Variance partitioning

R. Cassano-Coleman

2025-01-13

This notebook uses the musical and emotional features and performs variance partitioning to understand how these sets of features work together to explain similarity.

Load the data.

Keep the musical and emotional features.

With help from: https://www.davidzeleny.net/anadat-r/doku.php/en:varpart

Venn diagram parts:

- a: variance explained by musical features only
- b: variance explained by musical and emotional features together
- c: variance explained by emotional features only
- d: unexplained variance

For the musical features, we'll allow key, voice type, and MFCCs to interact because that was the best musical features model (see `musical_features.Rmd`).

Calculate manually

[a+b+c]

```
##
## Call:
## rda(formula = data$mean_sim ~ tempo_difference + key_difference *      voice_type_difference * mfccs
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.6426     0.6426
## Unconstrained  0.3574     0.3574
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.6426 0.3574
## Proportion Explained  0.6426 0.3574
## Cumulative Proportion 0.6426 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                           RDA1
## Eigenvalue            0.6426
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

Is this consistent with just a simple linear model?

Yes, the explained variance is 0.5821, which is similar to the "constrained" variance in the RDA model. For RDA, it seems like "unconstrained" variance is [d].

Model for musical features only [a+b]

```
##
## Call:
## rda(formula = data$mean_sim ~ tempo_difference + key_difference *      voice_type_difference * mfccs
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.5607     0.5607
## Unconstrained  0.4393     0.4393
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.5607 0.4393
## Proportion Explained  0.5607 0.4393
## Cumulative Proportion 0.5607 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                           RDA1
```

```
## Eigenvalue            0.5607
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

Model for emotional features only [b+c]

```
##
## Call:
## rda(formula = data$mean_sim ~ mean_valence_difference + mean_arousal_difference,      data = data)
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.3944     0.3944
## Unconstrained  0.6056     0.6056
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.3944 0.6056
## Proportion Explained  0.3944 0.6056
## Cumulative Proportion 0.3944 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                           RDA1
## Eigenvalue            0.3944
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

Extract the R^2 values.

Subtract to find individual fractions.

- variance explained by musical features only (a): 20.58%
- shared variance explained (b): 29.74%
- variance explained by emotional features only (c): 7.89%
- unexplained variance (d): 41.79%

Musical features

- a: variance explained by tempo only
- b: shared variance explained by the musical features
- c: variance explained by key, voice type, and MFCCs only
- d: unexplained variance

[a+b+c]

```
##
## Call:
## rda(formula = data$mean_sim ~ tempo_difference + key_difference *      voice_type_difference * mfccs_
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.5607     0.5607
## Unconstrained  0.4393     0.4393
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.5607 0.4393
## Proportion Explained  0.5607 0.4393
## Cumulative Proportion 0.5607 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                           RDA1
## Eigenvalue            0.5607
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

[a+b]

```
##
## Call:
## rda(formula = data$mean_sim ~ tempo_difference, data = data)
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.3386     0.3386
## Unconstrained  0.6614     0.6614
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.3386 0.6614
## Proportion Explained  0.3386 0.6614
## Cumulative Proportion 0.3386 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
```

```
##                          RDA1
## Eigenvalue            0.3386
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

[b+c]

```
##
## Call:
## rda(formula = data$mean_sim ~ key_difference * voice_type_difference *      mfccs_dist, data = data)
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.3174     0.3174
## Unconstrained  0.6826     0.6826
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                          RDA1    PC1
## Eigenvalue            0.3174 0.6826
## Proportion Explained  0.3174 0.6826
## Cumulative Proportion 0.3174 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                          RDA1
## Eigenvalue            0.3174
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

Extract the R^2 values.

Subtract to find individual fractions.

- variance explained by tempo only (a): 26.28%
- shared variance (b): 6.61%
- variance explained by key, voice type, and timbre only (c): 17.42%
- unexplained variance (d): 49.69%

Emotional features

- a: variance explained by valence only
- b: shared variance explained by the emotional features
- c: variance explained by arousal only
- d: unexplained variance

[a+b+c]

```
##
## Call:
## rda(formula = data$mean_sim ~ mean_valence_difference + mean_arousal_difference,      data = data)
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.3944     0.3944
## Unconstrained  0.6056     0.6056
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.3944 0.6056
## Proportion Explained  0.3944 0.6056
## Cumulative Proportion 0.3944 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                           RDA1
## Eigenvalue            0.3944
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

[a+b]

```
##
## Call:
## rda(formula = data$mean_sim ~ mean_valence_difference, data = data)
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.3045     0.3045
## Unconstrained  0.6955     0.6955
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                           RDA1    PC1
## Eigenvalue            0.3045 0.6955
## Proportion Explained  0.3045 0.6955
## Cumulative Proportion 0.3045 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
```

```
##                          RDA1
## Eigenvalue            0.3045
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

[b+c]

```
##
## Call:
## rda(formula = data$mean_sim ~ mean_arousal_difference, data = data)
##
## Partitioning of variance:
##               Inertia Proportion
## Total          1.0000     1.0000
## Constrained    0.2029     0.2029
## Unconstrained  0.7971     0.7971
##
## Eigenvalues, and their contribution to the variance
##
## Importance of components:
##                          RDA1    PC1
## Eigenvalue            0.2029 0.7971
## Proportion Explained  0.2029 0.7971
## Cumulative Proportion 0.2029 1.0000
##
## Accumulated constrained eigenvalues
## Importance of components:
##                          RDA1
## Eigenvalue            0.2029
## Proportion Explained  1.0000
## Cumulative Proportion 1.0000
```

Extract the R^2 values.

