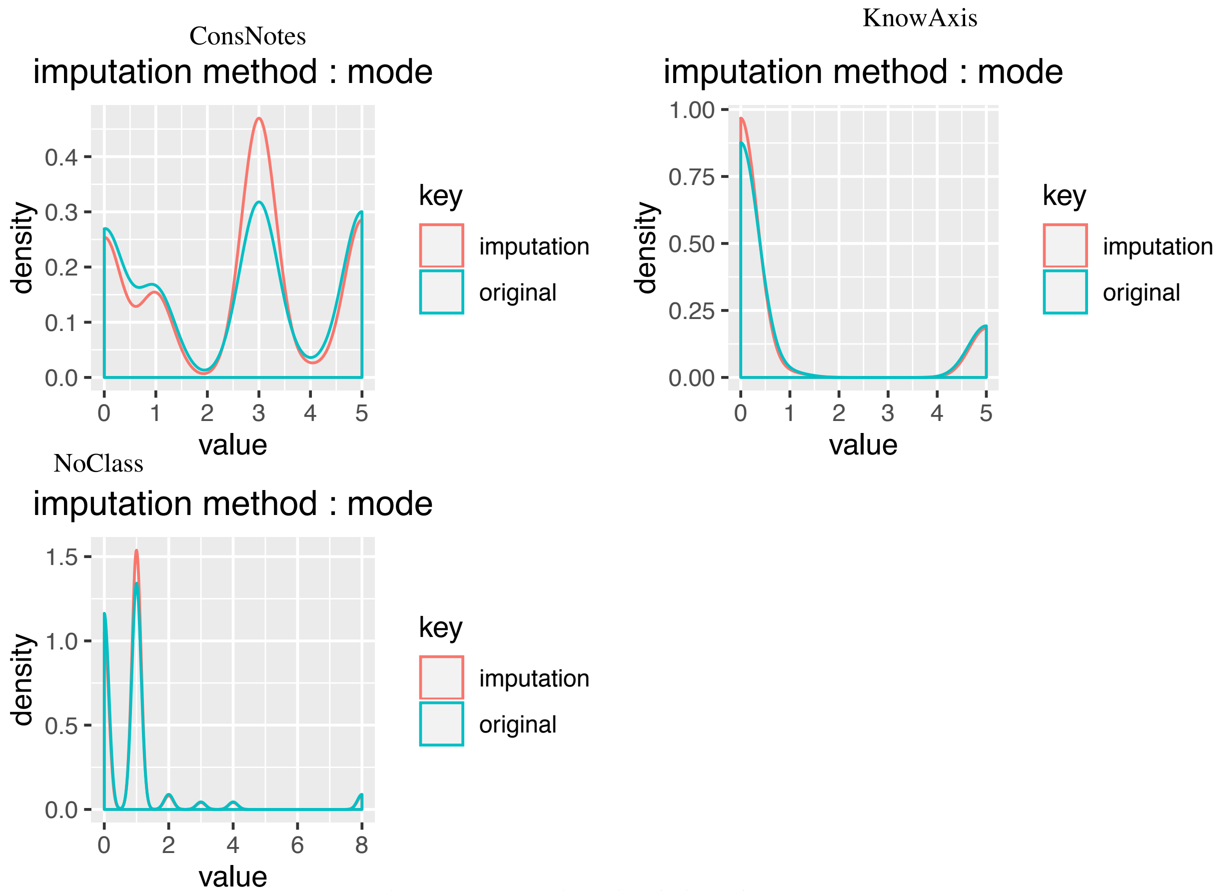

Figure 1: Imputation Simulation Plots

- Trivial missingness(Green, bottom 10 rows in Table 1): Missing_percent $< 10.0\%$
  - Strategy: Omitting the missing values in this variable

After the imputation, we reconstruted Instr.minus.Notes by using ConsInstr minus imputed ConsNotes.We cleaned the invalid data afterwards, which is recorded in Table 2.

Table 2: Clean-up Records

Variable	Clean_up
Classical	1. Filtered out 0s and decimal values 2. Corrected one observation from 19 to 10
Popular	Filtered out 0s and decimal values
X16.minus.17	Filtered out decimal values

[1] We assumed Classical ratings should be integer
[2] We assumed this is typing error because 9 and 0 are very close in the keyboard
[3] We assumed Popular ratings should be integer

