Many Ways of Hearing: Clustering Continuous Responses to Music

MARL
NYU Music and Audio Research Laboratory

Finn Upham

NYU Steinhardt
Steinhardt School of Culture, Education, and Human Development

Clustering emotional responses to the same music

The aim of this study is to evaluate whether continuous responses show evidence of distinct but repeatable temporal patterns of perception or experience to the same musical stimuli. In such cases as different patterns arise, there is the subsequent aim of evaluating the degree of difference and inform future discussion on the quantification of similarity and difference between individual continuous responses to music.

32 experimental collections of emotion ratings to music and 32 random collections were clustered using three versions of each: the ratings in their origional form, sampled at 1Hz, the ratings post lowpass filtering with cutoff of 0.1 Hz (zero phase), and the first order difference of the filtered responses after being down sampled to 0.2Hz. (see figure 1.)

These versions of the collections were then clustered hierarchically using pairwise distances of euclidean distance or Pearson correlations, and complete linkage. From the hierarchical clustering, clusters were extracted from the lowest cut which yielded two clusters each containing at least a quarter of the collections responses. Note that over the 32 experimental collections and the 32 random collections (see Data box), some collections failed to yield two such clusters under some clustering paradigms.

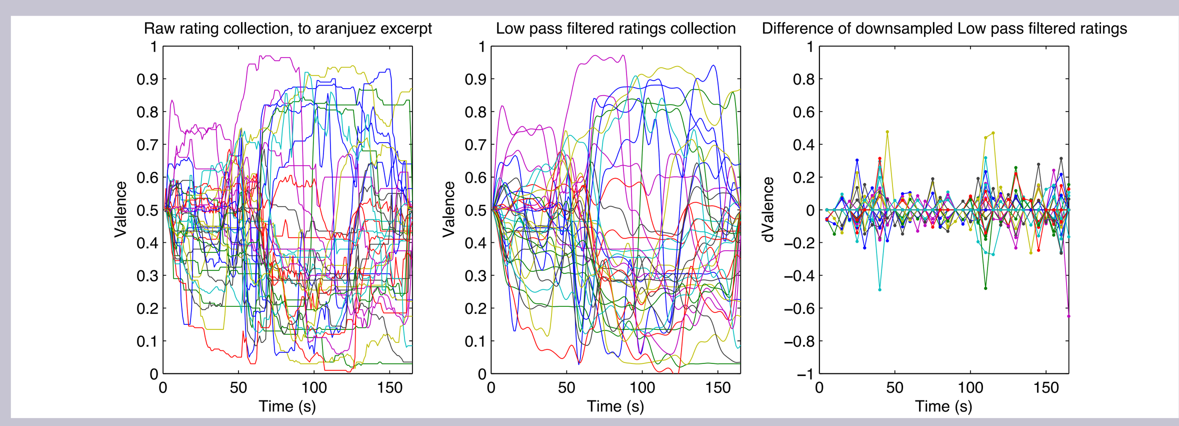

Figure 1: The collection erceived valence ratings to orchestral excerpt composed by Rodrigo, in their raw state (left), after low pass filtering (middle), and the subsequent downsampled first order difference.

The Data Collections

This project includes analysis of many collections of continuous responses to music. The experimental collections are thirty two sets of emotion ratings on a single dimension to a single musical stimulus. Though they come from several experiments, each set is sampled at 1 Hz, on a scale ranging from 0 to 1, and average 30 responses per collection. Twelve collections are of emotion perceived in the music, half valence, half arousal (collected concurrently), twelve are of emotion felt by the participants, half valence, half arousal, and eight contain ratings of emotional intensity. The stimuli are all concert music pieces, mostly of the classical and nationalistic eras.

For comparison, 32 random response collections were constructed from these experimental data sets by sampling randomly across collections. The resulting collections are composed of responses in different measures and of different stimuli, matching each experimental collection in number of responses. Note these collections are truncated to the shortest response included. These collections give a clue as to how unrelated response collection would perform under the same treatment, a check in inferring too much from the experimental data collections.

Clusters Coherence and Separation

Following the clustering, the clusters were the assessed for internal cohesion and separation from their within collection pair. Figure 2 shows two examples of the cohesion measure results using cohesion measures in forms alternate to the clustering system: correlation based cohesion for Euclidean built clusters and vis versa. The top graph shows many experimental showing higher intra-cluster correlation than the experimental responses (note these average r value are low because the information is very sparse in these series). The bottom graph shows a similar story with clustering by correlations have a consequence for some of having tighters clusters on the rating range as well. Keep in mind, though, that several experimental collections do not shore greater cross-criteria cohesion in their clusters than the random collections.

