Musical features analysis

R. Cassano-Coleman

2025-01-17

This script explores the musical features in the expanded corpus, specifically:

- correlations between feature differences
- correlations between individual feature differences and similarity between originals and covers

We include MFCCs distance as an overall measure of timbre.

First, load the features and feature differences (unscaled).

Load the scaled feature differences for later.

Select only columns with musical features.

For genre - we don't have a great way of measuring genre distance, but we do have some examples where the original and cover are in the same genre. Create "same genre" predictor.

How many examples are there where original-cover pairs are in the same genre?

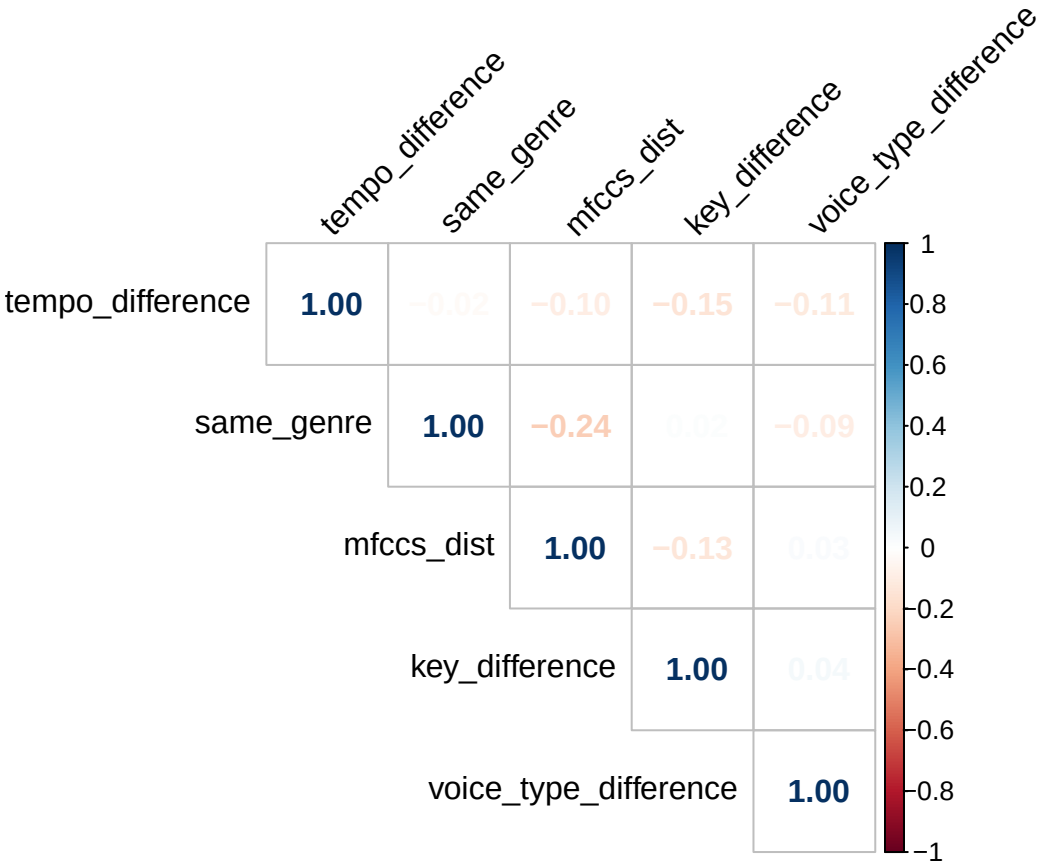
```
## # A tibble: 16 x 4
##    song_id                                Cover5S Original5S same_genre
##    <chr>                                    <dbl>      <dbl>      <dbl>
##  1 BlurredLines_RobinThicke                     1          1          1
##  2 Dilemma_NellyKellyRowland                   12         12          1
##  3 FlashdanceWhatAFeeling_IreneCara            13         13          1
##  4 GetUrFreakOn_MissyElliot                    14         14          1
##  5 GoodGollyMissMolly_LittleRichard             9          9          1
##  6 Happy_PharellWilliams                        1          1          1
##  7 IWannaDanceWithSomebody_WhitneyHouston       1          1          1
##  8 IWillAlwaysLoveYou_WhitneyHouston           10         10          1
##  9 LikeAPrayer_Madonna                          1          1          1
## 10 LikeARollingStone_BobDylan                   9          9          1
## 11 SmellsLikeTeenSpirit_Nirvana                 3          3          1
## 12 Superstition_StevieWonder                   12         12          1
## 13 TheBoyIsMine_BrandyMonica                   12         12          1
## 14 UptownFunk_MarkRonson                        1          1          1
## 15 Weak_SWV                                    12         12          1
## 16 YouLightUpMyLife_DebbyBoone                 10         10          1
```

16 (out of 70)

Join same genre with other musical feature differences.

Are there correlations between the feature differences?

Remove treble and TBB difference - summarized in VT difference



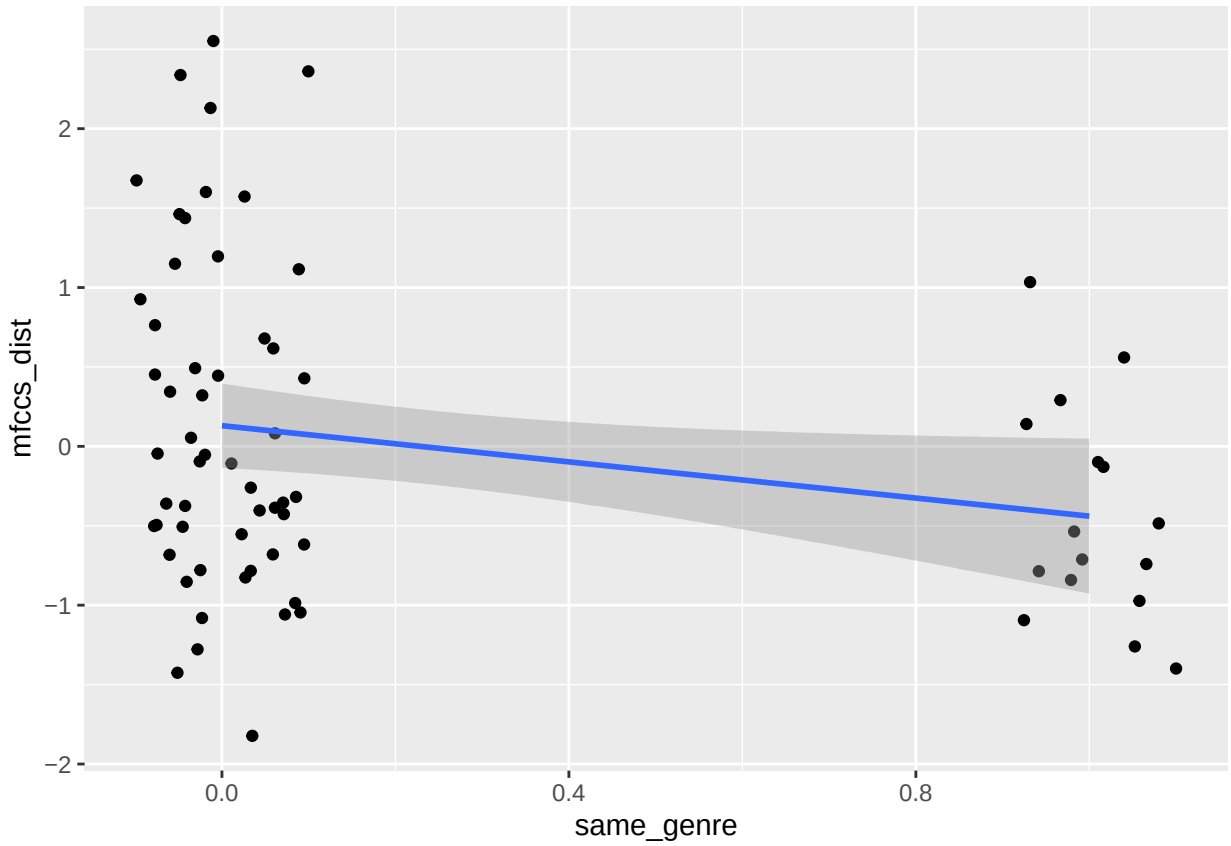

```
##
## Call:
## lm(formula = tempo_difference ~ same_genre, data = musical_diffs_scaled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9288 -0.7874 -0.3316  0.3976  3.3158
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1203     0.1336   0.900   0.3710
## same_genre   -0.5265     0.2795  -1.883   0.0639 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.982 on 68 degrees of freedom
## Multiple R-squared:  0.04958,    Adjusted R-squared:  0.03561
## F-statistic: 3.548 on 1 and 68 DF,  p-value: 0.06391
```

```
##
## Call:
## lm(formula = key_difference ~ same_genre, data = musical_diffs_scaled)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -0.7722 -0.7722 -0.1544  0.4633  2.9343
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01765    0.13701   0.129    0.898
## same_genre  -0.07722    0.28657  -0.269    0.788
##
## Residual standard error: 1.007 on 68 degrees of freedom
## Multiple R-squared:  0.001067,   Adjusted R-squared:  -0.01362
## F-statistic: 0.07261 on 1 and 68 DF,  p-value: 0.7884
```

```
##
## Call:
## lm(formula = voice_type_difference ~ same_genre, data = musical_diffs_scaled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7691 -0.7691 -0.5464  1.4167  1.6393
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05088    0.13647   0.373    0.710
## same_genre  -0.22262    0.28545  -0.780    0.438
##
## Residual standard error: 1.003 on 68 degrees of freedom
## Multiple R-squared:  0.008865,   Adjusted R-squared:  -0.00571
## F-statistic: 0.6082 on 1 and 68 DF,  p-value: 0.4382
```

```
##
## Call:
## lm(formula = mfccs_dist ~ same_genre, data = musical_diffs_scaled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9529 -0.6508 -0.2316  0.5724  2.4211
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1302     0.1330   0.979   0.3311
## same_genre   -0.5698     0.2783  -2.047   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9776 on 68 degrees of freedom
## Multiple R-squared:  0.05807,    Adjusted R-squared:  0.04422
## F-statistic: 4.192 on 1 and 68 DF,  p-value: 0.04448
```