Subtract to find individual fractions.

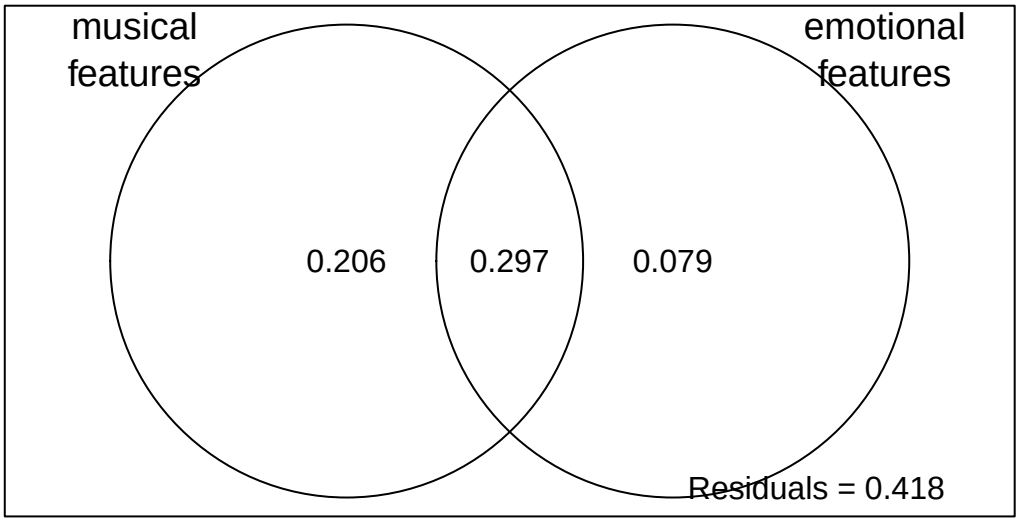
- variance explained by valence only (a): 18.51%
- shared variance by the emotional features (b): 10.92%
- variance explained by arousal only (c): 8.20%
- unexplained variance (d): 62.37%

Using `varpart`

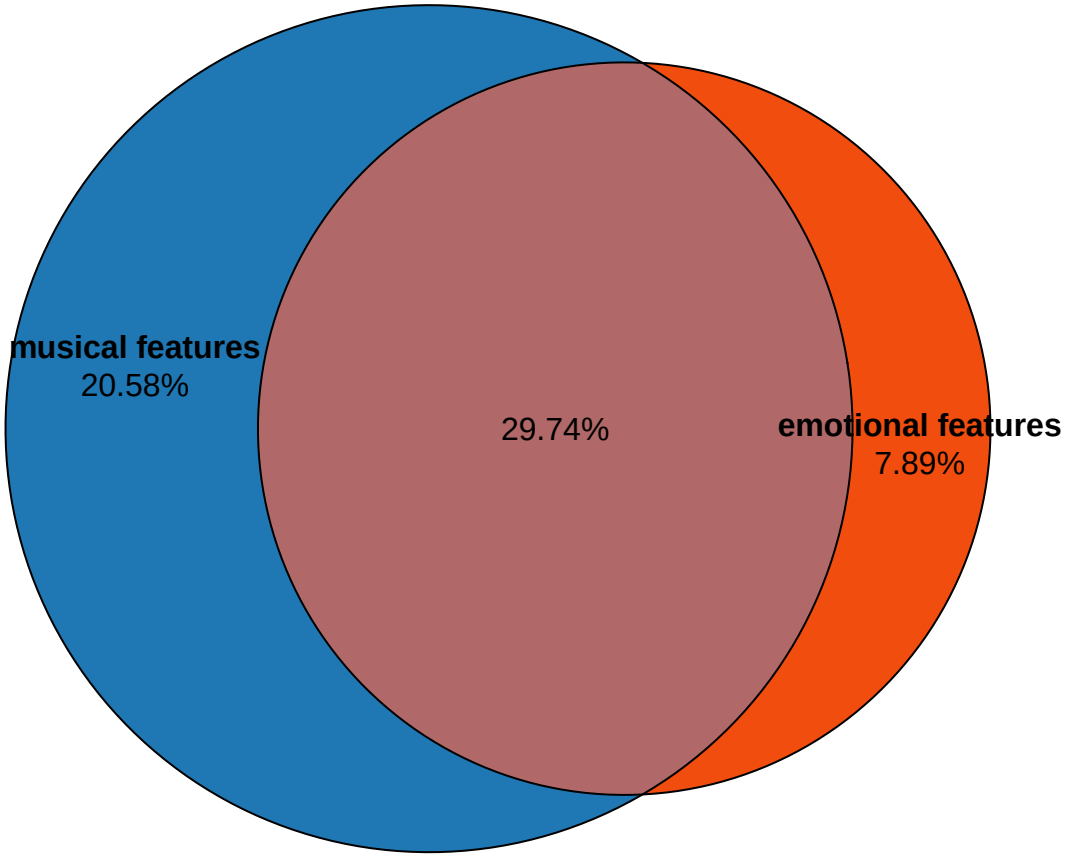
Musical and emotional features as a set

```
##
## Partition of variance in RDA
##
## Call: varpart(Y = data$mean_sim, X = ~tempo_difference + key_difference
## * voice_type_difference * mfccs_dist, ~mean_valence_difference +
## mean_arousal_difference, data = data)
##
## Explanatory tables:
## X1:  ~tempo_difference + key_difference * voice_type_difference * mfccs_dist
## X2:  ~mean_valence_difference + mean_arousal_difference
##
## No. of explanatory tables: 2
## Total variation (SS): 69
##              Variance: 1
## No. of observations: 70
##
## Partition table:
##                       Df R.squared Adj.R.squared Testable
## [a+c] = X1             8   0.56074       0.50313     TRUE
## [b+c] = X2             2   0.39437       0.37629     TRUE
## [a+b+c] = X1+X2       10   0.64264       0.58207     TRUE
## Individual fractions
## [a] = X1|X2            8                 0.20578     TRUE
## [b] = X2|X1            2                 0.07894     TRUE
## [c]                    0                 0.29736    FALSE
## [d] = Residuals                          0.41793    FALSE
## ---
## Use function 'rda' to test significance of fractions of interest
```