Also of importance is how different these clusters manage to be from each other. In this case, values similar to those of the random data is good, as they, by design, should contain more easily separable responses than the experimental collections. The top graph of figure 3 shows that few collections yielded clusters so different to cover the same range of low and negative correlations between the cluster means. In the bottom graph, there are only a few experimental collections have relatively high variance in the distance between cluster means. To get a sense of whether the cluster yielded any useful difference, below are select examples of experimental collections clusterings.

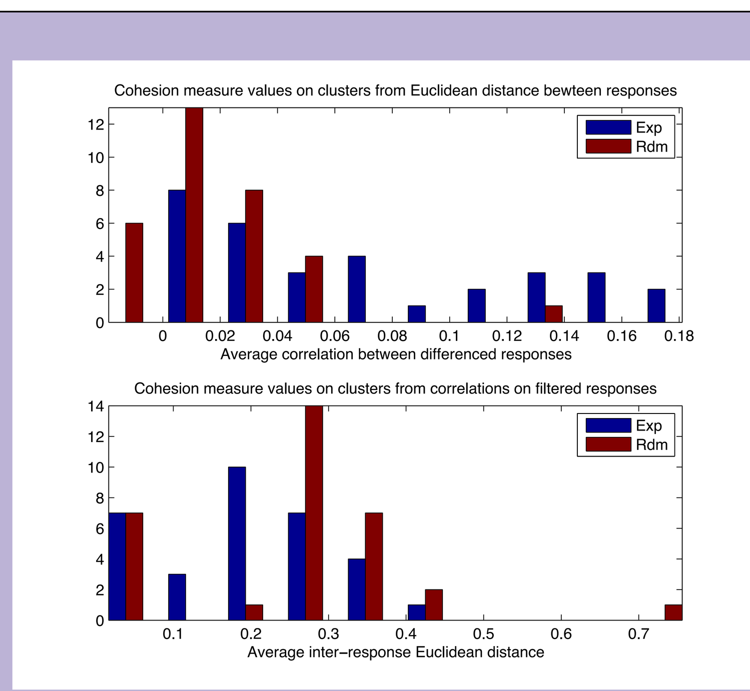

Figure 2: Figure 3: Two distributions of cohesion measures values per collection's pair of clusters, Above are the average Pearson correlations between differenced responses in each cluster on clusters formed using the euclidean distance between responses. Below are the average distances between responses per cluster on clusters formed by correlations between filtered responses

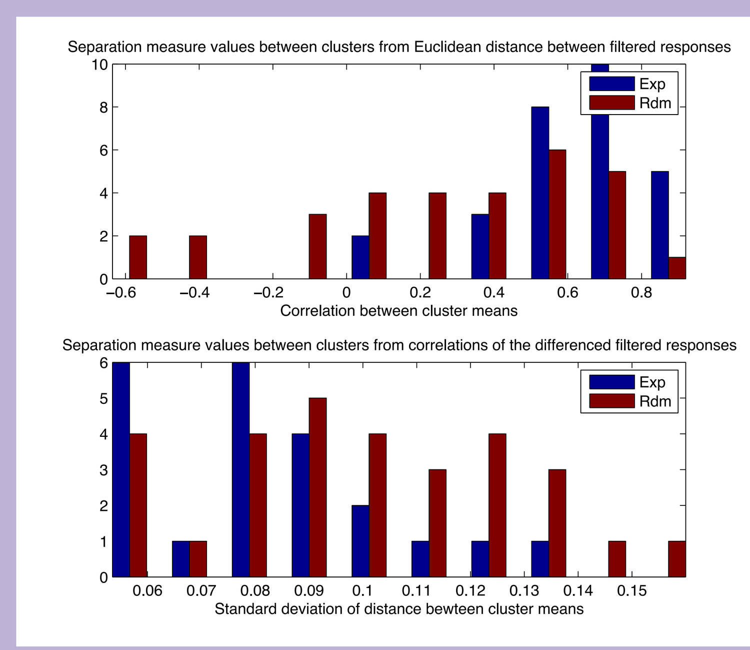

Figure 3: Two distributions of separation measures values per collection's pair of clusters,