mfccs_dist
2
1
0
−1
−2
0.0
0.4
0.8
same_genre

Do the individual features predict similarity?

We don't need to take the absolute value of the differences - we did that in `4_data_joining.R` before each feature difference was scaled.

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.93106 -0.67857  0.09402  0.64609  1.73546
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -4.429e-16  9.791e-02   0.000        1
## tempo_difference -5.819e-01  9.862e-02  -5.901 1.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8192 on 68 degrees of freedom
## Multiple R-squared:  0.3386, Adjusted R-squared:  0.3289
## F-statistic: 34.82 on 1 and 68 DF,  p-value: 1.267e-07
```

```
##
## Call:
## lm(formula = mean_sim ~ key_difference, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.89729 -0.68653  0.02209  0.73678  2.11831
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.566e-16  1.132e-01   0.000  1.00000
## key_difference -3.398e-01  1.141e-01  -2.979  0.00401 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9474 on 68 degrees of freedom
## Multiple R-squared:  0.1154, Adjusted R-squared:  0.1024
## F-statistic: 8.874 on 1 and 68 DF,  p-value: 0.004008
```

```
##
## Call:
## lm(formula = mean_sim ~ voice_type_difference, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92757 -0.75269 -0.04427  0.74178  1.94479
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -4.148e-16  1.146e-01   0.000  1.00000
## voice_type_difference -3.063e-01  1.154e-01  -2.654  0.00991 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9589 on 68 degrees of freedom
## Multiple R-squared:  0.09384,    Adjusted R-squared:  0.08051
## F-statistic: 7.042 on 1 and 68 DF,  p-value: 0.009906
```

```
##
## Call:
## lm(formula = mean_sim ~ mfccs_dist, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.91617 -0.79584  0.02915  0.88035  1.74669
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.961e-16  1.149e-01   0.000   1.0000
## mfccs_dist  -2.989e-01  1.157e-01  -2.583   0.0119 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9613 on 68 degrees of freedom
## Multiple R-squared:  0.08937,    Adjusted R-squared:  0.07598
## F-statistic: 6.673 on 1 and 68 DF,  p-value: 0.01194
```

```
##
## Call:
## lm(formula = mean_sim ~ same_genre, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.08500 -0.81692  0.01607  0.89446  1.63077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1582     0.1311  -1.207   0.2315
## same_genre    0.6922     0.2742   2.525   0.0139 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9632 on 68 degrees of freedom
## Multiple R-squared:  0.08572,    Adjusted R-squared:  0.07228
## F-statistic: 6.376 on 1 and 68 DF,  p-value: 0.01391
```

How do the individual features interact with genre?

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference * same_genre, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78382 -0.51894  0.06882  0.60330  1.19848
##
```

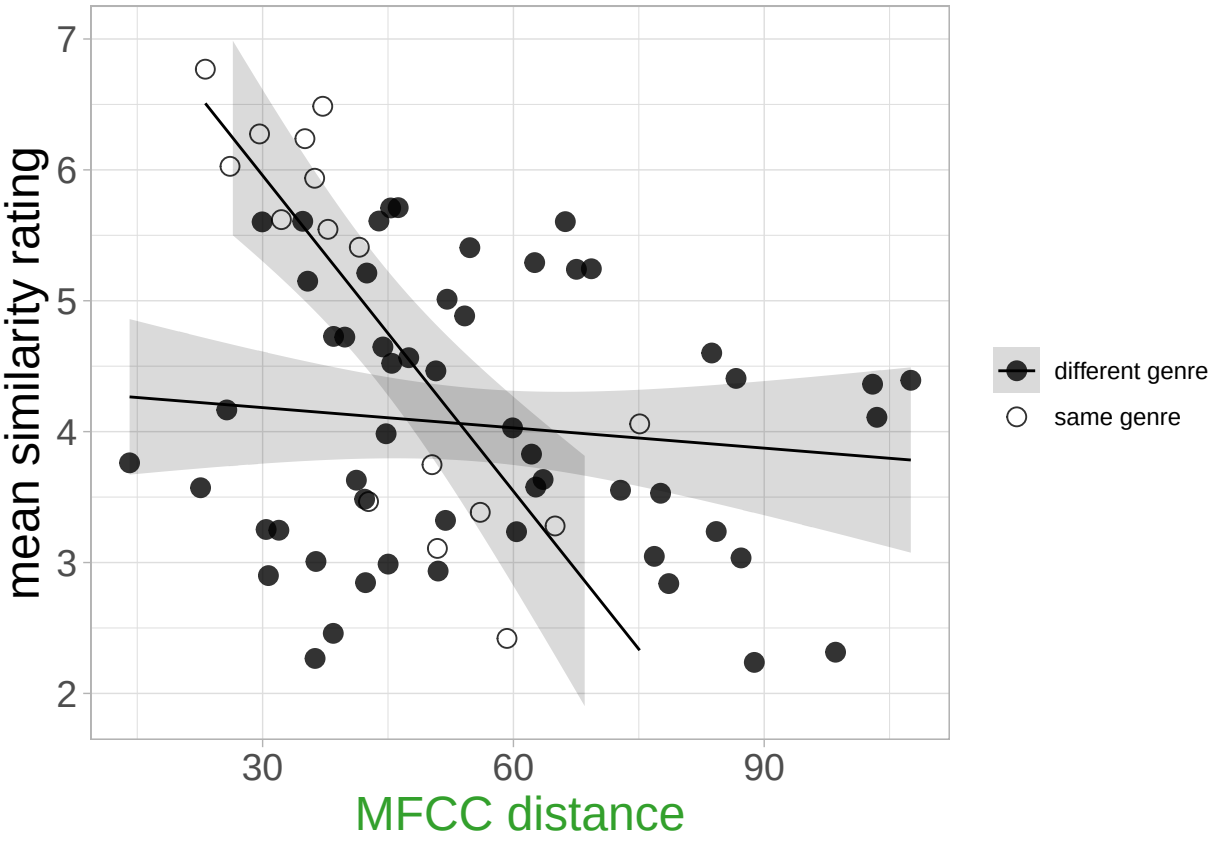
```
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -0.1023     0.1063  -0.962    0.340
## tempo_difference              -0.4649     0.1007  -4.616 1.86e-05 ***
## same_genre                     0.1069     0.2551   0.419    0.677
## tempo_difference:same_genre   -0.8387     0.3285  -2.553    0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7763 on 66 degrees of freedom
## Multiple R-squared:  0.4236, Adjusted R-squared:  0.3974
## F-statistic: 16.17 on 3 and 66 DF,  p-value: 5.528e-08
```

```
##
## Call:
## lm(formula = mean_sim ~ key_difference * same_genre, data = musical_diffs_scaled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8507 -0.7129  0.1056  0.7773  1.5973
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -0.1527     0.1247  -1.224   0.2252
## key_difference              -0.3134     0.1192  -2.629   0.0106 *
## same_genre                   0.6609     0.2614   2.528   0.0139 *
## key_difference:same_genre   -0.1200     0.3157  -0.380   0.7051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9165 on 66 degrees of freedom
## Multiple R-squared:  0.1966, Adjusted R-squared:  0.1601
## F-statistic: 5.384 on 3 and 66 DF,  p-value: 0.002238
```

