ISyE 6740 – Spring 2021
Final Report

Team Member Names: Samuel Lyon

Project Title: Clusters in Collected Riffs

Problem Statement

For most of my life, music has been a primary passion. Over the years I've written a good deal of material, some of which I've crafted into finished songs; however, as one might imagine, a different subset of what I've written is in fragments. I don't typically write an entire song at once, so I'll sometimes record 10-60 second samples of an idea on my phone to save for future use. These samples can then be developed further or added to other preexisting parts to form a more complete idea.

I have identified 167 samples of various sorts: acoustic guitar, electric guitar, voice, sound effects, etc., from my phone recordings. In this project, I transform the audio files of these samples into manipulatable data; cluster the samples using KMeans, Spectral Clustering, and Gaussian mixture models; pick the best method and cluster size using the Silhouette Score; and then use various methods of analysis to investigate the chosen model. My goals are threefold. First, I'd like to see if the clustering algorithms identify any groupings of samples that I believe can be turned into cohesive ideas; second, I would like to see if the clustering is generally in line with the labels that I assign to each sample; and third, using each cluster in my chosen model as a data point, I would like to see if I can predict certain features of the clusters using the other available features.

Data Source

As mentioned in the problem statement, my data comes from a collection of 167 audio samples that I have accrued over the years. I have assigned each sample a label that classifies it based on the primary instrument/sound used; whether or not a looping pedal (a songwriting and performance tool that allows one to quickly stack multiple "loops" on top of each other to make one overall loop) was used, and, if so, how many loops make up the sample; and which effects, if any, are used. I have also included an additional portion of the label that indicates the general "feel" I get from the sample, extra pieces of information about the sample, and/or whether the sample sounds like it is in a major or minor key.

I will outline this label nomenclature on the next page. First off is the general label structure, followed by some examples of labels with descriptions, and lastly the common label abbreviations used. I have included all of the samples with my project submission in case the reader would like to listen to any of them.

Label Structure: Label#. primary instrument, looping pedal - # of loops - reverse effect, other effects used, additional label portion

Example Label 1: *47. EGA, LP-3-Rev, Dis, major_intro-outro*

This is sample # 47. It uses an amplified electric guitar, a looping pedal with 3 loops and the reverse effect, plus the distortion effect. This sample sounds like it is in a major key, and to me it feels like it could be used for an intro or an outro to some type of media.

Example Label 2: *81. AG, minor_alt-tuning*

This is sample # 81. It uses an acoustic guitar. This sample sounds like it is in a minor key, and I used a non-standard tuning for the guitar.

Example Label 3: *102. AGA, LP-3, on your way*

This is sample # 102. It uses an amplified acoustic guitar and a looping pedal with 3 loops. To me, this sample feels like being on one's way through some type of journey.

Common Abbreviations:

Abbreviation	Definition
AG	Acoustic Guitar
AGA	Acoustic Guitar Amplified
EG	Electric Guitar
EGA	Electric Guitar Amplified
LP	Looping Pedal
Rev	Reverse Effect
Ch	Chorus Effect
Dis	Distortion Effect
Dir	Dirt Effect
Ec	Echo Effect
Re	Reverb Effect
Wa	Wah Effect

Note: Not all labels fit the exact structure laid out above. Some samples include words such as "Beatbox", "Mouth sounds", "Percussion", etc. Hopefully these will be self-explanatory for the reader, or they will be able to be deduced from a quick listen.

Methodology

Having a set of 167 labeled .wav file samples, I proceed to the data conversion. The first website cited at the end of this report {1} contains relevant information on one type of feature that can be extracted from audio data—the chroma vector. The chroma vector contains information about the frequency of the 12 musical notes of the chromatic scale, within each time window of a specified length, in a given audio sample (please see {2} for more information on the chromatic scale). Thus, the chroma vector can be used to classify the most prominent note(s) across a given window of time within a sample, as well as the most prominent note(s) within the entire sample. By manipulating the chroma vector for each sample, I get a count of the number of times that each note was the most prominent note for one of the windows within that sample. I can then normalize this to give a relative frequency for each of the 12 notes across the entire sample, resulting in a 1x12 vector for each sample. I use functions from the pyAudioAnalysis package for data conversion.

Below are the first few rows of my 167x12 data frame.

