Genre Visualization using the Free Music Archive

Erik Duus
MATH637 2021

Introduction

Music Information Retrieval (MIR) is a broad, interdisciplinary field concerned with organizing and searching music collections. Relevant disciplines include music theory, audio engineering, digital signal processing, computer science, and information retrieval. Potential areas of interest are varied, including feature extraction, recommender systems, classification systems, automated cataloging, and automated score transcription.

Music Genre Recognition is one such MIR classification problem. Given a set of genres or a genre hierarchy, can a system predict the genre of a particular track? This is not a novel problem, and diverse approaches have been taken to address it. Classic machine learning techniques have been applied, including k-nearest neighbors (KNN) [1], Hidden Markov Models (HMM) [2], and Support Vector Machines (SVM) [3]. Deep learning techniques have also been tried, including Convolutional Neural Networks (CNN) [4] and Long Short-Term Networks (LSTM) [5].

Interestingly, it seems that the question of *which* set of genres or genre hierarchy to use as the ground truth is often not directly addressed. Researchers have a limited set of music datasets available to them and are consequently bound to the associated genre meta-data. Inaccuracies in the genre hierarchy and in genre assignments would obviously impair the performance of any approach to the recognition task.

This project attempts to use some simple dimensionality reduction techniques to visualize the genre space of a popular MIR dataset, gain some insight into the quality of the associated genre hierarchy, and by extension, the relative difficulty of classifying individual genres.

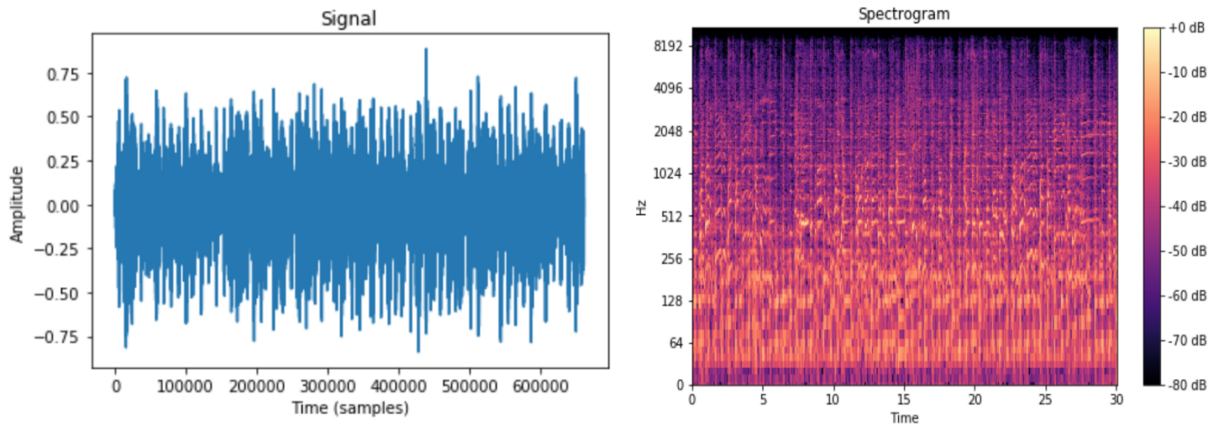
The FMA Dataset

The Free Music Archive (FMA) [6], founded in 2009 by radio station WFMU, provides over 100,000 audio tracks of Creative Commons-licensed original music, from over 16,000 articles arranged in a genre hierarchy of more than 160 genres. It also includes extensive track metadata, tags (comments, listens, favorites, etc.), and text such as artist biographies.