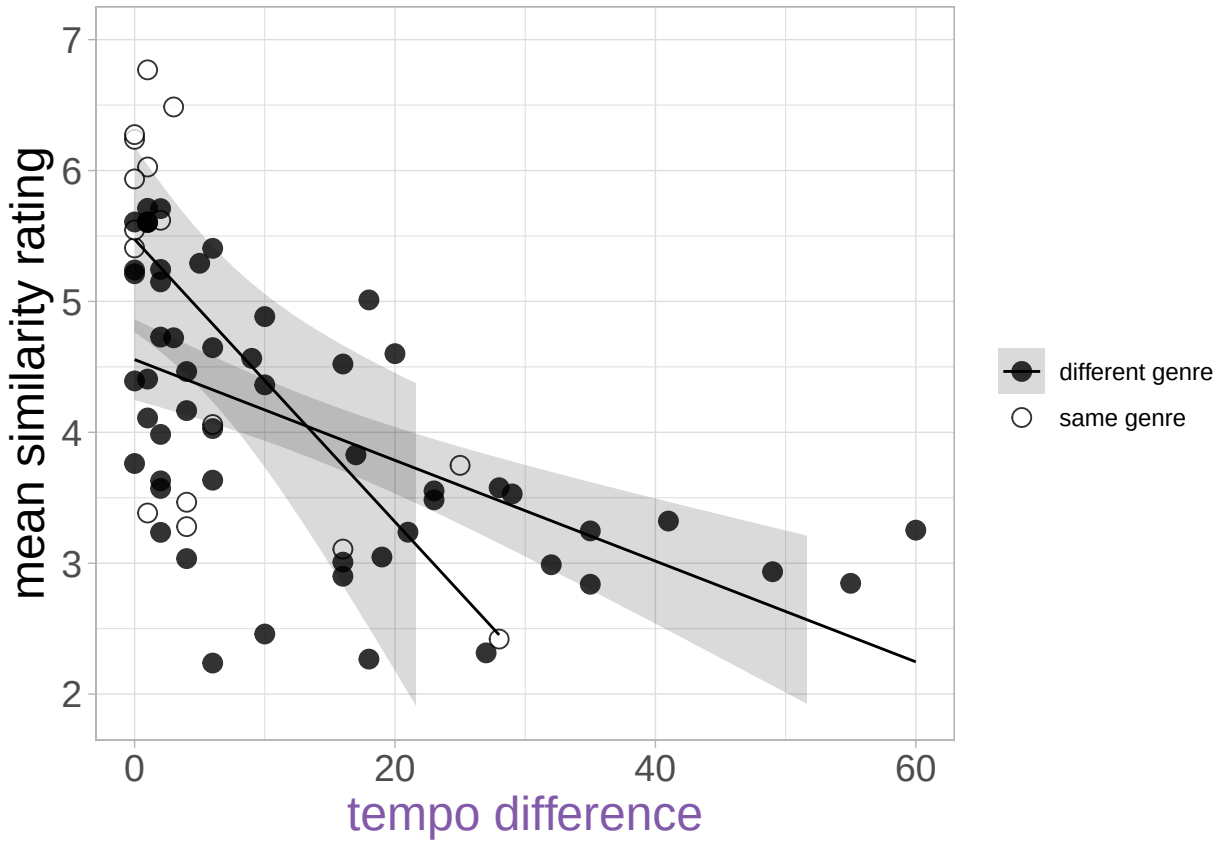
```
##
## Call:
## lm(formula = mean_sim ~ voice_type_difference * same_genre, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20122 -0.83964  0.06652  0.79946  1.51455
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -0.14299    0.12728  -1.123   0.2653
## voice_type_difference            -0.29944    0.12705  -2.357   0.0214 *
## same_genre                        0.64048    0.26932   2.378   0.0203 *
## voice_type_difference:same_genre  0.08676    0.27753   0.313   0.7556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9341 on 66 degrees of freedom
## Multiple R-squared:  0.1654, Adjusted R-squared:  0.1274
## F-statistic: 4.359 on 3 and 66 DF,  p-value: 0.007321
```

```
##
## Call:
## lm(formula = mean_sim ~ mfccs_dist * same_genre, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.60968 -0.66689 -0.01719  0.57109  1.47760
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -0.14596    0.11630  -1.255 0.213886
## mfccs_dist            -0.09420    0.11136  -0.846 0.400648
## same_genre             0.03475    0.27976   0.124 0.901514
## mfccs_dist:same_genre -1.37381    0.33898  -4.053 0.000136 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8479 on 66 degrees of freedom
## Multiple R-squared:  0.3123, Adjusted R-squared:  0.281
## F-statistic: 9.989 on 3 and 66 DF,  p-value: 1.631e-05
```