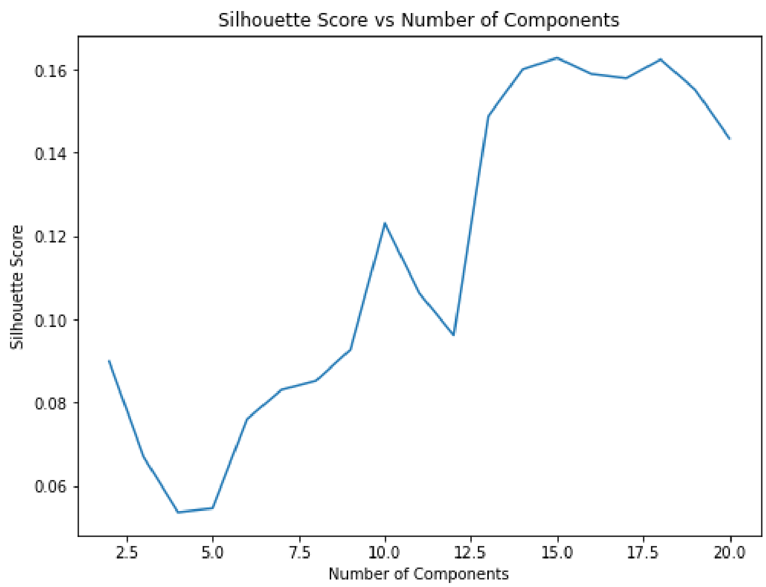
	G#	G	F#	F	E	D#	D	C#	C	B	A#	A
0	0.005525	0.000000	0.000000	0.005525	0.027624	0.000000	0.093923	0.325967	0.000000	0.535912	0.000000	0.005525
1	0.159869	0.063622	0.039152	0.008157	0.001631	0.011419	0.200653	0.249592	0.215334	0.024470	0.024470	0.001631
2	0.006993	0.000000	0.310023	0.002331	0.016317	0.018648	0.202797	0.004662	0.011655	0.006993	0.410256	0.009324
3	0.182222	0.008889	0.393333	0.008889	0.320000	0.000000	0.004444	0.000000	0.000000	0.000000	0.053333	0.028889
4	0.170526	0.004211	0.063158	0.054737	0.054737	0.075789	0.107368	0.006316	0.046316	0.254737	0.151579	0.010526

I now run KMeans, Spectral Clustering, and Gaussian mixture models on the data, using algorithms from the scikit-learn package. Below is the plot of Silhouette Score vs Number of Clusters for each method—I will choose the method/number of clusters that yields the highest Silhouette Score (more information on Silhouette Score at citation {3})