1.2 Data Transformation

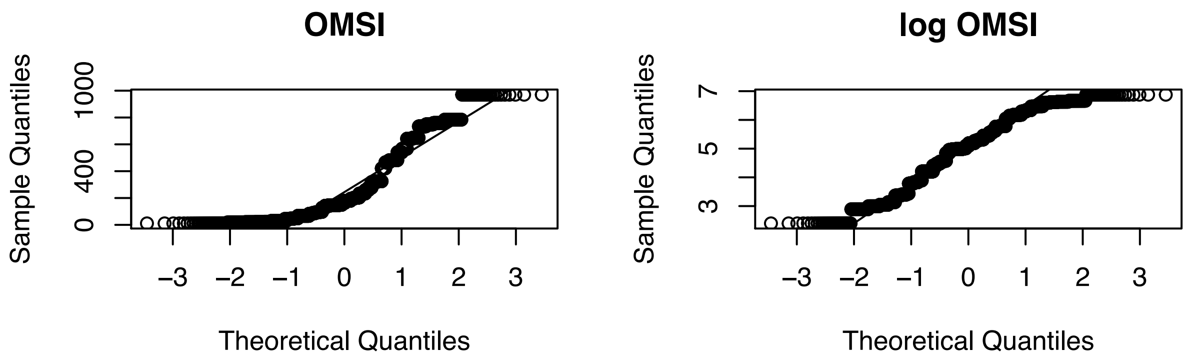

Figure 2.: OMSI Before and After log Transformation

2. Heteroscedasticity Detection of ANOVA of Mutiple Linear Models for Classical and Popular Ratings

Null Hypothesis for Table 3 and 5: ANOVA model has constant variance Alternative Hypothesis for Table 3 and 5: ANOVA model does not have constant variance

According to the results, the p-values in both Table 3 and 5 are significantly less than 5% significance level and thus the null hypothesis is rejected. This means ANOVA models for both classical and popular ratings violate the constant variance assumption.

Table 3: Levene Test for ANOVA of Classical

	Df	F value	Pr(>F)
group	35	2.005	0.0004652
	1762	NA	NA

Table 4: Normality Test for ANOVA of Classical

Test statistic	P value
0.9965	0.0004399 * * *

Null Hypothesis for Table 4 and 6: ANOVA model is not normally distributed Alternative Hypothesis for Table 4 and 6: ANOVA model is normally distributed

According to the results, the p-values in both Table 4 and 6 are significantly less than 5% signficance level and thus the null hypothesis is rejected. This means ANOVA models for both classical and popular ratings obey the normality assumption.

Table 5: Levene Test for ANOVA of POpular

	Df	F value	Pr(>F)
group	35	2.005	0.0004652
	1762	NA	NA

Table 6: Normality Test for ANOVA of Popular

Test statistic	P value
0.9955	3.138e-05 * * *

3. Choose Best Hierarchical Linear Model for Three Main Experimental Factors

For Classical

```
library(RLRsim)
library(lme4)

m0 <- lm(Classical ~ Instrument + Harmony*Voice, data = no_naratings)
lmer.b0 <- lmer(Classical ~ 1 + Instrument + Harmony*Voice + (1|Subject),
                data=no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))

######################## method: Anova ##############################
# Ho: there is no random intercept
# Ha: there should be random intercept
anova(lmer.b0,m0) # reject Ho

#Select Best Random Effects By Hand
lmer2 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Instrument + Harmony + Voice|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer3 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Instrument + Harmony|Subject),
              no_naratings,  REML=F, control = lmerControl(optimizer = 'bobyqa'))
              # smallest AIC and BIC and thus the best model

lmer4 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Instrument + Voice|Subject),
              no_naratings,  REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer5 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Harmony + Voice|Subject),
              no_naratings,  REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer6 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Instrument|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer7 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Harmony|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer8 <- lmer(Classical ~  1 + Instrument + Harmony*Voice + (1 + Voice|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
```

```r
anova(lmer.b0,lmer2,lmer3,lmer4,lmer5,lmer6,lmer7,lmer8)
```

For Popular

```r
m0 <- lm(Popular ~ Instrument + Harmony + Voice, data = no_naratings)
lmer.b0 <- lmer(Popular ~ 1 + Instrument + Harmony + Voice + (1|Subject),
                data=no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))

######################## method: Anova ##############################
# Ho: there is no random intercept
# Ha: there should be random intercept
anova(lmer.b0,m0) # reject Ho

#Select Best Random Effects By Hand

lmer2 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Instrument + Harmony + Voice|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer3 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Instrument + Harmony|Subject),
              no_naratings,  REML=F, control = lmerControl(optimizer = 'bobyqa'))
              # smallest AIC and BIC and thus the best model

lmer4 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Instrument + Voice|Subject),
              no_naratings,  REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer5 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Harmony + Voice|Subject),
              no_naratings,  REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer6 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Instrument|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer7 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Harmony|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))
lmer8 <- lmer(Popular ~  1 + Instrument + Harmony + Voice + (1 + Voice|Subject),
              no_naratings, REML=F, control = lmerControl(optimizer = 'bobyqa'))

anova(lmer.b0,lmer2,lmer3,lmer4,lmer5,lmer6,lmer7,lmer8)
```