Plot the results with a Venn diagram



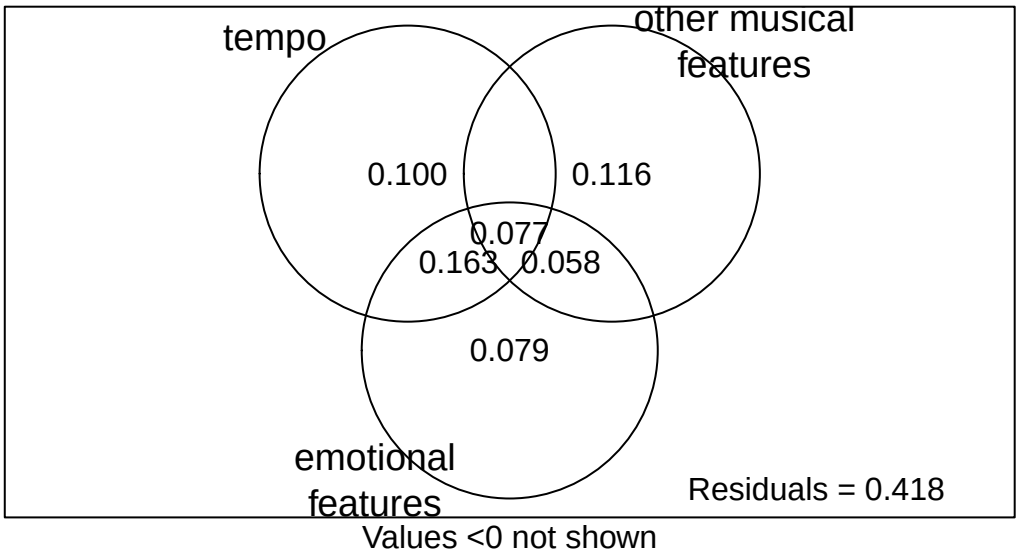
Plot an area-proportional Venn diagram

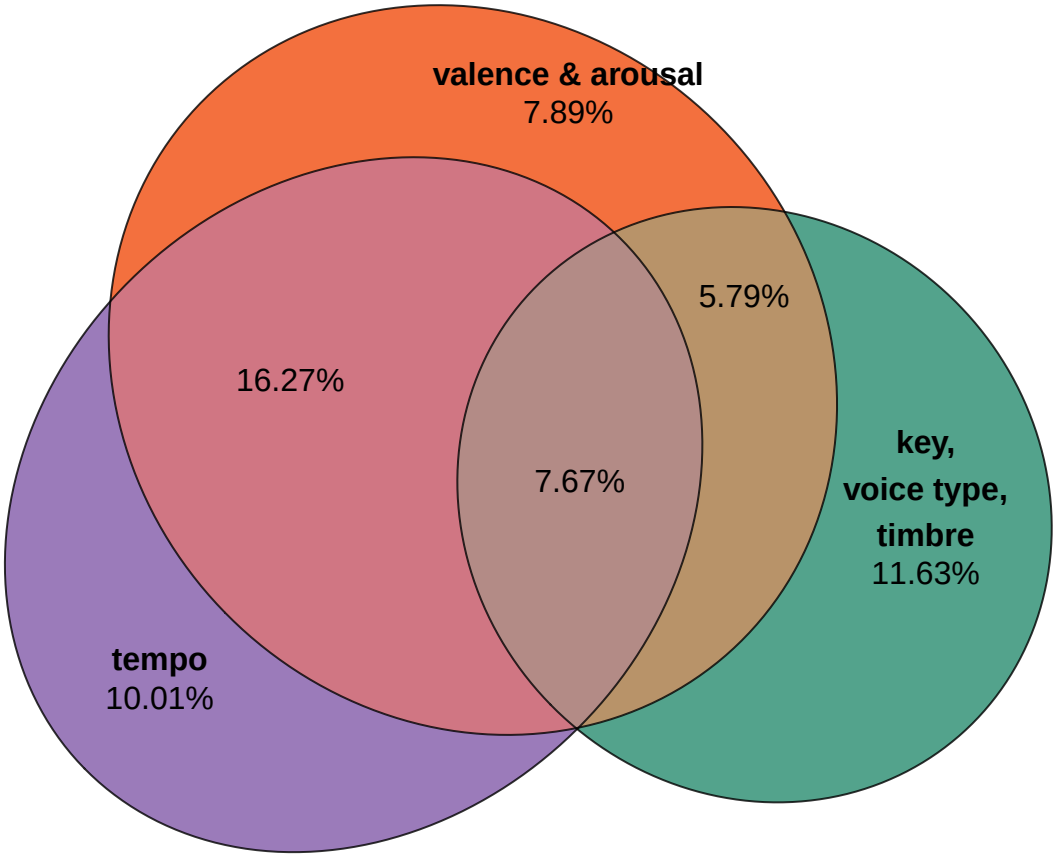


Separating into three sets (tempo, all other musical features, emotional features)

```
##
## Partition of variance in RDA
##
## Call: varpart(Y = data$mean_sim, X = ~tempo_difference, ~key_difference
## * voice_type_difference * mfccs_dist, ~mean_valence_difference +
## mean_arousal_difference, data = data)
##
## Explanatory tables:
## X1:  ~tempo_difference
## X2:  ~key_difference * voice_type_difference * mfccs_dist
## X3:  ~mean_valence_difference + mean_arousal_difference
##
## No. of explanatory tables: 3
## Total variation (SS): 69
##              Variance: 1
## No. of observations: 70
##
## Partition table:
##                         Df R.square Adj.R.square Testable
## [a+d+f+g] = X1           1  0.33864      0.32892     TRUE
## [b+d+e+g] = X2           7  0.31736      0.24029     TRUE
## [c+e+f+g] = X3           2  0.39437      0.37629     TRUE
## [a+b+d+e+f+g] = X1+X2    8  0.56074      0.50313     TRUE
## [a+c+d+e+f+g] = X1+X3    3  0.48900      0.46578     TRUE
## [b+c+d+e+f+g] = X2+X3    9  0.54953      0.48195     TRUE
```