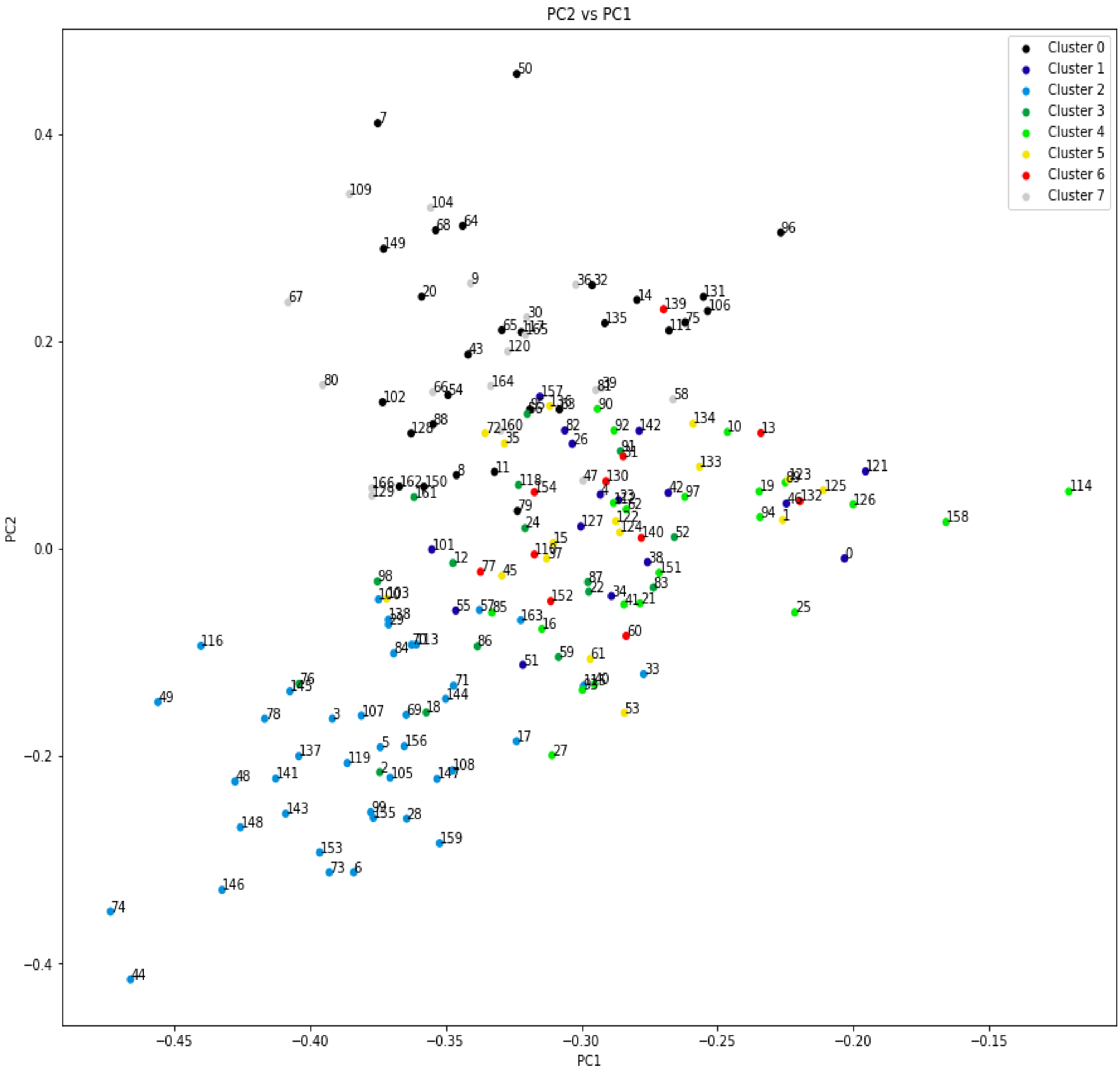
KMeans

Spectral Clustering

Gaussian mixture models



Based on the plots above, the highest Silhouette Score occurs using KMeans with 8 clusters. On the next page, I plot the sample data projected onto its first 2 principal components. Each point (i.e., sample) is numbered with its label number; each of the 8 clusters resulting from KMeans on the full dataset is differently colored.


PC2 vs PC1
Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
0.4
0.2
0.0
−0.2
−0.4
PC2
−0.45
−0.40
−0.35
−0.30
−0.25
−0.20
−0.15
PC1

Evaluation and Final Results

Below, and on the following pages, I provide: a list of the samples in each cluster, a table summarizing relevant categorical statistics for each cluster, and a brief analysis of the preceding information. There is a Total table at the end that describes the categorical statistics across all clusters, which I use for comparisons between clusters. The *% of cluster* value for the Total table can be used as an expected value for the *% of cluster* value for each individual cluster. First, the definitions of each statistic.

Count: The count of samples in the specified cluster in the given category.

% of cluster: Count / total number of samples in the specified cluster

% of total category: Count / total number of samples in the given category

Cluster 0

[7, 8, 11, 14, 20, 32, 43, 50, 54, 63, 64, 65, 68, 75, 79, 88, 95, 96, 102, 106, 111, 117, 128, 131, 135, 149, 150, 162]

	Cluster 0	28								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	11	6	1	6	4	8	5	6	7	16
% of cluster	0.39	0.21	0.04	0.21	0.14	0.29	0.18	0.21	0.25	0.57
% of total category	0.16	0.20	0.50	0.12	0.24	0.14	0.10	0.14	0.25	0.17

Cluster 0 has a higher Other *Count* (all beatbox/mouth sounds) than any other cluster. It has less Effects than expected (based on the Total table), and there are more Minors than expected; however, there are still a good number of Major/happy labeled songs. Samples 65 and 106 are both in this cluster, and they go together well—I actually originally wrote 106 to go with 65.

