Mediation analysis

R. Cassano-Coleman

2025-01-13

This notebook performs the necessary modeling for the mediation analysis (musical features mediating relationships between emotion and similarity).

We confirmed in the other notebook `emotional_features.Rmd` that both valence and arousal (separately) predict similarity. In this notebook we add more steps to do a mediation analysis:

1. We find which of the musical features are predicted by valence or arousal. As far as interpretations go this is backwards (emotion -> music), but that's ok because the regression is symmetric. If the musical feature is predicted by the emotional feature, then it is a candidate to be a mediator.
2. We then build a model with the emotional feature and the potential mediator(s).
3. We compare the emotion only model to emotion + music model. If the significance of the emotion coefficient disappears completely, then it is a full mediation. If it disappears partially, then it is a partial mediation.

Each of these steps is conducted with valence and arousal separately.

Load the data.

Keep the musical and emotional features.

Valence

```
##
## Call:
## lm(formula = mean_sim ~ mean_valence_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6144 -0.5870 -0.1789  0.6614  1.8885
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.363e-16  1.004e-01   0.000        1
## mean_valence_difference -5.519e-01  1.011e-01  -5.457 7.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.84 on 68 degrees of freedom
## Multiple R-squared:  0.3045, Adjusted R-squared:  0.2943
## F-statistic: 29.78 on 1 and 68 DF,  p-value: 7.348e-07
```

Potential mediators

```
##
## Call:
## lm(formula = tempo_difference ~ mean_valence_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9627 -0.5191 -0.3352  0.4114  2.9400
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -3.455e-17  1.110e-01    0.00 1.000000
## mean_valence_difference  3.879e-01  1.118e-01    3.47 0.000908 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9285 on 68 degrees of freedom
## Multiple R-squared:  0.1505, Adjusted R-squared:  0.138
## F-statistic: 12.04 on 1 and 68 DF,  p-value: 0.0009076
```

```
##
## Call:
## lm(formula = key_difference ~ mean_valence_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2846 -0.6876 -0.5336  0.4414  2.6648
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.196e-16  1.185e-01   0.000     1.00
## mean_valence_difference  1.781e-01  1.193e-01   1.493     0.14
##
```

```
## Residual standard error: 0.9912 on 68 degrees of freedom
## Multiple R-squared:  0.03173,    Adjusted R-squared:  0.01749
## F-statistic: 2.228 on 1 and 68 DF,  p-value: 0.1401
```

```
##
## Call:
## lm(formula = voice_type_difference ~ mean_valence_difference,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0053 -0.7001 -0.6345  1.3263  1.5755
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -2.885e-19  1.198e-01   0.000    1.000
## mean_valence_difference  9.648e-02  1.207e-01   0.799    0.427
##
## Residual standard error: 1.003 on 68 degrees of freedom
## Multiple R-squared:  0.009309,   Adjusted R-squared:  -0.00526
## F-statistic: 0.6389 on 1 and 68 DF,  p-value: 0.4269
```

```
##
## Call:
## lm(formula = mfccs_dist ~ mean_valence_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0022 -0.7276 -0.2768  0.5609  2.6190
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              3.951e-16  1.202e-01   0.000     1.00
## mean_valence_difference 6.033e-02  1.210e-01   0.498     0.62
##
## Residual standard error: 1.005 on 68 degrees of freedom
## Multiple R-squared:  0.00364,    Adjusted R-squared:  -0.01101
## F-statistic: 0.2484 on 1 and 68 DF,  p-value: 0.6198
```

Tempo is the only musical feature with a relationship with valence, so it is the only potential mediator.