```
## [a+b+c+d+e+f+g] = All 10  0.64264      0.58207     TRUE
## Individual fractions
## [a] = X1 | X2+X3       1               0.10012     TRUE
## [b] = X2 | X1+X3       7               0.11629     TRUE
## [c] = X3 | X1+X2       2               0.07894     TRUE
## [d]                    0              -0.01063    FALSE
## [e]                    0               0.05793    FALSE
## [f]                    0               0.16273    FALSE
## [g]                    0               0.07670    FALSE
## [h] = Residuals                        0.41793    FALSE
## Controlling 1 table X
## [a+d] = X1 | X3        1               0.08949     TRUE
## [a+f] = X1 | X2        1               0.26285     TRUE
## [b+d] = X2 | X3        7               0.10566     TRUE
## [b+e] = X2 | X1        7               0.17422     TRUE
## [c+e] = X3 | X1        2               0.13686     TRUE
## [c+f] = X3 | X2        2               0.24167     TRUE
## ---
## Use function 'rda' to test significance of fractions of interest
```



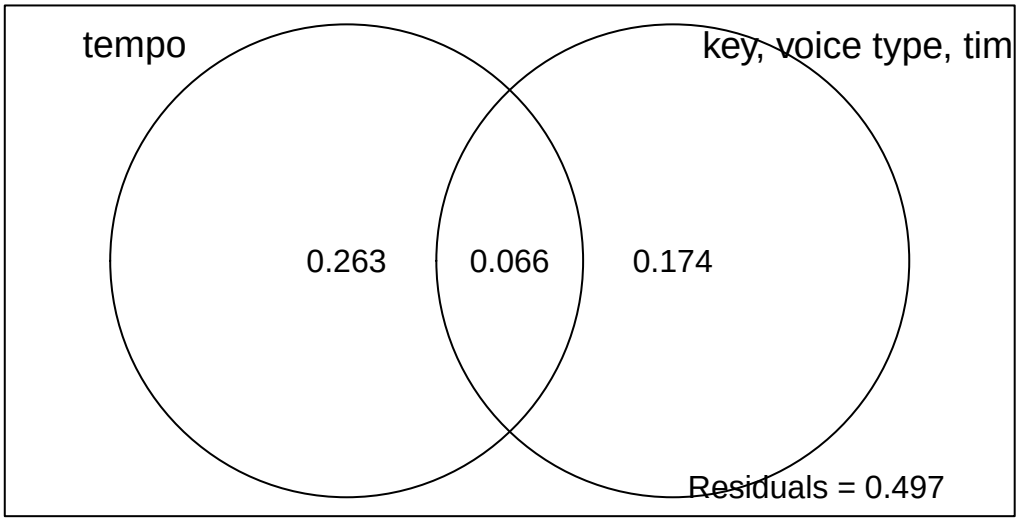

valence & arousal
7.89%
5.79%
16.27%
7.67%
key,
voice type,
timbre
11.63%
tempo
10.01%

Venn diagrams for musical features only and emotional features only.