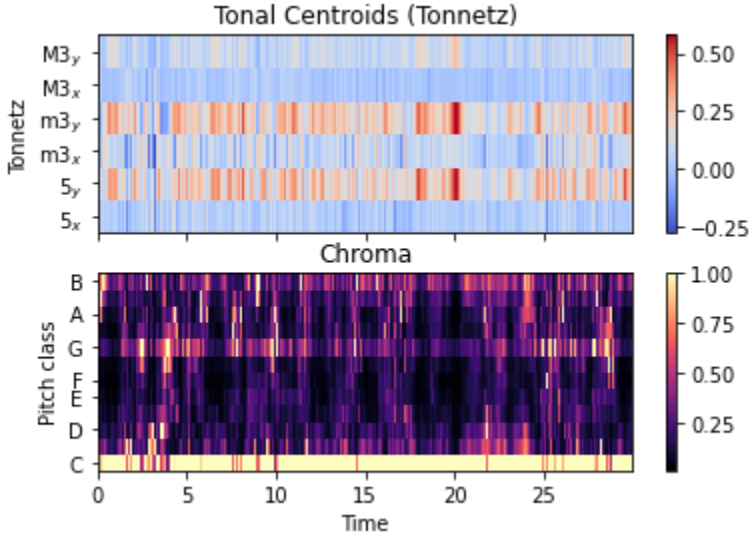
The FMA Dataset [7] is the result of an initiative to create a benchmark MIR dataset using the FMA as a basis. Existing MIR datasets suffer from a number of drawbacks, including small numbers of tracks, poor quality audio, lack of metadata, licensing, and lack of accessibility. The FMA Dataset is large, has extensive metadata (including genres), and has high-quality, freely licensed mp3s that are available for download as part of the dataset.

Audio Features

The dataset is enriched with a set of audio features extracted using Librosa[]. While the details of audio processing algorithms are beyond the scope of this paper, they generally attempt to extract information from a raw audio signal. For example, a spectrogram converts an audio signal into a map of frequency intensities over time:



Other features refine the basic spectogram into components related to the human perception of music, such as tonal centroids and chroma:



The FMA dataset, however, does not contain the extracted audio features. Instead, it includes a set of summary statistics computed over the binned frequencies. It is important to note that this eliminates the time component of these features. In total, there are 518 audio features consisting of 7 statistics (min, max, mean, median, standard deviation, skew, and kurtosis) across the following Librosa-extracted features:

- Chroma (STFT, CQT, CENS variants): 12 frequency bins
- Tonnetz: 6 frequency bins
- Mel Frequency Cepstral Coefficients (MFCC): 20 frequency bins
- Spectral centroid

- Spectral bandwidth
- Spectral contrast: 7 frequency bins
- Spectral roll-off
- RMS energy
- Zero-crossing rate

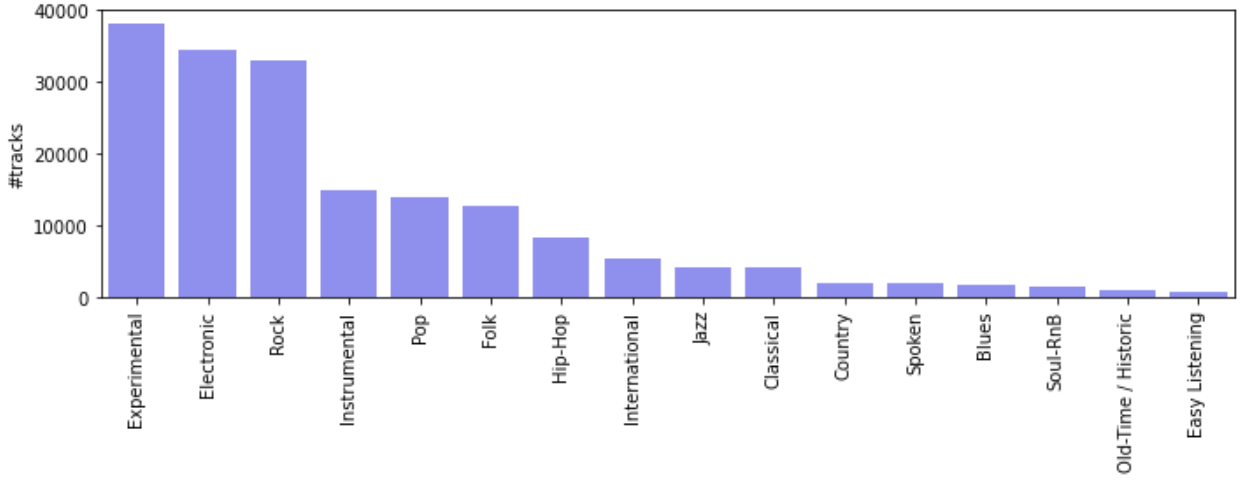
Subsets

The FMA dataset is also enhanced with several pre-computed subsets to simplify analysis and model creation:

- Full: the complete dataset, with all tracks, full audio clips, metadata, and features.
- Large: the full dataset, but with 30-second audio clips
- Medium: a subset of the large subset, selecting only those tracks where the assigned genres share the same root genre. The subset includes approximately 25,000 tracks with unbalanced top-level genres.
- Small: a balanced subset of the medium subset, consisting of 8,000 tracks from 8 genres.

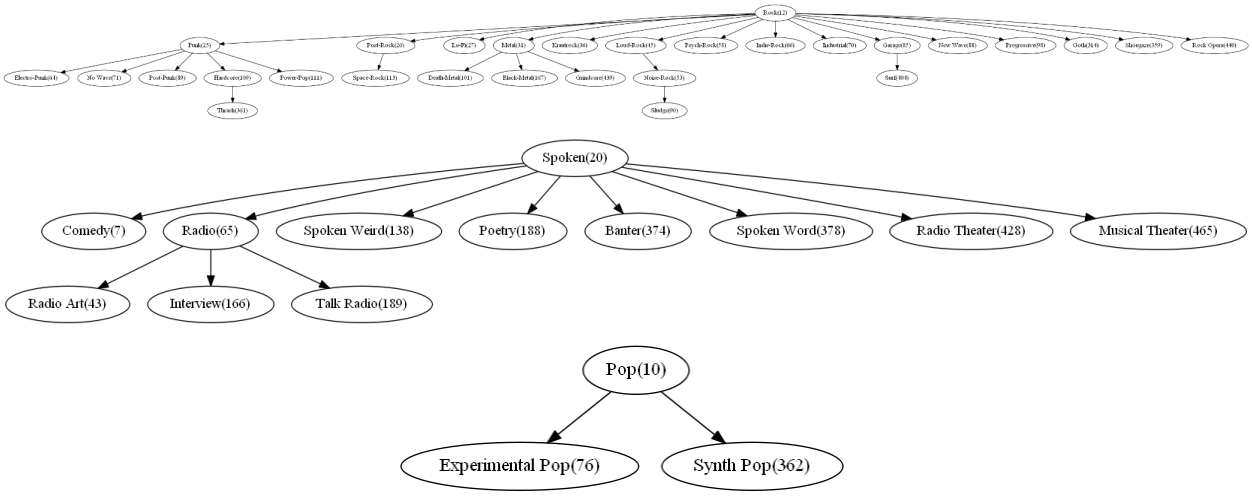
Genre Hierarchy

FMA tracks are tagged with genres by the artists themselves from a predefined list of 161 genres in a hierarchy. There are 16 root genres:



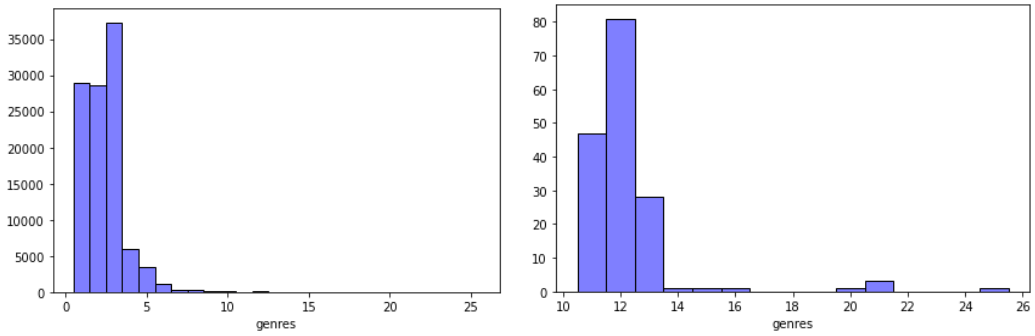
Note that the counts overlap since some tracks are tagged with multiple root genres. The FMA dataset appears to be substantially overweight in Experimental, Electronic, and Rock genres. The imbalance could be related to the nature of the archive itself (free music from lesser-known artists. It could also be because genres are often ambiguous (for example, it would be easy to tag a Pop song as Electronic).