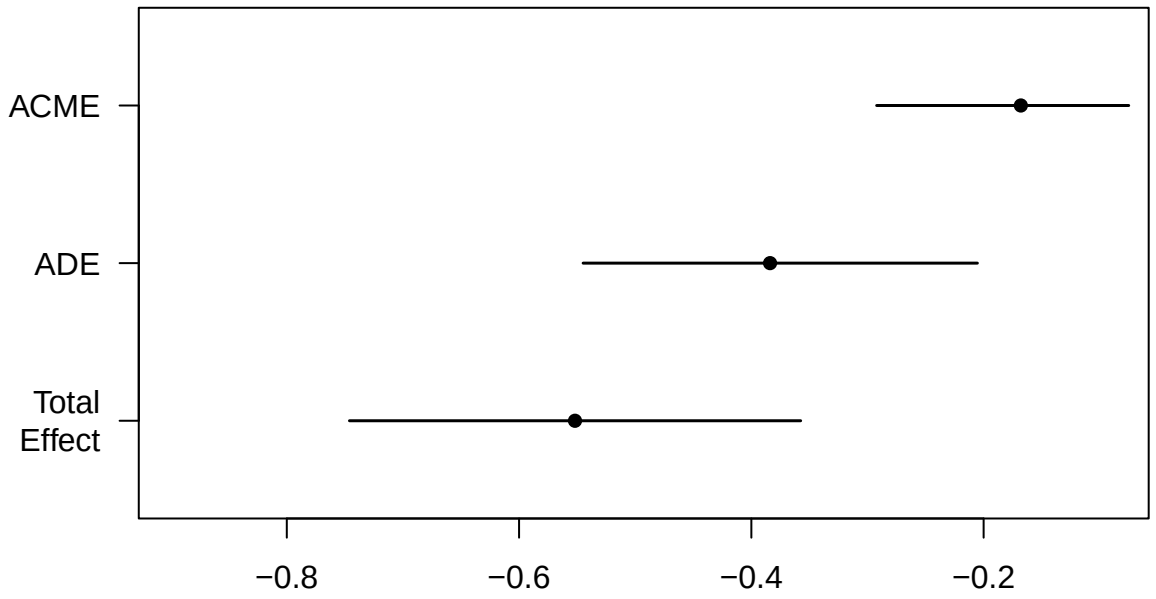
```
##
## Call:
## lm(formula = tempo_difference ~ mean_valence_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9627 -0.5191 -0.3352  0.4114  2.9400
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -3.455e-17  1.110e-01    0.00 1.000000
## mean_valence_difference  3.879e-01  1.118e-01    3.47 0.000908 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.9285 on 68 degrees of freedom
## Multiple R-squared:  0.1505, Adjusted R-squared:  0.138
## F-statistic: 12.04 on 1 and 68 DF,  p-value: 0.0009076
```

Mediation

```
##
## Call:
## lm(formula = mean_sim ~ mean_valence_difference + tempo_difference,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.62884 -0.52130  0.04045  0.50548  1.65313
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.513e-16  8.881e-02   0.000 1.000000
## mean_valence_difference -3.839e-01  9.705e-02  -3.955 0.000187 ***
## tempo_difference        -4.330e-01  9.705e-02  -4.462 3.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7431 on 67 degrees of freedom
## Multiple R-squared:  0.4638, Adjusted R-squared:  0.4478
## F-statistic: 28.98 on 2 and 67 DF,  p-value: 8.535e-10
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
## ACME             -0.168       -0.292        -0.08  <2e-16 ***
## ADE              -0.384       -0.545        -0.21  <2e-16 ***
## Total Effect     -0.552       -0.746        -0.36  <2e-16 ***
## Prop. Mediated    0.304        0.158         0.53  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 70
##
##
## Simulations: 500
```



ACME: Average causal mediation effects ADE: Average direct effects

Note that total effect is the same as the coefficient for the valence only model (`mean_sim ~ mean_valence_difference`).

Since ACME is different from zero, there is at least some mediation. However, the ADE is also different from zero, so the mediation is only partial. (If the ADE was zero, then it would be full mediation).