Categorical Difference in Responses

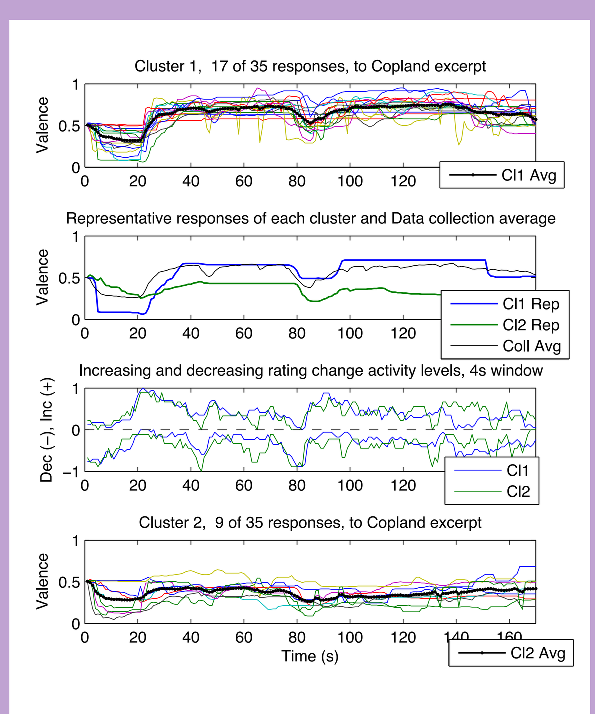

Figure 4: Clusters from euclidean distance between filtered ratings of perceived emotional valence in the beginning of Copland's Fanfare for the Common Man. The top graph is the first cluster with the average in black, second is the center most response in each cluster plotted beside the whole collections average response. Third is the rating change activity levels, which reports the proportion of each cluster changing ratings over in the 4 second sliding time window, and last is the second cluster of responses and the average rating.

The clustering in this example separates primarily differences of rating range use. Valence is a bipolar scale, often reported using an interface which marks the border between the positive and negative rating range. Here, nearly half of the participants report perceiving this piece as being mostly of positive valence, while another quarter reports mostly negative valence. While their contours and activity are pretty similar (with possible interesting differences around seconds 63 and 98) this difference could be interpreted as categorical. This supports the idea that valence in music is often ambiguous and context and listener dependent.

Sensitivity Difference in Responses

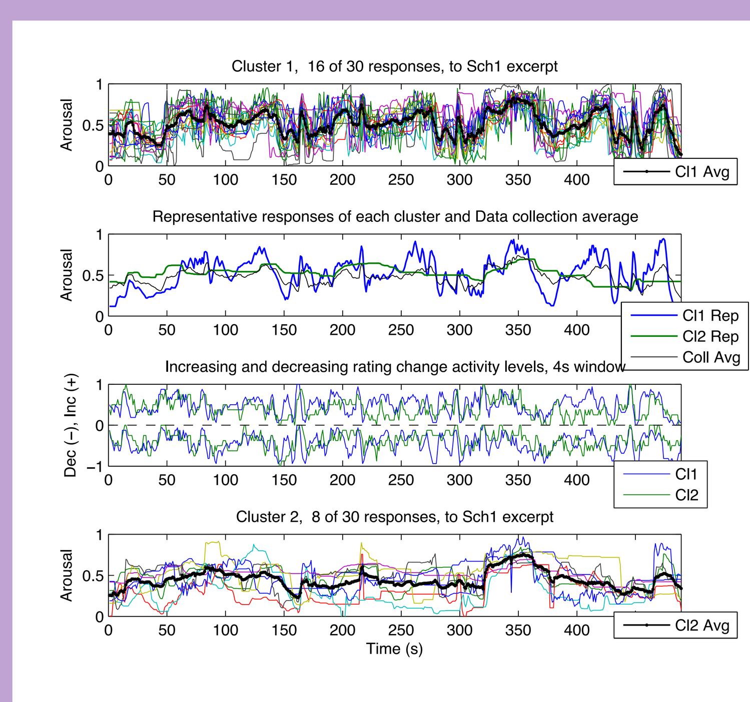

Figure 5: Clusters from correlation between filtered ratings of felt emotional arousal to the first movement of Robert Schumann's string quartet in A major, presented in the same formate as figure 4.