3.1 Assumption Check for Classical ~ 1 + Instrument + Harmony*Voice + (1 + Instrument + Harmony|Subject)

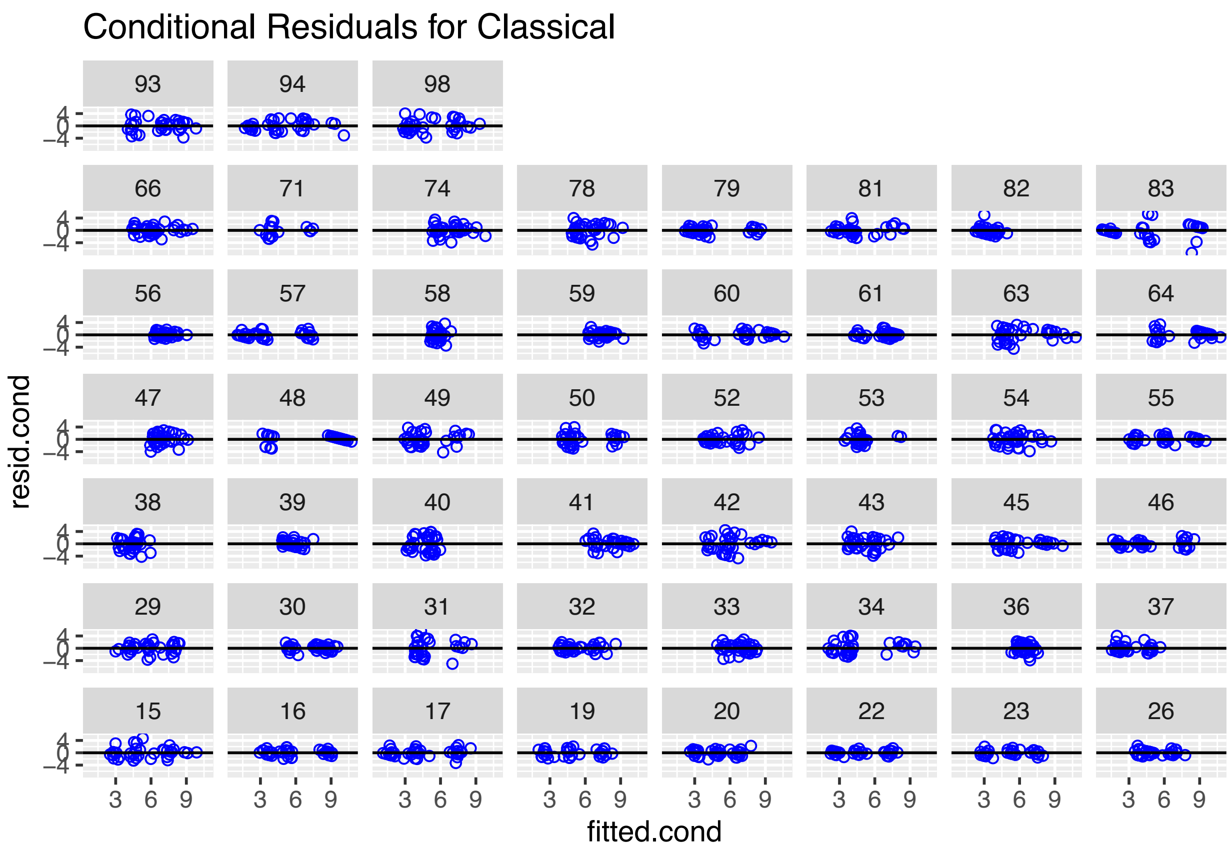

Figure 3: Conditional Residuals for Classical

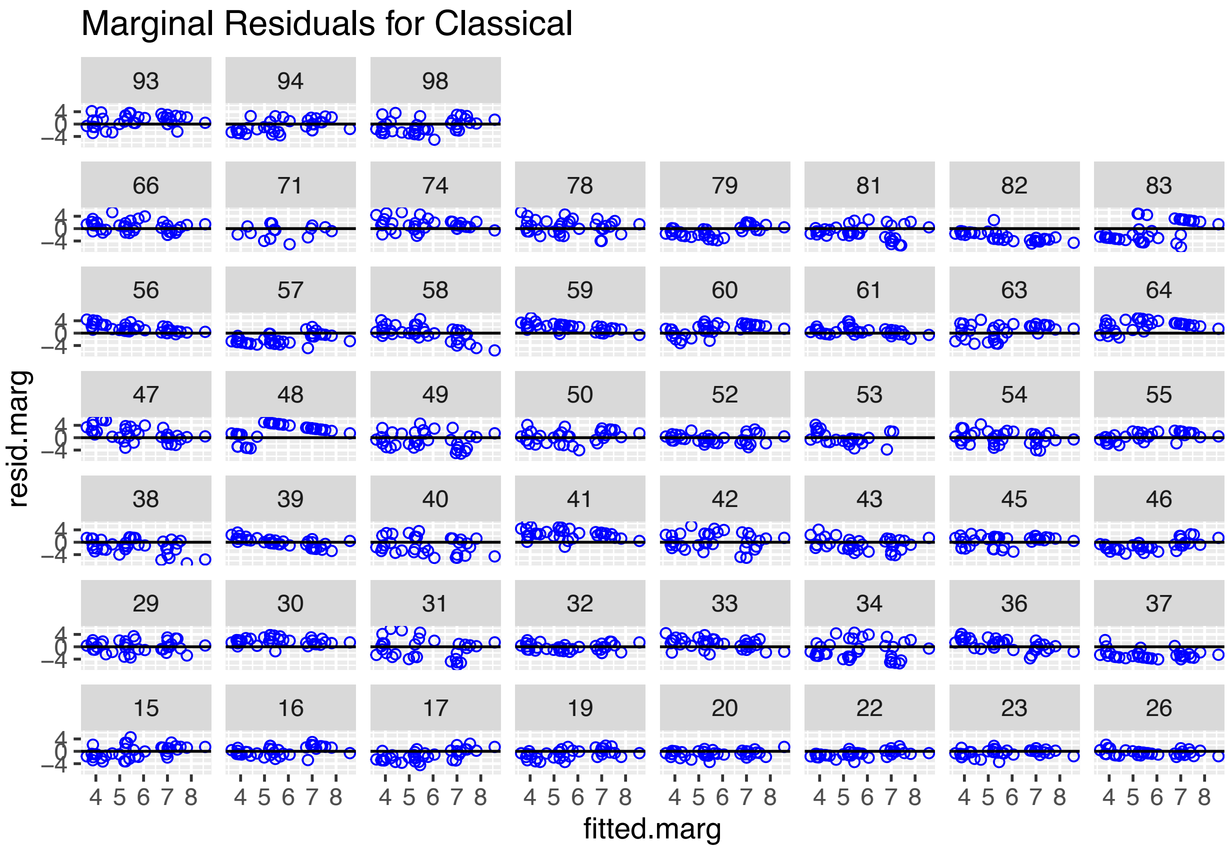