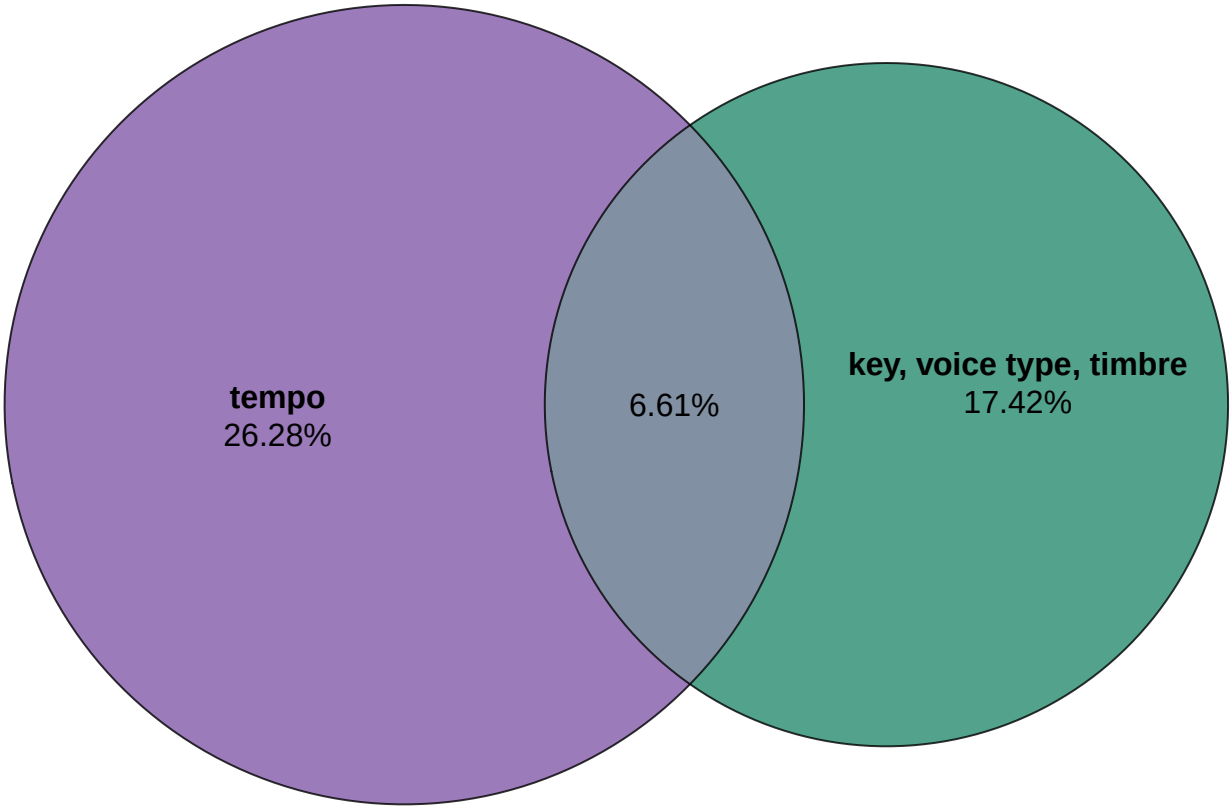
Musical features only

```
##
## Partition of variance in RDA
##
## Call: varpart(Y = data$mean_sim, X = ~tempo_difference, ~key_difference
## * voice_type_difference * mfccs_dist, data = data)
##
## Explanatory tables:
## X1:  ~tempo_difference
## X2:  ~key_difference * voice_type_difference * mfccs_dist
##
## No. of explanatory tables: 2
## Total variation (SS): 69
##              Variance: 1
## No. of observations: 70
##
## Partition table:
##                       Df R.squared Adj.R.squared Testable
## [a+c] = X1             1   0.33864       0.32892     TRUE
## [b+c] = X2             7   0.31736       0.24029     TRUE
## [a+b+c] = X1+X2        8   0.56074       0.50313     TRUE
## Individual fractions
## [a] = X1|X2            1                 0.26285     TRUE
## [b] = X2|X1            7                 0.17422     TRUE
## [c]                    0                 0.06607    FALSE
## [d] = Residuals                          0.49687    FALSE
## ---
## Use function 'rda' to test significance of fractions of interest
```