Visualize timbre and genre



Visualize tempo and genre

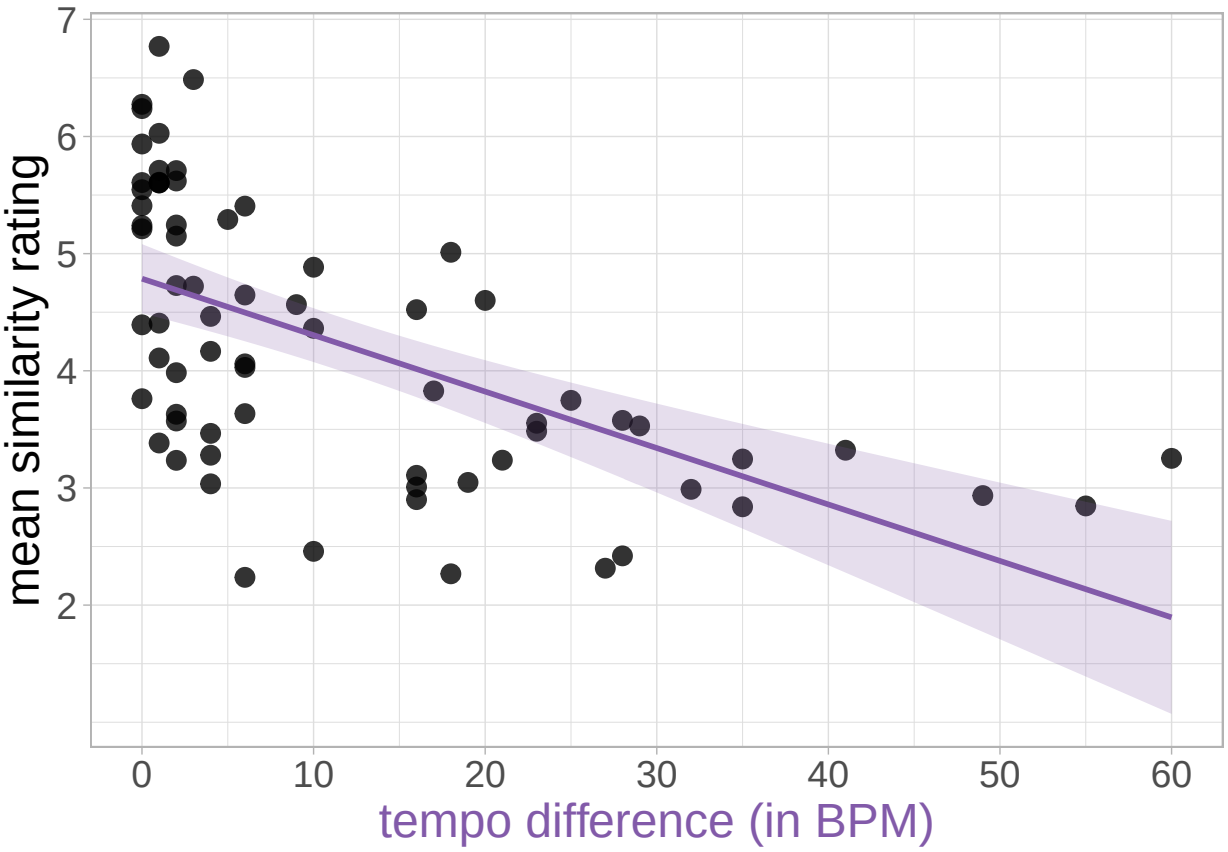

mean similarity rating
7
6
5
4
3
2
0
20
40
60
tempo difference
different genre
same genre

Visualize features individually.

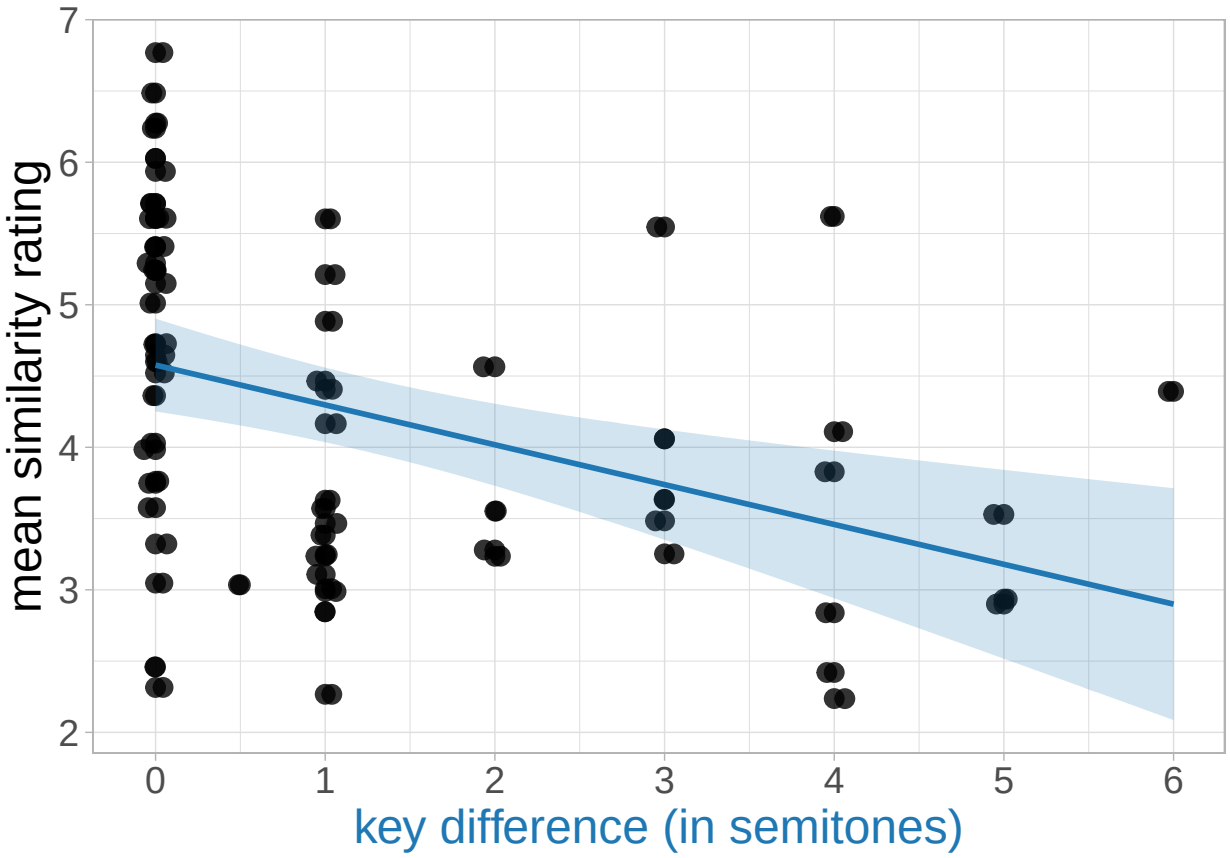
Color palette adapted from paired palette in RColorBrewer. Blended two colors using https://meyerweb.com /eric/tools/color-blend/ For the main colors: used darker version in paired palette (previously midpoint 3, but that was too light for text) For teal, blended blue and green ^

Tempo: purple #825AA9 Key: blue #1F78B4 Voice type: teal #298C70 Timbre: green #33A02C