Cluster 1

[0, 4, 23, 26, 34, 38, 42, 46, 51, 55, 82, 101, 121, 127, 142, 157]

	Cluster 1	16								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	3	3	0	8	2	7	9	8	1	7
% of cluster	0.19	0.19	0.00	0.50	0.13	0.44	0.56	0.50	0.06	0.44
% of total category	0.04	0.10	0.00	0.16	0.12	0.12	0.19	0.18	0.04	0.07

Cluster 1 has more EGA than expected (50% of cluster), which I believe led to more LP and Effects than expected. There are more Major and less Minor than expected, and multiple samples have labels including the word “harmonics”. Samples 46 and 51 are both in this cluster, and they go together well—I actually originally wrote 51 to go with 46.

Cluster 2

[3, 5, 6, 17, 28, 29, 33, 44, 48, 49, 57, 69, 70, 71, 73, 74, 78, 84, 99, 100, 105, 107, 108, 113, 115, 116, 119, 137, 138, 141, 143, 144, 145, 146, 147, 148, 153, 155, 156, 159, 163]

	Cluster 2	41								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	22	10	1	6	2	12	8	10	3	28
% of cluster	0.54	0.24	0.02	0.15	0.05	0.29	0.20	0.24	0.07	0.68
% of total category	0.33	0.33	0.50	0.12	0.12	0.21	0.17	0.23	0.11	0.29

Cluster 2 is mostly acoustic, having more AG and AGA than expected. Many of the samples are not labeled with Major/Minor, leading to more Neither samples. There are 4 samples with “contemplative” in the label, 5 with “hopeful”; also, words like “determined”, “resolve”, “hardheaded”, “stubborn”, “intended outcome”, “gentle resolution”, “driving”, etc. In general, this cluster seems to indicate a feeling of effort and struggle that has paid off or will eventually pay off. Samples 69 and 70 go well together and were originally written to be in the same song. Samples 159 and 153 also go well together and were originally written to be in the same song.

Cluster 3

[2, 12, 18, 22, 24, 52, 56, 59, 76, 83, 86, 87, 91, 98, 118, 161]

	Cluster 3	16								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	7	1	0	6	2	5	5	3	5	8
% of cluster	0.44	0.06	0.00	0.38	0.13	0.31	0.31	0.19	0.31	0.50
% of total category	0.10	0.03	0.00	0.12	0.12	0.09	0.10	0.07	0.18	0.08

Cluster 3 has more Minor than expected, and the labels include words like “fragile”, “loss”, “dissonant_strange”, “wit’s end”. Samples 56 and 76 could be combined to form a more comprehensive idea, although they were not originally written to go together.

Cluster 4

[10, 16, 19, 21, 25, 27, 40, 41, 62, 85, 89, 90, 92, 93, 94, 97, 112, 114, 126, 151, 158]

	Cluster 4	21								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	4	4	0	10	3	8	8	6	3	12
% of cluster	0.19	0.19	0.00	0.48	0.14	0.38	0.38	0.29	0.14	0.57
% of total category	0.06	0.13	0.00	0.20	0.18	0.14	0.17	0.14	0.11	0.13

Cluster 4 has more EGA than expected. Samples 40 and 41 could go together, maybe as back-to-back songs. Samples 62 and 85 are similar, but one is Major and one Minor.

Cluster 5

[1, 15, 35, 37, 45, 53, 61, 72, 103, 122, 123, 124, 125, 133, 134, 136]

	Cluster 5	16								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	4	3	0	6	3	8	4	2	5	9
% of cluster	0.25	0.19	0.00	0.38	0.19	0.50	0.25	0.13	0.31	0.56
% of total category	0.06	0.10	0.00	0.12	0.18	0.14	0.08	0.05	0.18	0.09

In cluster 5, Others make up a bigger portion of the cluster than in other clusters. There are more samples with LP than expected and more Minor than expected. Both rain/windshield wiper samples are in this cluster, and there are 2 guitar samples that reference “water” in their labels.

Cluster 6

[13, 31, 60, 77, 110, 130, 132, 139, 140, 152, 154]

	Cluster 6	11								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	6	0	0	5	0	4	4	5	1	5
% of cluster	0.55	0.00	0.00	0.45	0.00	0.36	0.36	0.45	0.09	0.45
% of total category	0.09	0.00	0.00	0.10	0.00	0.07	0.08	0.11	0.04	0.05

In cluster 6, there are more Major than expected, and all samples are either AG or EGA.