Plot the results with a Venn diagram



Plot an area-proportional Venn diagram

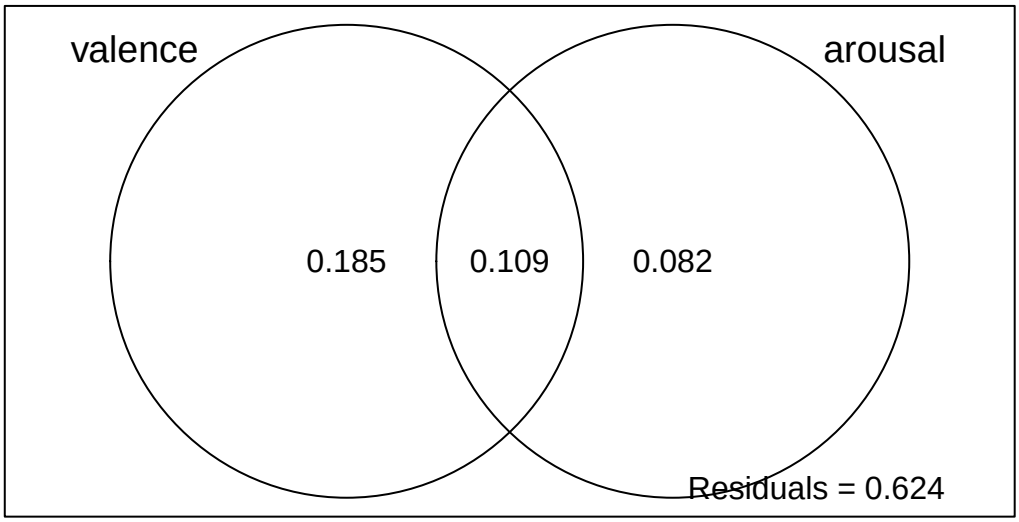


Emotional features only

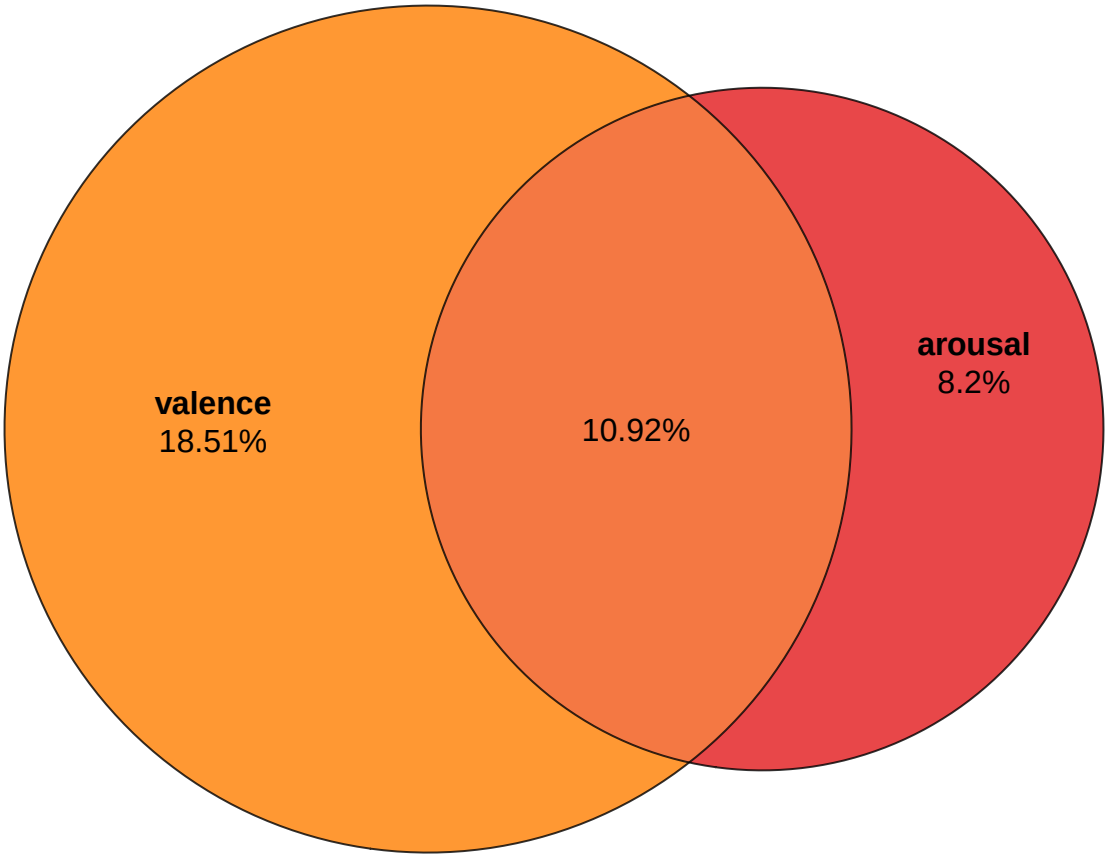
```
##
## Partition of variance in RDA
##
## Call: varpart(Y = data$mean_sim, X = ~mean_valence_difference,
## ~mean_arousal_difference, data = data)
##
## Explanatory tables:
## X1:  ~mean_valence_difference
## X2:  ~mean_arousal_difference
##
## No. of explanatory tables: 2
## Total variation (SS): 69
##              Variance: 1
## No. of observations: 70
##
## Partition table:
##                       Df R.squared Adj.R.squared Testable
## [a+c] = X1             1   0.30455       0.29432     TRUE
## [b+c] = X2             1   0.20288       0.19116     TRUE
## [a+b+c] = X1+X2        2   0.39437       0.37629     TRUE
## Individual fractions
## [a] = X1|X2            1                 0.18513     TRUE
## [b] = X2|X1            1                 0.08197     TRUE
## [c]                    0                 0.10919    FALSE
## [d] = Residuals                          0.62371    FALSE
## ---
```

```
## Use function 'rda' to test significance of fractions of interest
```