Figure 4: Marginal Residuals for Classical

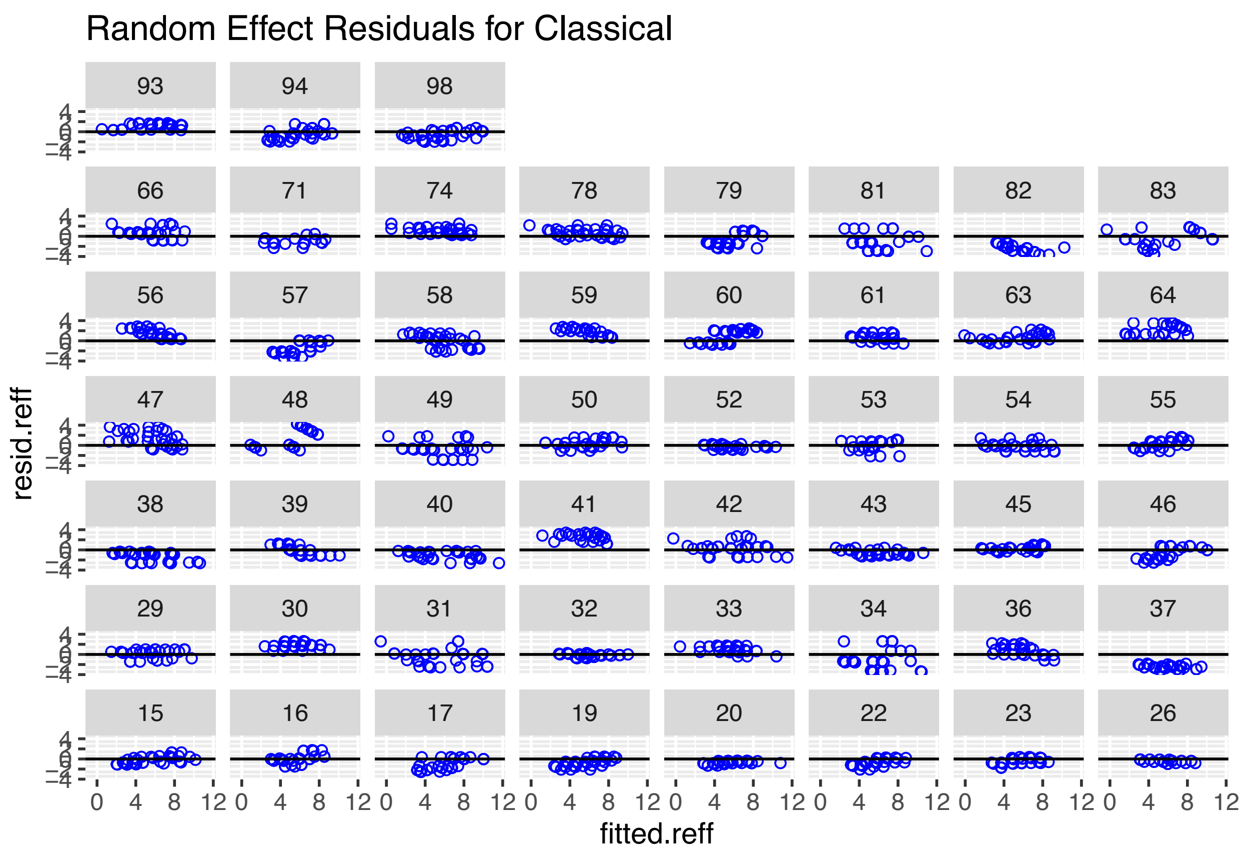

Figure 5: Random Effects Residuals for Classical

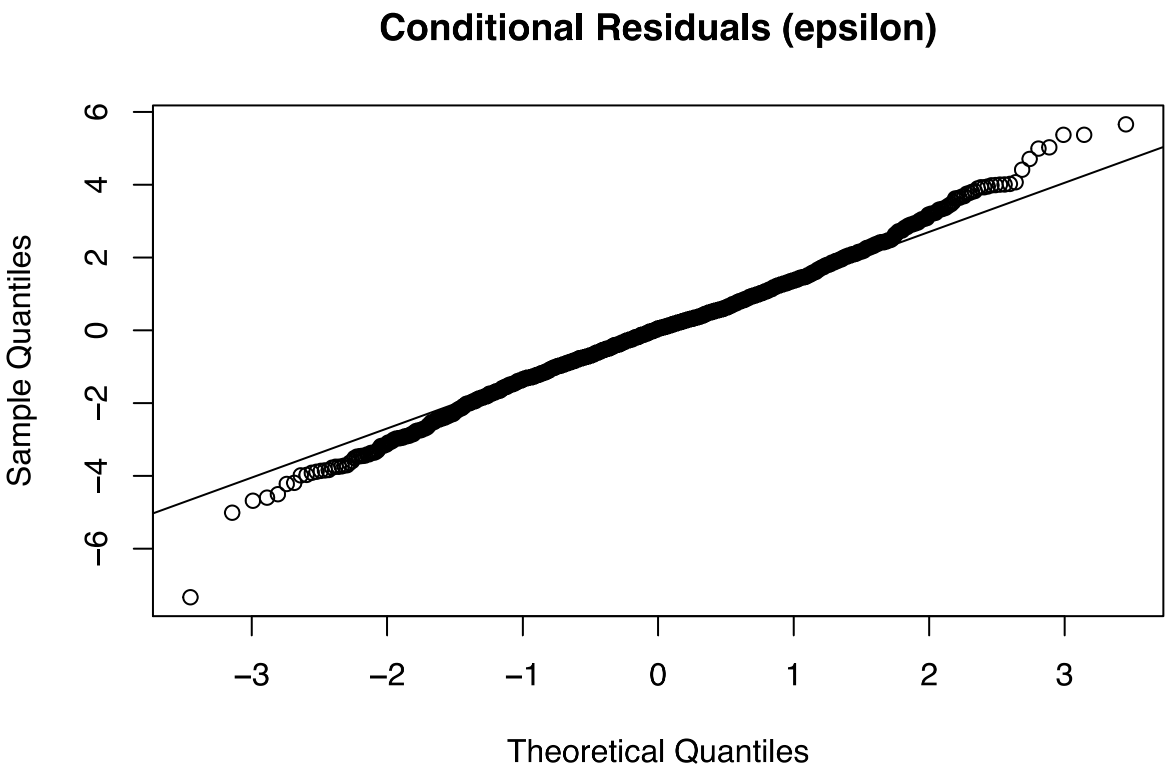

Figure 6: Q-Q plot of Conditional Residuals for Classical

All these figures look good since there is no specific pattern appear in Figure 3,4,5 and almost all points lie on the diagnoal line in Figure 6. Thus we can conclude that the multilevel model we tested for classical model is a valid model.