Clustered using low pass filtered data, the most obvious difference between these clusters would have been somewhat dampened, but looking at the activity levels of the two clusters, there are many moments in which the first cluster reports moving up and down and up again in arousal while the second cluster continues a slower trend (ex: 150s, 290s), Over the course of this longer excerpt, these two groups rarely disagree, however the first cluster appears to report more variation more quickly than cluster two, suggesting that the difference here is not a matter of interpretation as much as differences in sensitivity, either of feeling or reporting. The distinction between these clusters is relevant, particularly when model responses from acoustical features which change at different rates.

Active Opposition in Responses

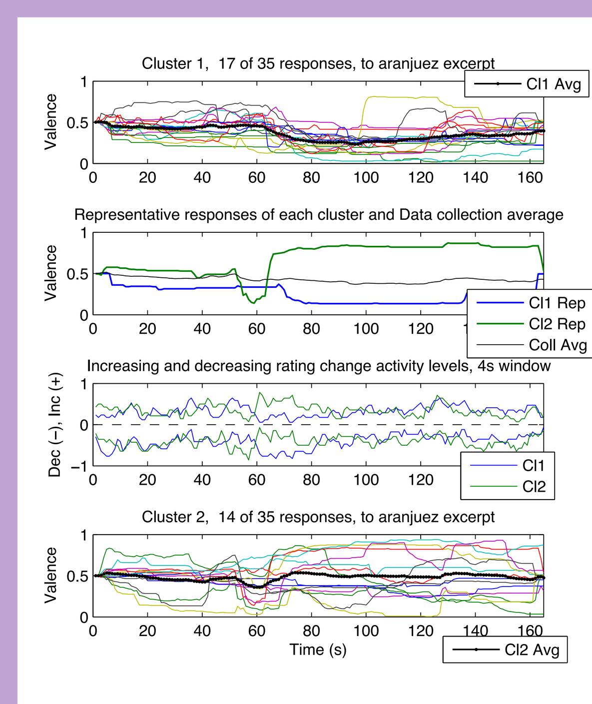

Figure 6: Clusters from the euclidean distance between filtered and differenced ratings of perceived emotional valence to the first few moments of the Adagio of Rodrigo's Concierto de Aranjuez, presented in the same formate as figure 4.

Across the set of 32 collections, few showed disagreement as strongly as the perceived emotional valence ratings to this excerpt of the Concierto de Aranjuez. The cluster activity shows strong oppositional activity from around 50s to 80s, the interval in which, for this recording, the oboe enters. The two clusters are composed of nearly all the responses, but are split so evenly that the average response across the collection barely moves a smudge. In this case, the average is a very poor representation of these activty but contradictory reports of emotion perceived in the music. Previous attempts of using this average for training and testing of models of emotional valence using musical features have been thwarted by this miss-match of high activity and mean stability.

Conclusions

- Clustering revels occasions of dissent in emotional rating responses to music. Though not discussed here, different clustering criteria results in separations which emphasise different time points of contention.
- When clusters show strong dissent this undermines the presumed utility of the average response of a collection.
- Differences in sensitivity can be exposed by clustering, and should be explore more systematically in relation to traits of the individual participant.

This analysis is only a beginning to exploring the differences in response to the same musical stimuli, and there is still much to be done to improve statistics for cohesions, separation, and proper clustering criteria.

Bibliography

Coutinho, E. and Cangelosi, A. (2009). The use of spatio-temporal connectionist models in psychological studies of musical emotions. Music Perception, 27(1):1–15.
Davis, A. (2003). Aesthetic Response to Choral Music: Response Comparisons of Performer- Participants and Non-Performer Respondents. International Journal of Research in Choral Singing, 1 (1):60–14.
Fredrickson, W. (1995). A comparison of perceived musical tension and aesthetic response. Psychology of Music, 23(1):81–87.
Lychner, J. (2008). A comparison of non-musicians' and musicians' aesthetic response to music experienced with and without video. International Journal of Music Education, 26(1):21–32.
Juslin, P. and Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. Behavioral and Brain Sciences, 31(05):559–575.
Korhonen, M., Clausi, D., and Jernigan, M. (2005). Modeling emotional content of music using system identification. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 36(3):588–599.
McAdams, S., Vines, B., Vieillard, S., Smith, B., and Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. Music Perception, 22(2):297–350.

Acknowledgements

My thanks to Stephen McAdams and Mark Korhonen for sharing their continuous rating data; to NSERC and SSHERC, funding bodies which made the collection of these data possible, and the many people involved in the BSO and CARS experiments; to Stephen McAdams again for supervising the beginning of this research as part of my master's thesis, and to my current advisor, Mary Farbood, and the Steinhardt School at NYU for supporting its continuation.