Plot the results with a Venn diagram



Plot an area-proportional Venn diagram

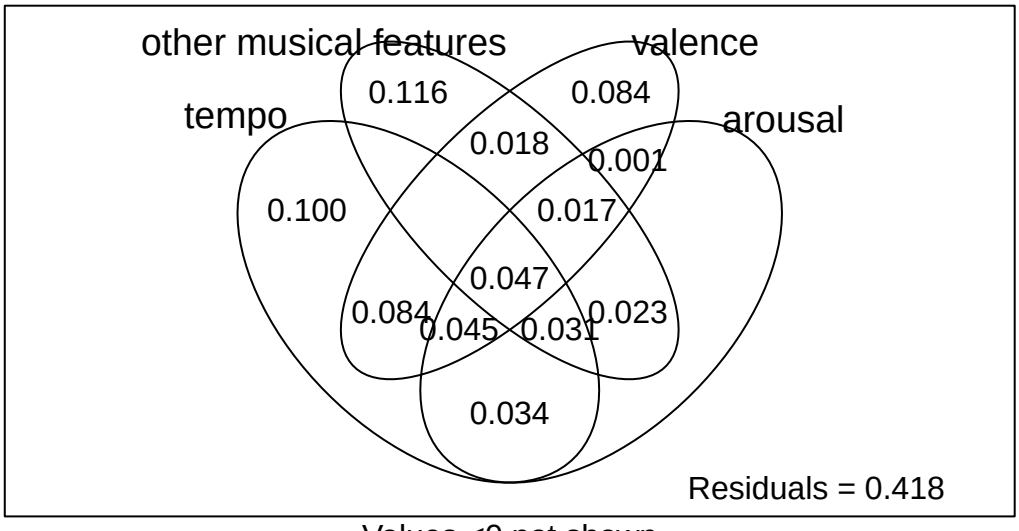


Can we do this with valence and arousal separately in a four-way model?

```
##
## Partition of variance in RDA
##
## Call: varpart(Y = data$mean_sim, X = ~tempo_difference, ~key_difference
## * voice_type_difference * mfccs_dist, ~mean_valence_difference,
## ~mean_arousal_difference, data = data)
##
## Explanatory tables:
## X1:  ~tempo_difference
## X2:  ~key_difference * voice_type_difference * mfccs_dist
## X3:  ~mean_valence_difference
## X4:  ~mean_arousal_difference
##
## No. of explanatory tables: 4
## Total variation (SS): 69
##             Variance: 1
## No. of observations: 70
##
## Partition table:
##                              Df R.square Adj.R.square Testable
## [aeghklno] = X1               1  0.33864      0.32892     TRUE
## [befiklmo] = X2               7  0.31736      0.24029     TRUE
## [cfgjlmno] = X3               1  0.30455      0.29432     TRUE
## [dhijkmno] = X4               1  0.20288      0.19116     TRUE
## [abefghiklmno] = X1+X2        8  0.56074      0.50313     TRUE
## [acefghjklmno] = X1+X3        2  0.46385      0.44784     TRUE
## [adeghijklmno] = X1+X4        2  0.38248      0.36404     TRUE
## [bcefgijklmno] = X2+X3        8  0.51711      0.45378     TRUE
## [bdefhijklmno] = X2+X4        8  0.39341      0.31386     TRUE
## [cdfghijklmno] = X3+X4        2  0.39437      0.37629     TRUE
## [abcefghijklmno] = X1+X2+X3   9  0.64124      0.58743     TRUE
## [abdefghijklmno] = X1+X2+X4   9  0.56386      0.49844     TRUE
## [acdefghijklmno] = X1+X3+X4   3  0.48900      0.46578     TRUE
## [bcdefghijklmno] = X2+X3+X4   9  0.54953      0.48195     TRUE
## [abcdefghijklmno] = All      10  0.64264      0.58207     TRUE
## Individual fractions
## [a] = X1 | X2+X3+X4           1               0.10012     TRUE
## [b] = X2 | X1+X3+X4           7               0.11629     TRUE
## [c] = X3 | X1+X2+X4           1               0.08363     TRUE
## [d] = X4 | X1+X2+X3           1              -0.00536     TRUE
## [e]                           0              -0.01063    FALSE
## [f]                           0               0.01811    FALSE
## [g]                           0               0.08447    FALSE
## [h]                           0               0.03353    FALSE
## [i]                           0               0.02329    FALSE
## [j]                           0               0.00067    FALSE
## [k]                           0               0.03051    FALSE
## [l]                           0              -0.00107    FALSE
## [m]                           0               0.01652    FALSE
## [n]                           0               0.04473    FALSE
## [o]                           0               0.04727    FALSE
## [p] = Residuals               0               0.41793    FALSE
```