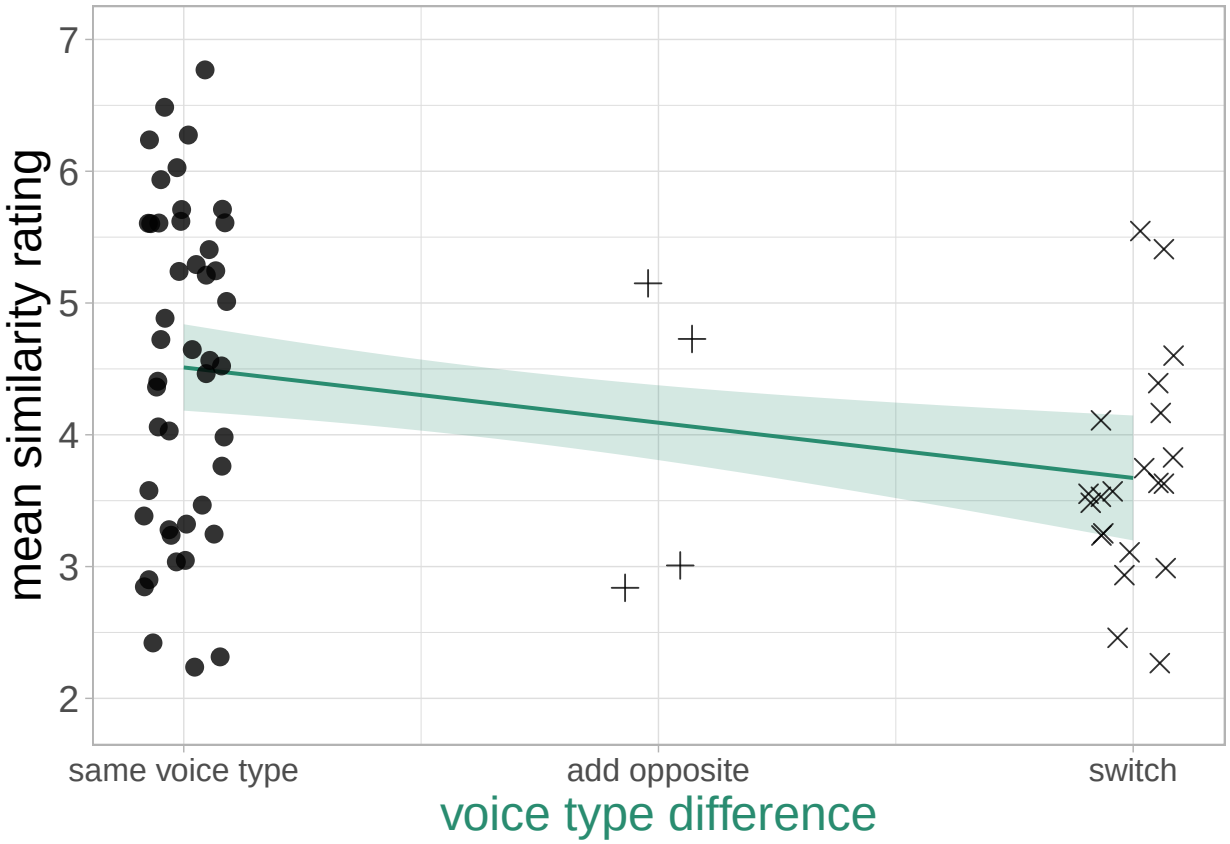
Tempo



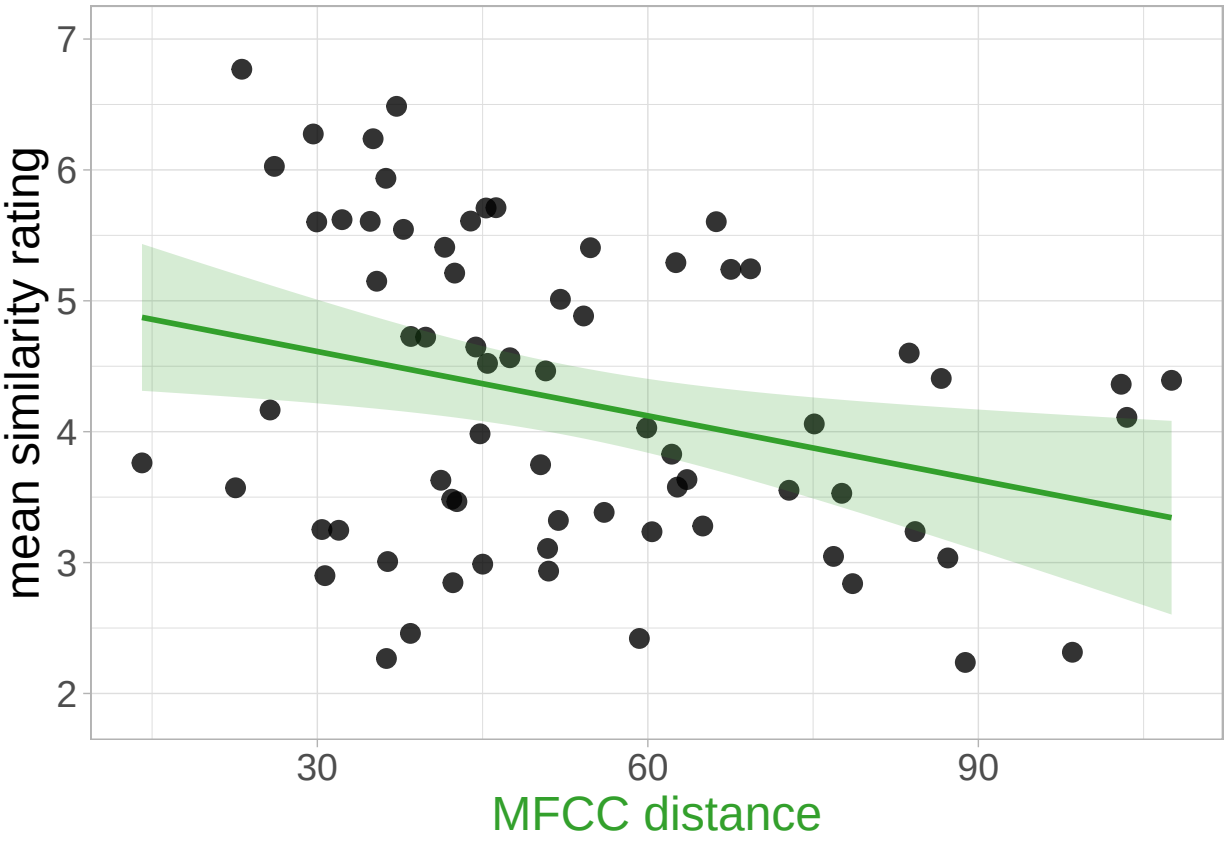
Key


mean similarity rating
7
6
5
4
3
2
0
1
2
3
4
5
6
key difference (in semitones)

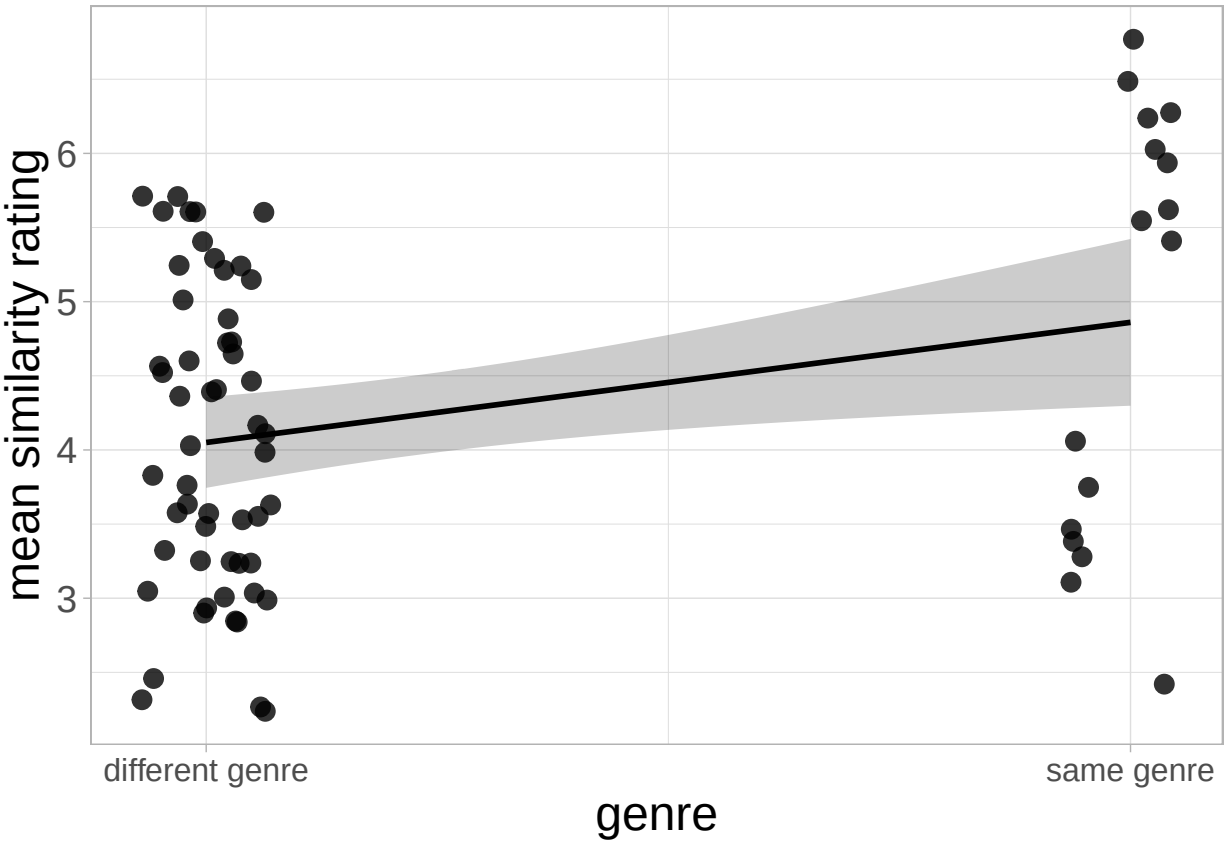
Voice type


mean similarity rating
7
6
5
4
3
2
same voice type
add opposite
switch
voice type difference

Timbre



Genre


mean similarity rating
6
5
4
3
different genre
same genre
genre

How do the musical features together predict similarity?