3.2 Assumption Check for Popular ~ 1 + Instrument + Harmony + Voice + (1 + Instrument + Harmony|Subject)

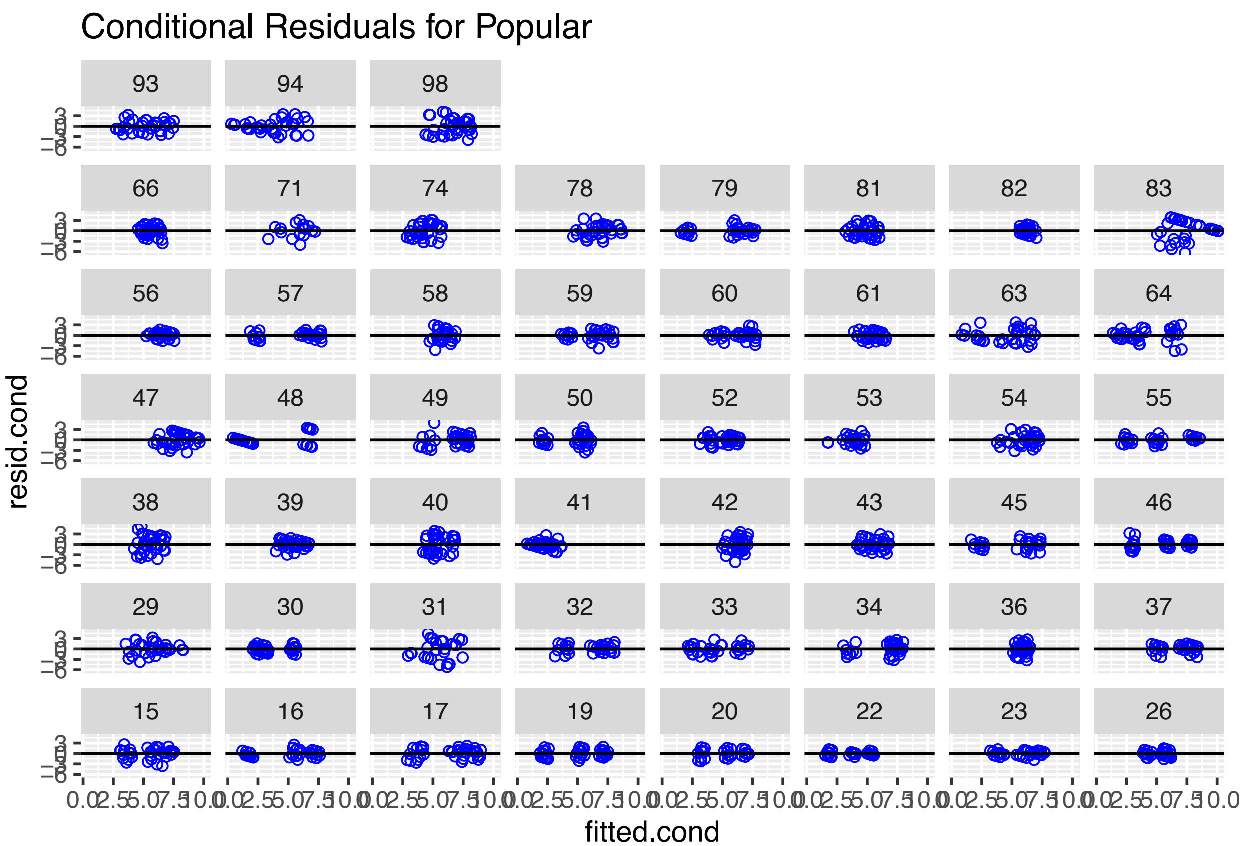

Figure 7: Conditional Residuals for Popular

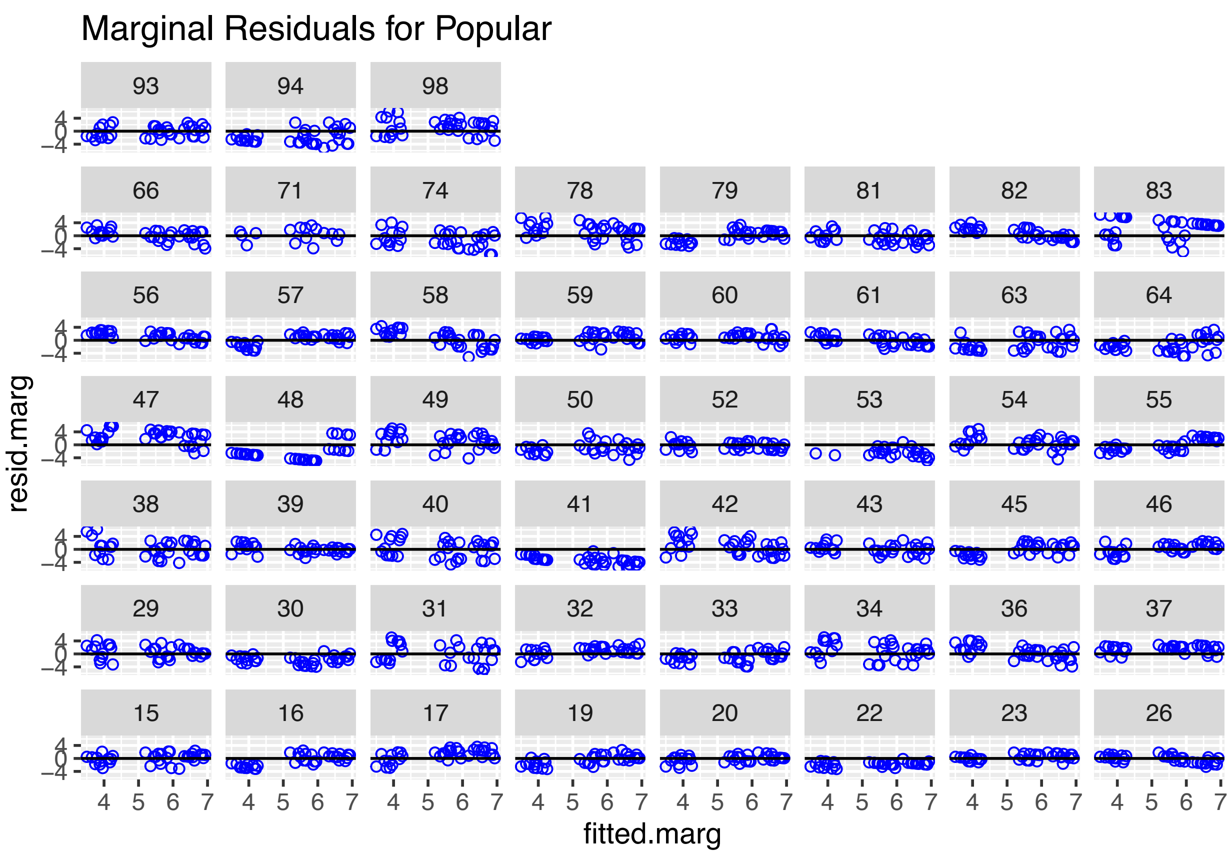

Figure 8: Marginal Residuals for Popular

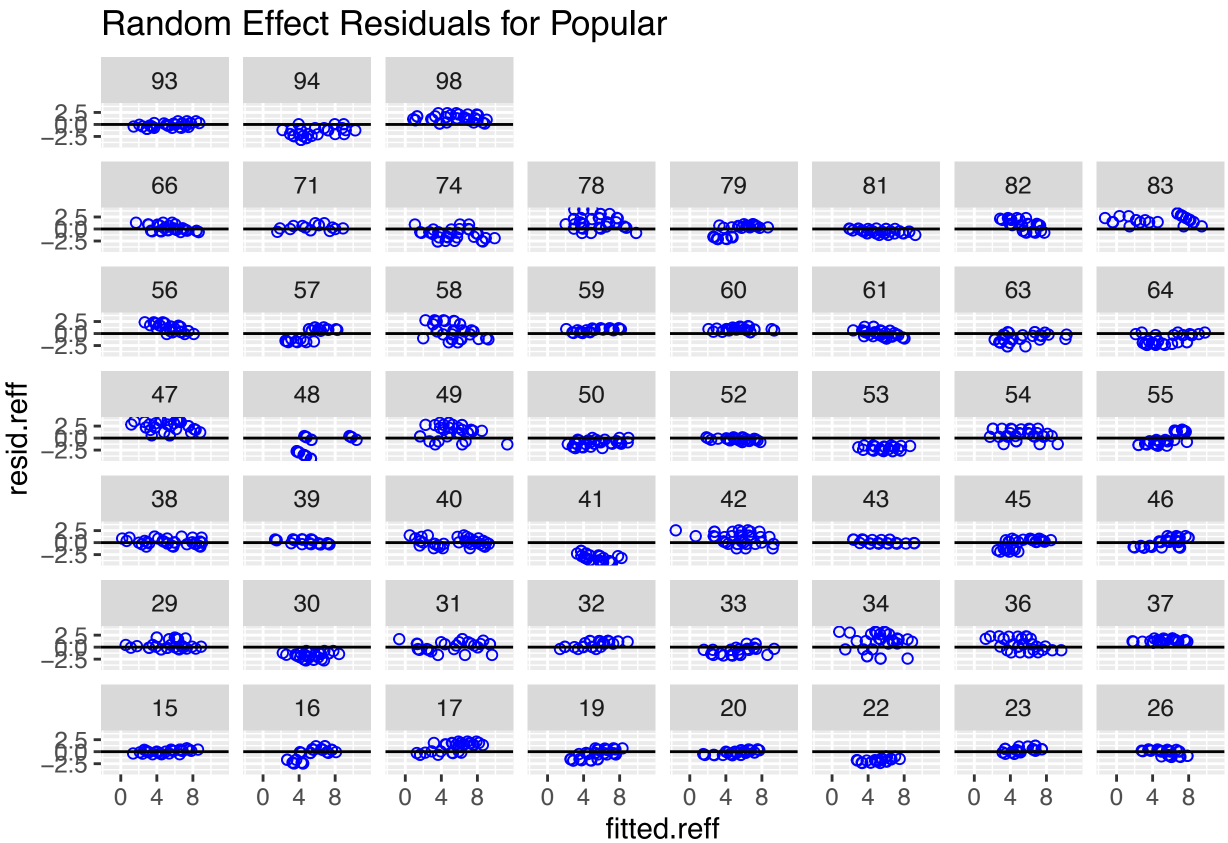

Figure 9: Random Effects Residuals for Popular

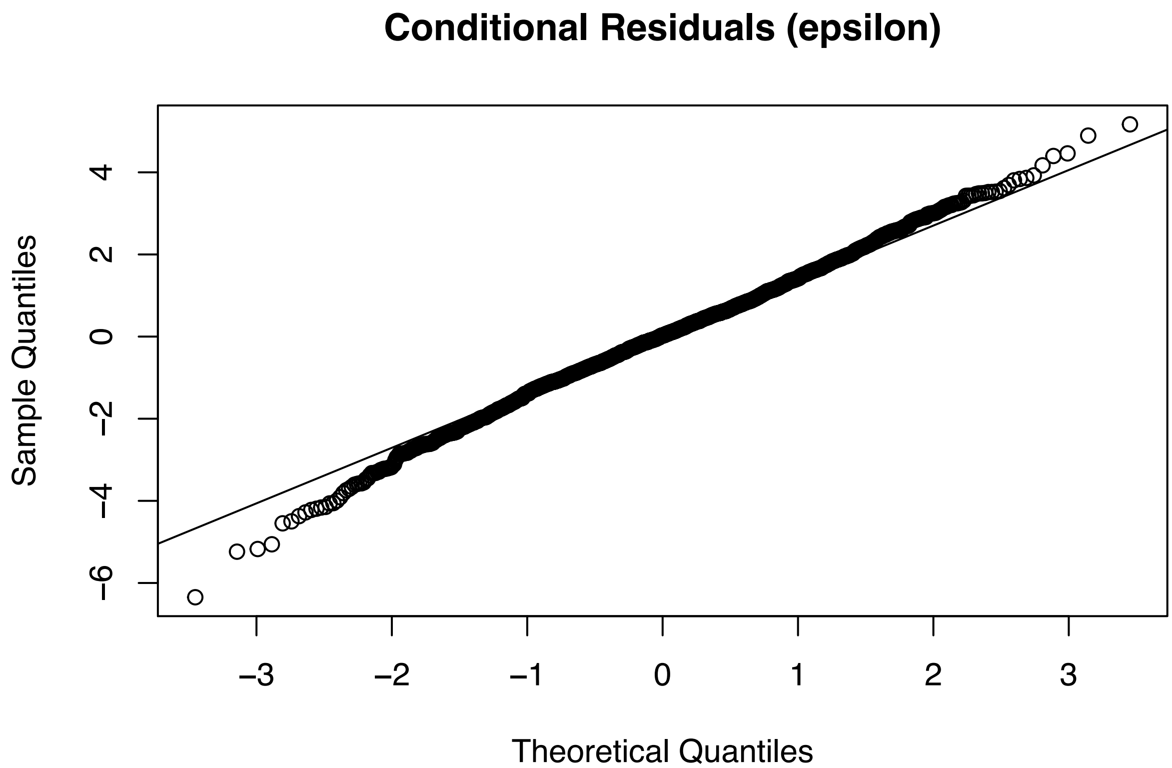

Figure 10: Q-Q plot of Conditional Residuals for Popular

All these figures look good since there is no specific pattern appear in Figure 7,8 and 9 and almost all points lie on the diagnoal line in Figure 10. Thus we can conclude that the multilevel model we tested for popular model is a valid model.

4. Automatic Model Selection

4.1 For Classical Ratings

Select for Fixed Effects

```
X.cont <- names(data)
X.cont <- X.cont[-c(1,19,20)]

(max.c <- as.formula(paste("Classical ~",
                          paste(c(X.cont,"(1|Subject)","Harmony:Voice"),
                                collapse="+"))))

summary(max.c.1 <- lmer(max.c, data, REML = F, control = lmerControl(optimizer = 'bobyqa')))

llrt.c <- fitLMER.fnc(max.c.1, method="llrt")
```

Select for Random Effects

```
vars <- attr(terms(formula(llrt.c)),"term.labels")
vars <- vars[-c(8,9)]
```

```r
## this call to fitLMER.fnc() forces the corrrelation between each
## random slope and the random intercept to be zero

llrt.c.1.0 <- fitLMER.fnc(llrt.c,
                           ran.effects=
                             list(slopes=vars, by.vars="Subject",
                                  corr=rep(0,length(vars))), method = 'llrt')