Arousal

```
##
## Call:
## lm(formula = mean_sim ~ mean_arousal_difference, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.91540 -0.48391 -0.00222  0.78581  1.78578
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -3.633e-16  1.075e-01    0.00        1
## mean_arousal_difference -4.504e-01  1.083e-01   -4.16 9.13e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8994 on 68 degrees of freedom
## Multiple R-squared:  0.2029, Adjusted R-squared:  0.1912
## F-statistic: 17.31 on 1 and 68 DF,  p-value: 9.131e-05
```

Potential mediators

```
##
## Call:
## lm(formula = tempo_difference ~ mean_arousal_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3011 -0.5481 -0.3076  0.4985  3.4359
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.011e-16  1.073e-01   0.000        1
## mean_arousal_difference  4.533e-01  1.081e-01   4.194 8.11e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8979 on 68 degrees of freedom
## Multiple R-squared:  0.2055, Adjusted R-squared:  0.1938
## F-statistic: 17.59 on 1 and 68 DF,  p-value: 8.113e-05
```

```
##
## Call:
## lm(formula = key_difference ~ mean_arousal_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3499 -0.6873 -0.4587  0.2569  2.5711
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.482e-16  1.182e-01   0.000    1.000
## mean_arousal_difference  1.908e-01  1.190e-01   1.603    0.114
##
```

```
## Residual standard error: 0.9888 on 68 degrees of freedom
## Multiple R-squared:  0.03639,    Adjusted R-squared:  0.02222
## F-statistic: 2.568 on 1 and 68 DF,  p-value: 0.1137
```

```
##
## Call:
## lm(formula = voice_type_difference ~ mean_arousal_difference,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7185 -0.7182 -0.7181  1.4675  1.4677
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.302e-18  1.204e-01   0.000    1.000
## mean_arousal_difference  1.037e-04  1.213e-01   0.001    0.999
##
## Residual standard error: 1.007 on 68 degrees of freedom
## Multiple R-squared:  1.075e-08,  Adjusted R-squared:  -0.01471
## F-statistic: 7.308e-07 on 1 and 68 DF,  p-value: 0.9993
```

```
##
## Call:
## lm(formula = mfccs_dist ~ mean_arousal_difference, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7573 -0.5964 -0.2386  0.5492  2.6315
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              3.530e-16  1.125e-01   0.000  1.00000
## mean_arousal_difference 3.561e-01  1.133e-01   3.142  0.00248 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9413 on 68 degrees of freedom
## Multiple R-squared:  0.1268, Adjusted R-squared:  0.1139
## F-statistic: 9.874 on 1 and 68 DF,  p-value: 0.002484
```

Tempo and timbre are the musical features with a relationship with arousal, so these are the potential mediators.