Make voice type a factor, and include "add opposite" in "switch"

Full model

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference + key_difference + voice_type_factor +
##     mfccs_dist + same_genre, data = musical_diffs_scaled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5531 -0.5818  0.2457  0.5737  1.2014
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.10076    0.10888  -0.925   0.3582
## tempo_difference   -0.48499    0.09911  -4.893 7.02e-06 ***
## key_difference     -0.10577    0.10796  -0.980   0.3309
## voice_type_factor1  0.13139    0.10651   1.234   0.2219
## mfccs_dist         -0.20297    0.10092  -2.011   0.0485 *
## same_genre          0.27660    0.23409   1.182   0.2417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7705 on 64 degrees of freedom
## Multiple R-squared:  0.4494, Adjusted R-squared:  0.4064
## F-statistic: 10.45 on 5 and 64 DF,  p-value: 2.341e-07
```

Full model, with interactions

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference * key_difference * voice_type_factor *
##     mfccs_dist * same_genre, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41170 -0.22929  0.04826  0.38981  0.81194
##
## Coefficients: (4 not defined because of singularities)
##                                                                          Estimate
## (Intercept)                                                             -0.34333
## tempo_difference                                                        -0.46355
## key_difference                                                          -0.32316
## voice_type_factor1                                                       0.19839
## mfccs_dist                                                              -0.07654
## same_genre                                                               1.08880
## tempo_difference:key_difference                                         -0.10287
## tempo_difference:voice_type_factor1                                     -0.31624
## key_difference:voice_type_factor1                                       -0.23055
## tempo_difference:mfccs_dist                                              0.02915
## key_difference:mfccs_dist                                               -0.05822
## voice_type_factor1:mfccs_dist                                           -0.20812
## tempo_difference:same_genre                                             -3.78812
## key_difference:same_genre                                               -3.54514
## voice_type_factor1:same_genre                                           -2.24335
## mfccs_dist:same_genre                                                   12.74823
## tempo_difference:key_difference:voice_type_factor1                      -0.15624
## tempo_difference:key_difference:mfccs_dist                              -0.24690
## tempo_difference:voice_type_factor1:mfccs_dist                          -0.41899
## key_difference:voice_type_factor1:mfccs_dist                            -0.08954
## tempo_difference:key_difference:same_genre                              -4.66055
## tempo_difference:voice_type_factor1:same_genre                           2.80440
## key_difference:voice_type_factor1:same_genre                            -0.13279
## tempo_difference:mfccs_dist:same_genre                                   9.88731
## key_difference:mfccs_dist:same_genre                                     4.75340
## voice_type_factor1:mfccs_dist:same_genre                                -7.50120
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist            0.13358
## tempo_difference:key_difference:voice_type_factor1:same_genre                 NA
## tempo_difference:key_difference:mfccs_dist:same_genre                    8.19297
## tempo_difference:voice_type_factor1:mfccs_dist:same_genre                     NA
## key_difference:voice_type_factor1:mfccs_dist:same_genre                       NA
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist:same_genre      NA
##                                                                         Std. Error
## (Intercept)                                                                0.14472
## tempo_difference                                                           0.18839
## key_difference                                                             0.16900
## voice_type_factor1                                                         0.14472
## mfccs_dist                                                                 0.14464
## same_genre                                                                 3.68484
## tempo_difference:key_difference                                            0.22898
## tempo_difference:voice_type_factor1                                        0.18839
```

```
## key_difference:voice_type_factor1                                          0.16900
## tempo_difference:mfccs_dist                                                0.21333
## key_difference:mfccs_dist                                                  0.13150
## voice_type_factor1:mfccs_dist                                              0.14464
## tempo_difference:same_genre                                                5.08055
## key_difference:same_genre                                                  5.71194
## voice_type_factor1:same_genre                                              2.68571
## mfccs_dist:same_genre                                                     19.93977
## tempo_difference:key_difference:voice_type_factor1                         0.22898
## tempo_difference:key_difference:mfccs_dist                                 0.18255
## tempo_difference:voice_type_factor1:mfccs_dist                             0.21333
## key_difference:voice_type_factor1:mfccs_dist                               0.13150
## tempo_difference:key_difference:same_genre                                 8.66251
## tempo_difference:voice_type_factor1:same_genre                             6.39526
## key_difference:voice_type_factor1:same_genre                               0.76041
## tempo_difference:mfccs_dist:same_genre                                     9.82990
## key_difference:mfccs_dist:same_genre                                       2.82497
## voice_type_factor1:mfccs_dist:same_genre                                  13.27018
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist              0.18255
## tempo_difference:key_difference:voice_type_factor1:same_genre                   NA
## tempo_difference:key_difference:mfccs_dist:same_genre                      5.23995
## tempo_difference:voice_type_factor1:mfccs_dist:same_genre                       NA
## key_difference:voice_type_factor1:mfccs_dist:same_genre                         NA
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist:same_genre        NA
##                                                                          t value
## (Intercept)                                                               -2.372
## tempo_difference                                                          -2.461
## key_difference                                                            -1.912
## voice_type_factor1                                                         1.371
## mfccs_dist                                                                -0.529
## same_genre                                                                 0.295
## tempo_difference:key_difference                                           -0.449
## tempo_difference:voice_type_factor1                                       -1.679
## key_difference:voice_type_factor1                                         -1.364
## tempo_difference:mfccs_dist                                                0.137
## key_difference:mfccs_dist                                                 -0.443
## voice_type_factor1:mfccs_dist                                             -1.439
## tempo_difference:same_genre                                               -0.746
## key_difference:same_genre                                                 -0.621
## voice_type_factor1:same_genre                                             -0.835
## mfccs_dist:same_genre                                                      0.639
## tempo_difference:key_difference:voice_type_factor1                        -0.682
## tempo_difference:key_difference:mfccs_dist                                -1.353
## tempo_difference:voice_type_factor1:mfccs_dist                            -1.964
## key_difference:voice_type_factor1:mfccs_dist                              -0.681
## tempo_difference:key_difference:same_genre                                -0.538
## tempo_difference:voice_type_factor1:same_genre                             0.439
## key_difference:voice_type_factor1:same_genre                              -0.175
## tempo_difference:mfccs_dist:same_genre                                     1.006
## key_difference:mfccs_dist:same_genre                                       1.683
## voice_type_factor1:mfccs_dist:same_genre                                  -0.565
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist              0.732
## tempo_difference:key_difference:voice_type_factor1:same_genre                 NA
## tempo_difference:key_difference:mfccs_dist:same_genre                      1.564
```

```
## tempo_difference:voice_type_factor1:mfccs_dist:same_genre                    NA
## key_difference:voice_type_factor1:mfccs_dist:same_genre                      NA
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist:same_genre     NA
##                                                                          Pr(>|t|)
## (Intercept)                                                                0.0223
## tempo_difference                                                           0.0181
## key_difference                                                             0.0627
## voice_type_factor1                                                         0.1777
## mfccs_dist                                                                 0.5995
## same_genre                                                                 0.7691
## tempo_difference:key_difference                                            0.6556
## tempo_difference:voice_type_factor1                                        0.1006
## key_difference:voice_type_factor1                                          0.1798
## tempo_difference:mfccs_dist                                                0.8920
## key_difference:mfccs_dist                                                  0.6603
## voice_type_factor1:mfccs_dist                                              0.1576
## tempo_difference:same_genre                                                0.4601
## key_difference:same_genre                                                  0.5382
## voice_type_factor1:same_genre                                              0.4083
## mfccs_dist:same_genre                                                      0.5261
## tempo_difference:key_difference:voice_type_factor1                         0.4988
## tempo_difference:key_difference:mfccs_dist                                 0.1835
## tempo_difference:voice_type_factor1:mfccs_dist                             0.0562
## key_difference:voice_type_factor1:mfccs_dist                               0.4997
## tempo_difference:key_difference:same_genre                                 0.5934
## tempo_difference:voice_type_factor1:same_genre                             0.6633
## key_difference:voice_type_factor1:same_genre                               0.8622
## tempo_difference:mfccs_dist:same_genre                                     0.3203
## key_difference:mfccs_dist:same_genre                                       0.0999
## voice_type_factor1:mfccs_dist:same_genre                                   0.5749
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist              0.4684
## tempo_difference:key_difference:voice_type_factor1:same_genre                 NA
## tempo_difference:key_difference:mfccs_dist:same_genre                      0.1254
## tempo_difference:voice_type_factor1:mfccs_dist:same_genre                     NA
## key_difference:voice_type_factor1:mfccs_dist:same_genre                       NA
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist:same_genre      NA
##
## (Intercept)                                                              *
## tempo_difference                                                         *
## key_difference                                                           .
## voice_type_factor1
## mfccs_dist
## same_genre
## tempo_difference:key_difference
## tempo_difference:voice_type_factor1
## key_difference:voice_type_factor1
## tempo_difference:mfccs_dist
## key_difference:mfccs_dist
## voice_type_factor1:mfccs_dist
## tempo_difference:same_genre
## key_difference:same_genre
## voice_type_factor1:same_genre
## mfccs_dist:same_genre
## tempo_difference:key_difference:voice_type_factor1
```

```
## tempo_difference:key_difference:mfccs_dist
## tempo_difference:voice_type_factor1:mfccs_dist                          .
## key_difference:voice_type_factor1:mfccs_dist
## tempo_difference:key_difference:same_genre
## tempo_difference:voice_type_factor1:same_genre
## key_difference:voice_type_factor1:same_genre
## tempo_difference:mfccs_dist:same_genre
## key_difference:mfccs_dist:same_genre                                    .
## voice_type_factor1:mfccs_dist:same_genre
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist
## tempo_difference:key_difference:voice_type_factor1:same_genre
## tempo_difference:key_difference:mfccs_dist:same_genre
## tempo_difference:voice_type_factor1:mfccs_dist:same_genre
## key_difference:voice_type_factor1:mfccs_dist:same_genre
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist:same_genre
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6416 on 42 degrees of freedom
## Multiple R-squared:  0.7494, Adjusted R-squared:  0.5883
## F-statistic: 4.652 on 27 and 42 DF,  p-value: 4.602e-06
```