## this call to fitLMER.fnc() allows a correlation between each random
## slope and the random intercept on state

llrt.c.1.1 <- fitLMER.fnc(llrt.c,
                           ran.effects=
                             list(slopes=vars, by.vars="Subject",
                                  corr=rep(1,length(vars))), method = 'llrt')

anova(llrt.c.1.0, llrt.c.1.1)  #  this end up with the same model

llrt.c.1.1.2 <- lmer(Classical ~ Harmony + Instrument + Voice +   X16.minus.17+ (1 +
                     Instrument | Subject) + Harmony:Voice, data = data,
                      REML = FALSE, control = lmerControl(optimizer = "bobyqa"))

llrt.c.1.1.3 <- lmer(Classical ~ Harmony + Instrument + Voice +   X16.minus.17+ (1 +
                     Harmony + Instrument | Subject) + Harmony:Voice, data = data,
                      REML = FALSE, control = lmerControl(optimizer = "bobyqa"))
                      # smallest AIC and BIC and thus the best model

anova(llrt.c.1.1, llrt.c.1.1.2, llrt.c.1.1.3)
```

4.2 For Popular Ratings

Select Fixed Effects

```r
X.cont <- names(data)
X.cont <- X.cont[-c(1,19,20)]

(max.c <- as.formula(paste("Popular ~",
                             paste(c(X.cont,"(1|Subject)"),
                                   collapse="+"))))

summary(max.c.1 <- lmer(max.c, data, REML = F, control = lmerControl(optimizer = 'bobyqa')))

llrt.c <- LMERConvenienceFunctions::fitLMER.fnc(max.c.1, method="llrt")
```

Select Random Effects

```r
llrt.c.1.0 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 | Subject),
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

llrt.c.1.1 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 + Harmony + Instrument + Voice| Subject)
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

llrt.c.1.2 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 + Harmony + Instrument| Subject),
```

```
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

llrt.c.1.3 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 + Harmony + Voice| Subject),
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))
                  # smallest AIC and BIC and thus the final model

llrt.c.1.4 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 + Instrument + Voice| Subject),
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

llrt.c.1.5 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 + Harmony| Subject),
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

llrt.c.1.6 <- lmer(Popular ~ Harmony + Instrument + Voice + (1  + Instrument | Subject),
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

llrt.c.1.7 <- lmer(Popular ~ Harmony + Instrument + Voice + (1 + Voice| Subject),
                   data, REML = F, control = lmerControl(optimizer = 'bobyqa'))

anova(llrt.c.1.0, llrt.c.1.1, llrt.c.1.2,llrt.c.1.3,
      llrt.c.1.4,llrt.c.1.5,llrt.c.1.6,llrt.c.1.7)
```

References

R Core Team (2017), R: A language and environment for statistical computing.

Ivan J , Vincent R (2013) The Influence of Timbre, Harmony, and Voice Leading on Listeners' Distinction between Popular and Classical Music. Pittsburgh: University of Pittsburgh.