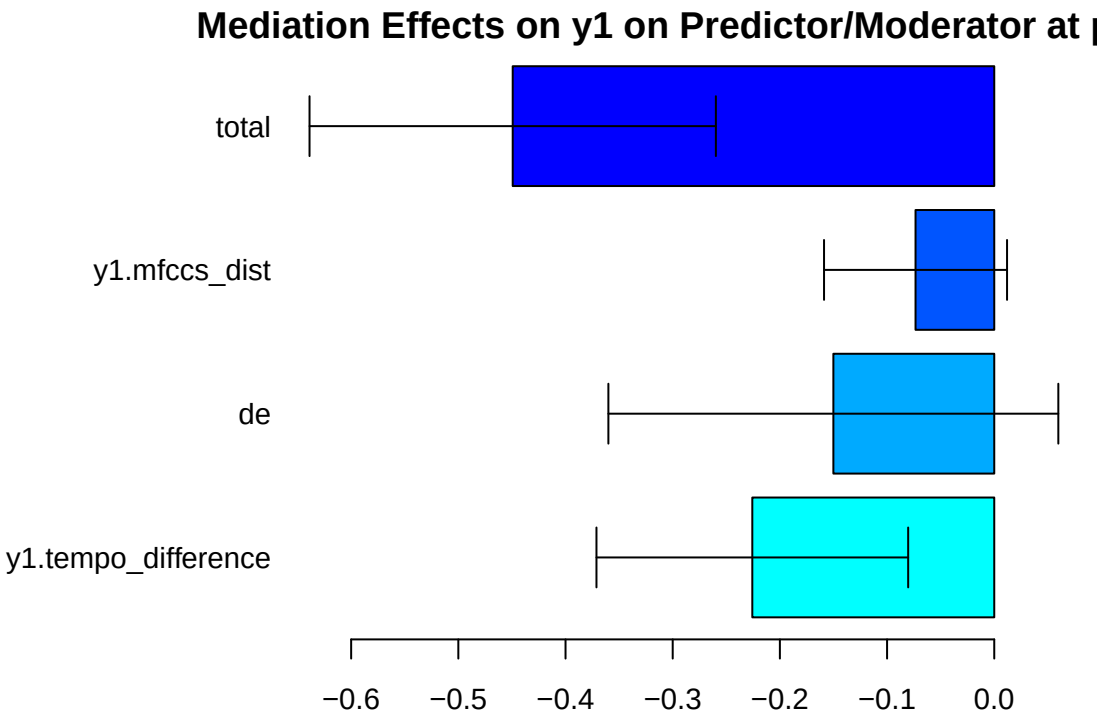
Mediation

```
##
## Call:
## lm(formula = mean_sim ~ mean_arousal_difference + tempo_difference +
##     mfccs_dist, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6946 -0.5084  0.1322  0.5956  1.3830
##
## Coefficients:
```

```
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -3.410e-16  9.315e-02   0.000   1.0000
## mean_arousal_difference -1.521e-01  1.128e-01  -1.348   0.1823
## tempo_difference        -4.967e-01  1.058e-01  -4.696 1.39e-05 ***
## mfccs_dist              -2.054e-01  1.009e-01  -2.036   0.0458 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7794 on 66 degrees of freedom
## Multiple R-squared:  0.419,  Adjusted R-squared:  0.3926
## F-statistic: 15.86 on 3 and 66 DF,  p-value: 7.158e-08
```

Because we have two potential mediators, we have to use a different package (`mma`).

```
## MMA Analysis: Estimated Mediation Effects Using GLM
## For Predictor/Moderator at pred
## $total.effect
##       est      mean        sd      upbd      lwbd    upbd_q    lwbd_q upbd_bcbi
##    -0.463    -0.449     0.097    -0.260    -0.639    -0.271    -0.629    -0.323
## lwbd_bcbi    upbd_b    lwbd_b  upbd_win  lwbd_win    p_norm    p_quan
##    -0.634    -0.227    -0.640    -0.271    -0.629     0.000     0.000
##
## $direct.effect
##       est      mean        sd      upbd      lwbd    upbd_q    lwbd_q upbd_bcbi
##    -0.168    -0.150     0.107     0.060    -0.360     0.066    -0.322     0.012
## lwbd_bcbi    upbd_b    lwbd_b  upbd_win  lwbd_win    p_norm    p_quan
##    -0.445     0.021    -0.324     0.066    -0.330     0.161     0.200
##
## $indirect.effect
##           y1.all y1.tempo_difference y1.mfccs_dist
## est       -0.295              -0.215        -0.081
## mean      -0.299              -0.226        -0.073
## sd         0.099               0.074         0.044
## upbd      -0.105              -0.080         0.012
## lwbd      -0.494              -0.371        -0.159
## upbd_q    -0.121              -0.095        -0.007
## lwbd_q    -0.452              -0.370        -0.149
## upbd_bcbi -0.103              -0.087        -0.008
## lwbd_bcbi -0.431              -0.367        -0.163
## upbd_b    -0.090              -0.063         0.023
## lwbd_b    -0.455              -0.373        -0.181
## upbd_win  -0.113              -0.105        -0.007
## lwbd_win  -0.442              -0.394        -0.151
## p_norm     0.003               0.002         0.092
## p_quan     0.000               0.000         0.040
```



In this case the indirect effect (of tempo and timbre) is different from zero, and the direct effect is not different from zero. Therefore, tempo and timbre fully mediate the relationship between arousal and similarity.