Is the benefit of better R^2 worth all of these degrees of freedom? Use ANOVA to test the nested models.

```
## Analysis of Variance Table
##
## Model 1: mean_sim ~ tempo_difference + key_difference + voice_type_factor +
##     mfccs_dist + same_genre
## Model 2: mean_sim ~ tempo_difference * key_difference * voice_type_factor *
##     mfccs_dist * same_genre
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     64 37.993
## 2     42 17.292 22    20.701 2.2855 0.01048 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All of these interactions are worth it ($p = 0.010$).

Without genre, there were marginally significant interactions are among key, voice type, and timbre.

What if we separate tempo and just let the other features interact?

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference + key_difference * voice_type_factor *
##     mfccs_dist * same_genre, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41225 -0.22629  0.00173  0.43445  1.11461
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                             -0.174773   0.113413
## tempo_difference                                        -0.502285   0.100552
## key_difference                                          -0.128508   0.119142
## voice_type_factor1                                       0.090087   0.113067
## mfccs_dist                                               0.003803   0.113380
## same_genre                                              -0.785494   0.516087
## key_difference:voice_type_factor1                       -0.272067   0.115237
## key_difference:mfccs_dist                               -0.092782   0.099941
## voice_type_factor1:mfccs_dist                           -0.184224   0.113397
## key_difference:same_genre                               -0.310297   0.962270
## voice_type_factor1:same_genre                            0.220087   0.516449
## mfccs_dist:same_genre                                   -2.285368   0.861996
## key_difference:voice_type_factor1:mfccs_dist             0.132692   0.092905
## key_difference:voice_type_factor1:same_genre             1.380184   0.961213
## key_difference:mfccs_dist:same_genre                    -0.485214   1.502033
## voice_type_factor1:mfccs_dist:same_genre                 0.782622   0.857706
## key_difference:voice_type_factor1:mfccs_dist:same_genre  1.377735   1.502972
##                                                         t value Pr(>|t|)
## (Intercept)                                              -1.541   0.1293
## tempo_difference                                         -4.995 6.75e-06 ***
## key_difference                                           -1.079   0.2856
## voice_type_factor1                                        0.797   0.4291
## mfccs_dist                                                0.034   0.9734
## same_genre                                               -1.522   0.1339
## key_difference:voice_type_factor1                        -2.361   0.0219 *
## key_difference:mfccs_dist                                -0.928   0.3574
## voice_type_factor1:mfccs_dist                            -1.625   0.1102
## key_difference:same_genre                                -0.322   0.7484
## voice_type_factor1:same_genre                             0.426   0.6717
## mfccs_dist:same_genre                                    -2.651   0.0106 *
## key_difference:voice_type_factor1:mfccs_dist              1.428   0.1591
## key_difference:voice_type_factor1:same_genre              1.436   0.1569
## key_difference:mfccs_dist:same_genre                     -0.323   0.7479
## voice_type_factor1:mfccs_dist:same_genre                  0.912   0.3657
## key_difference:voice_type_factor1:mfccs_dist:same_genre   0.917   0.3635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6578 on 53 degrees of freedom
## Multiple R-squared:  0.6677, Adjusted R-squared:  0.5673
```

```
## F-statistic: 6.655 on 16 and 53 DF,  p-value: 6.792e-08
```

Are these interactions worth it?

```
## Analysis of Variance Table
##
## Model 1: mean_sim ~ tempo_difference + key_difference + voice_type_factor +
##     mfccs_dist + same_genre
## Model 2: mean_sim ~ tempo_difference + key_difference * voice_type_factor *
##     mfccs_dist * same_genre
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     64 37.993
## 2     53 22.931 11    15.063 3.1649 0.002356 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, adding these interactions adds sufficient explanatory benefit to merit including them.

How do the musical features together predict similarity WITHOUT genre?

Full model

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference + key_difference + voice_type_factor +
##     mfccs_dist, data = musical_diffs_scaled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6065 -0.6007  0.1769  0.5988  1.3440
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.04212    0.09721  -0.433   0.6662
## tempo_difference   -0.50963    0.09719  -5.244 1.84e-06 ***
## key_difference     -0.08862    0.10731  -0.826   0.4119
## voice_type_factor1  0.14743    0.10597   1.391   0.1689
## mfccs_dist         -0.23407    0.09772  -2.395   0.0195 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7728 on 65 degrees of freedom
## Multiple R-squared:  0.4374, Adjusted R-squared:  0.4027
## F-statistic: 12.63 on 4 and 65 DF,  p-value: 1.161e-07
```

Full model, with interactions

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference * key_difference * voice_type_factor *
##     mfccs_dist, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.66949 -0.40896  0.05532  0.45298  1.13879
##
## Coefficients:
##                                                                Estimate
## (Intercept)                                                    -0.18133
## tempo_difference                                               -0.57328
## key_difference                                                 -0.19727
## voice_type_factor1                                              0.16466
## mfccs_dist                                                     -0.17036
## tempo_difference:key_difference                                -0.15575
## tempo_difference:voice_type_factor1                            -0.25173
## key_difference:voice_type_factor1                              -0.26866
## tempo_difference:mfccs_dist                                     0.10496
## key_difference:mfccs_dist                                      -0.10763
## voice_type_factor1:mfccs_dist                                  -0.26139
## tempo_difference:key_difference:voice_type_factor1             -0.17997
## tempo_difference:key_difference:mfccs_dist                     -0.19361
## tempo_difference:voice_type_factor1:mfccs_dist                 -0.29154
## key_difference:voice_type_factor1:mfccs_dist                    0.01613
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist   0.12043
##                                                                Std. Error
## (Intercept)                                                       0.13470
## tempo_difference                                                  0.17459
## key_difference                                                    0.14300
## voice_type_factor1                                                0.13470
## mfccs_dist                                                        0.15202
## tempo_difference:key_difference                                   0.16240
## tempo_difference:voice_type_factor1                               0.17459
## key_difference:voice_type_factor1                                 0.14300
## tempo_difference:mfccs_dist                                       0.23150
## key_difference:mfccs_dist                                         0.12587
## voice_type_factor1:mfccs_dist                                     0.15202
## tempo_difference:key_difference:voice_type_factor1                0.16240
## tempo_difference:key_difference:mfccs_dist                        0.18003
## tempo_difference:voice_type_factor1:mfccs_dist                    0.23150
## key_difference:voice_type_factor1:mfccs_dist                      0.12587
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist     0.18003
##                                                                t value Pr(>|t|)
## (Intercept)                                                     -1.346   0.1839
## tempo_difference                                                -3.284   0.0018
## key_difference                                                  -1.379   0.1734
## voice_type_factor1                                               1.222   0.2269
## mfccs_dist                                                      -1.121   0.2674
## tempo_difference:key_difference                                 -0.959   0.3418
## tempo_difference:voice_type_factor1                             -1.442   0.1551
```