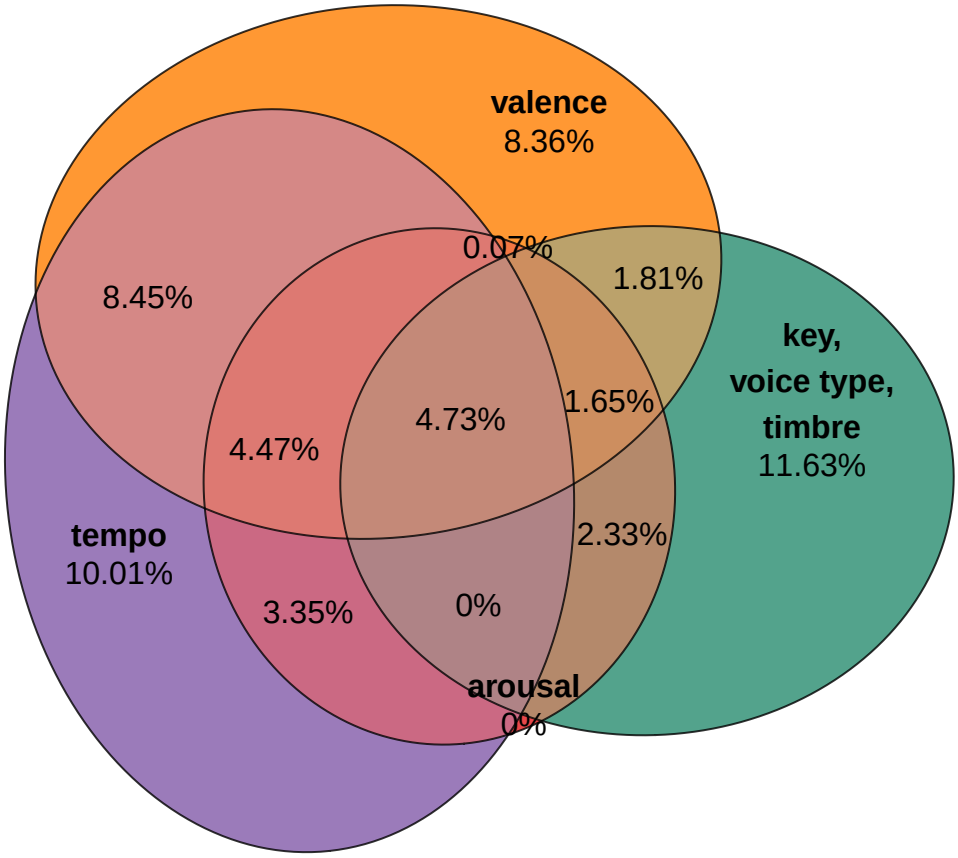
```
## Controlling 2 tables X
## [ae] = X1 | X3+X4            1              0.08949     TRUE
## [ag] = X1 | X2+X4            1              0.18459     TRUE
## [ah] = X1 | X2+X3            1              0.13364     TRUE
## [be] = X2 | X3+X4            7              0.10566     TRUE
## [bf] = X2 | X1+X4            7              0.13440     TRUE
## [bi] = X2 | X1+X3            7              0.13959     TRUE
## [cf] = X3 | X1+X4            1              0.10174     TRUE
## [cg] = X3 | X2+X4            1              0.16810     TRUE
## [cj] = X3 | X1+X2            1              0.08429     TRUE
## [dh] = X4 | X2+X3            1              0.02817     TRUE
## [di] = X4 | X1+X3            1              0.01794     TRUE
## [dj] = X4 | X1+X2            1             -0.00469     TRUE
## Controlling 1 table X
## [aghn] = X1 | X2             1              0.26285     TRUE
## [aehk] = X1 | X3             1              0.15352     TRUE
## [aegl] = X1 | X4             1              0.17288     TRUE
## [bfim] = X2 | X1             7              0.17422     TRUE
## [beik] = X2 | X3             7              0.15946     TRUE
## [befl] = X2 | X4             7              0.12270     TRUE
## [cfjm] = X3 | X1             1              0.11892     TRUE
## [cgjn] = X3 | X2             1              0.21349     TRUE
## [cfgl] = X3 | X4             1              0.18513     TRUE
## [dijm] = X4 | X1             1              0.03513     TRUE
## [dhjn] = X4 | X2             1              0.07357     TRUE
## [dhik] = X4 | X3             1              0.08197     TRUE
## ---
## Use function 'rda' to test significance of fractions of interest
```

Plot the results with a Venn diagram



Values <0 not shown

Try to plot an area-proportional Venn diagram


valence
8.36%
0.07%
1.81%
8.45%
key,
voice type,
timbre
11.63%
4.73%
1.65%
4.47%
2.33%
tempo
10.01%
3.35%
0%
arousal
0%