Cluster 7

[9, 30, 36, 39, 47, 58, 66, 67, 80, 81, 104, 109, 120, 129, 160, 164, 165, 166]

	Cluster 7	18								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	10	3	0	4	1	6	5	4	3	11
% of cluster	0.56	0.17	0.00	0.22	0.06	0.33	0.28	0.22	0.17	0.61
% of total category	0.15	0.10	0.00	0.08	0.06	0.10	0.10	0.09	0.11	0.11

Cluster 7 has more AG than expected and less EGA than expected; the cluster is mostly acoustic. Both of the samples whose labels reference "battle" are in this cluster, and there are 2 samples with "alt-tuning" in the label. Samples 164 and 166 go well together and were originally written to be in the same song. This is also the case for samples 66 and 81.

Total

	Total	167								
	Primary Instrument									
	# AG	# AGA	# EG	# EGA	# Other	# with LP	# with >= 1 Effects	# Major	# Minor	# Neither Major/Minor
Count	67	30	2	51	17	58	48	44	28	96
% of cluster	0.40	0.18	0.01	0.31	0.10	0.35	0.29	0.26	0.17	0.57
% of total category	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Prediction

So far, I have explored the first 2 goals from my Problem Statement. I will now explore the final goal, i.e., using each cluster in my chosen model as a data point, I would like to see if I can predict certain features of the clusters using the other available features. First, I turn my cluster Count data into an array, with each cluster as a row. I use the features from AG through Effects, because I wish to predict Major/Minor/Neither.

```
array([[11,  6,  1,  6,  4,  8,  5],
       [ 3,  3,  0,  8,  2,  7,  9],
       [22, 10,  1,  6,  2, 12,  8],
       [ 7,  1,  0,  6,  2,  5,  5],
       [ 4,  4,  0, 10,  3,  8,  8],
       [ 4,  3,  0,  6,  3,  8,  4],
       [ 6,  0,  0,  5,  0,  4,  4],
       [10,  3,  0,  4,  1,  6,  5]])
```

I will try both RidgeCV and LassoCV (from the scikit-learn package). RidgeCV uses Leave-One-Out cross-validation, and LassoCV uses 5-fold cross-validation. In each algorithm, I use the first 6 rows of my array for cross-validation and the last 2 rows for predictions.

RidgeCV

- Predicting the Major Count, I get the following: [2.49, 4.28]. This compares to the actual values of [5, 4].
- Predicting the Minor Count, I get the following: [4.02, 3.94]. This compares to the actual values of [1, 3].

- Predicting the Neither Count, I get the following: [5, 10.1]. This compares to the actual values of [5, 11].

If I only use the features AG through Other (i.e., only the primary instrument features), I get the following predictions.

- Predicting the Major Count, I get the following: [3.37, 4.44]. This compares to the actual values of [5, 4].
- Predicting the Minor Count, I get the following: [.65, 2.61]. This compares to the actual values of [1, 3].
- Predicting the Neither Count, I get the following: [4.96, 10.73]. This compares to the actual values of [5, 11].

Using only the primary instrument features yields better predictions than using all features.

Lasso

- Predicting the Major Count, I get the following: [2.37, 4.39]. This compares to the actual values of [5, 4]. The coefficient values are as follows: [.12, .21, 0, 0, 0, 0, .88].
- Predicting the Minor Count, I get the following: [4, 4]. This compares to the actual values of [1, 3]. The coefficient values are as follows: [0, 0, 0, 0, 0, 0, 0].
- Predicting the Neither Count, I get the following: [5.49, 10.07]. This compares to the actual values of [5, 11]. The coefficient values are as follows: [.81, .62, 0, 1.08, 0, .54, -.52].

In this case, RidgeCV generally performs better than LassoCV for predictions.

Conclusion

Above, I have explored the 3 goals that I set out at the beginning of this project. I intended to: see if the clustering algorithms identify any groupings of samples that I believe can be turned into cohesive ideas, see if the clustering is generally in line with the labels that I assign to each sample, and, using each cluster in my chosen model as a data point, see if I can predict certain features of the clusters using the other available features. In the future, I could add more samples to this data set to see how that changes my clusters and prediction results; I could also try to predict different features than the ones I have already and/or use new features such as the within-cluster-count of each specific effect.

Citations/Information

{1}

https://medium.com/heuristics/audio-signal-feature-extraction-and-clustering-935319d2225

{2}

https://en.wikipedia.org/wiki/Chromatic_scale#:~:text=The%20chromatic%20scale%20or%20twelve,12%20of%20the%20available%20pitches.

{3}

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html#sklearn.metrics.silhouette_score