```
## key_difference:voice_type_factor1                                -1.879   0.0657
## tempo_difference:mfccs_dist                                       0.453   0.6521
## key_difference:mfccs_dist                                        -0.855   0.3963
## voice_type_factor1:mfccs_dist                                    -1.719   0.0913
## tempo_difference:key_difference:voice_type_factor1               -1.108   0.2727
## tempo_difference:key_difference:mfccs_dist                       -1.075   0.2870
## tempo_difference:voice_type_factor1:mfccs_dist                   -1.259   0.2133
## key_difference:voice_type_factor1:mfccs_dist                      0.128   0.8985
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist     0.669   0.5064
##
## (Intercept)
## tempo_difference                                                 **
## key_difference
## voice_type_factor1
## mfccs_dist
## tempo_difference:key_difference
## tempo_difference:voice_type_factor1
## key_difference:voice_type_factor1                                .
## tempo_difference:mfccs_dist
## key_difference:mfccs_dist
## voice_type_factor1:mfccs_dist                                    .
## tempo_difference:key_difference:voice_type_factor1
## tempo_difference:key_difference:mfccs_dist
## tempo_difference:voice_type_factor1:mfccs_dist
## key_difference:voice_type_factor1:mfccs_dist
## tempo_difference:key_difference:voice_type_factor1:mfccs_dist
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7313 on 54 degrees of freedom
## Multiple R-squared:  0.5815, Adjusted R-squared:  0.4652
## F-statistic: 5.002 on 15 and 54 DF,  p-value: 5.322e-06
```

Is the benefit of better R^2 worth all of these degrees of freedom? Use ANOVA to test the nested models.

```
## Analysis of Variance Table
##
## Model 1: mean_sim ~ tempo_difference + key_difference + voice_type_factor +
##     mfccs_dist
## Model 2: mean_sim ~ tempo_difference * key_difference * voice_type_factor *
##     mfccs_dist
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     65 38.822
## 2     54 28.878 11    9.9443 1.6905 0.1006
```

All of these interactions are not worth it ($p = 0.10$).

Without genre, there are marginally significant interactions are among key, voice type, and timbre.

What if we separate tempo and just let the other features interact?

```
##
## Call:
## lm(formula = mean_sim ~ tempo_difference + key_difference * voice_type_factor *
##     mfccs_dist, data = musical_diffs_scaled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82231 -0.50140  0.08851  0.46115  1.18612
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                -0.06846    0.10630  -0.644
## tempo_difference                           -0.57908    0.10056  -5.759
## key_difference                             -0.09452    0.10829  -0.873
## voice_type_factor1                          0.08508    0.10798   0.788
## mfccs_dist                                 -0.09019    0.12054  -0.748
## key_difference:voice_type_factor1          -0.25654    0.10725  -2.392
## key_difference:mfccs_dist                  -0.11799    0.10125  -1.165
## voice_type_factor1:mfccs_dist              -0.25383    0.11973  -2.120
## key_difference:voice_type_factor1:mfccs_dist  0.14761    0.09571   1.542
##                                            Pr(>|t|)
## (Intercept)                                  0.5220
## tempo_difference                           2.96e-07 ***
## key_difference                               0.3862
## voice_type_factor1                           0.4338
## mfccs_dist                                   0.4572
## key_difference:voice_type_factor1            0.0199 *
## key_difference:mfccs_dist                    0.2485
## voice_type_factor1:mfccs_dist                0.0381 *
## key_difference:voice_type_factor1:mfccs_dist   0.1282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7219 on 61 degrees of freedom
## Multiple R-squared:  0.5393, Adjusted R-squared:  0.4789
## F-statistic: 8.926 on 8 and 61 DF,  p-value: 5.275e-08
```

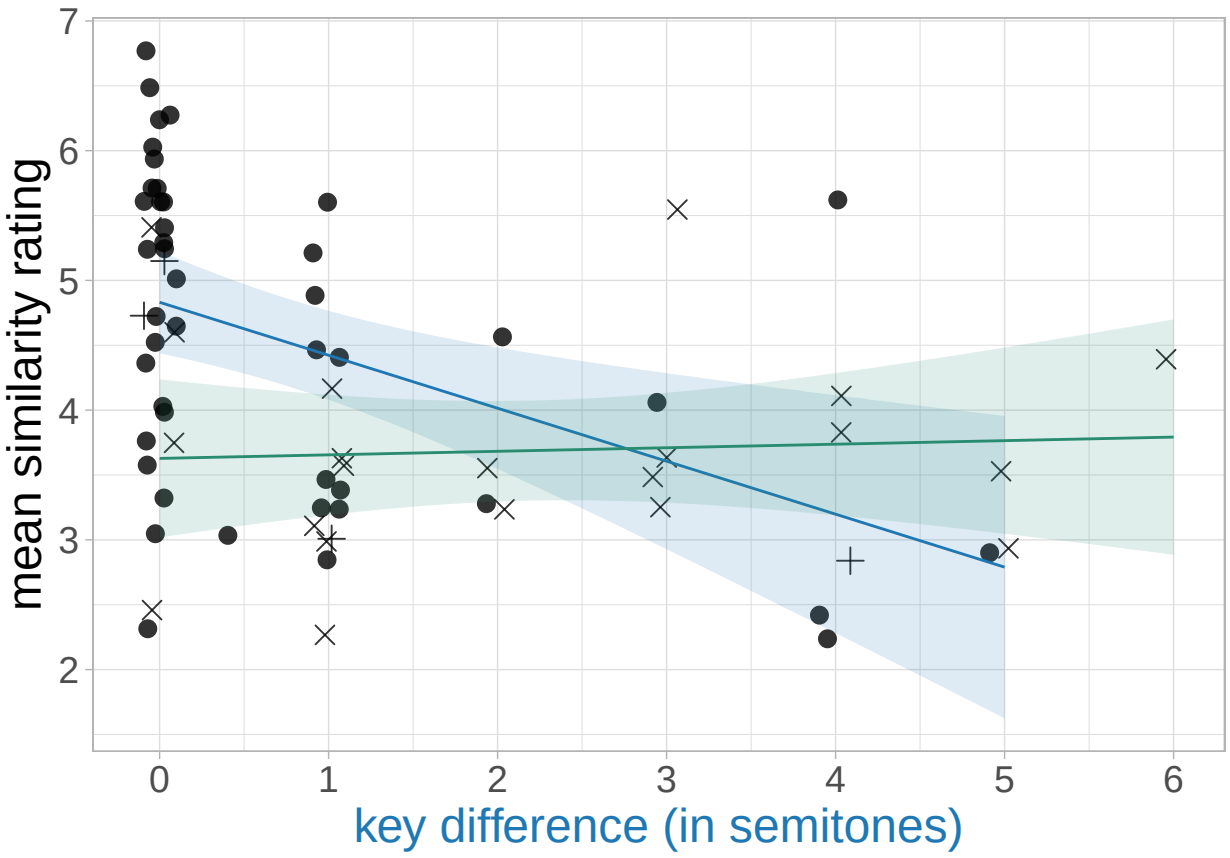
Are these interactions worth it?

```
## Analysis of Variance Table
##
## Model 1: mean_sim ~ tempo_difference + key_difference + voice_type_factor +
##     mfccs_dist
## Model 2: mean_sim ~ tempo_difference + key_difference * voice_type_factor *
##     mfccs_dist
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     65 38.822
## 2     61 31.787  4    7.0348 3.3749 0.01467 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, adding these interactions adds sufficient explanatory benefit to merit including them.

Visualize interactions

Key and voice type



Timbre and voice type