The genre hierarchy is large, so only some illustrative snippets (Rock, Folk, and Pop) are highlighted here.



These examples illustrate that the hierarchy seems inconsistent. The Rock hierarchy is extensive, while the Pop hierarchy is minimal. There are substantially more Pop tracks than Spoken, yet the Spoken hierarchy is far more comprehensive. Additionally, the root genres don't reflect relationships with each other (for example, the relationship between Blues, R&B, and Rock).

Examining the number of genres per track, the majority of the tracks have 5 or less, yet some have up to 25 assigned genres:



This again speaks to inconsistency in the assignment of genres to tracks or the genre hierarchy's shortcomings. The small subset is used to simplify the subsequent analysis:

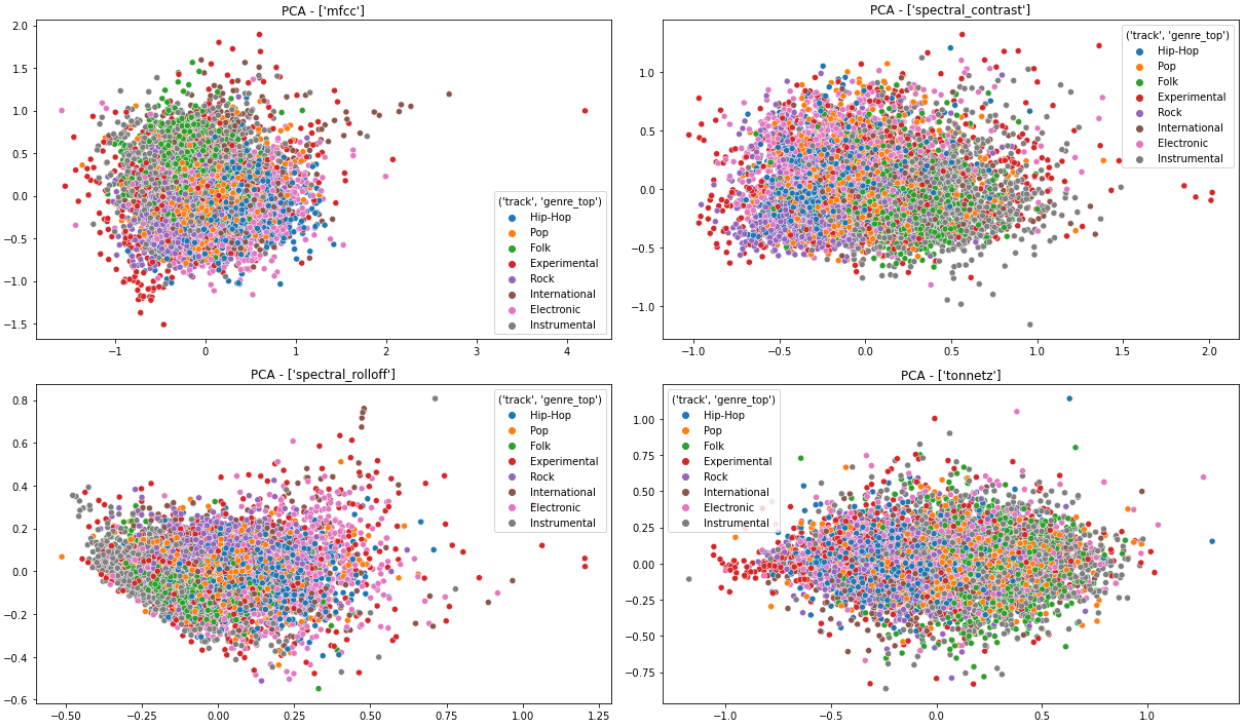
- The subset only includes tracks where the assigned genres share a single root (genre_top). A track encoded as both Rock and Pop is not included.
- The dataset is balanced with 8 genres and 1,000 tracks per genre.

Exploring genre structure through dimensionality reduction

PCA

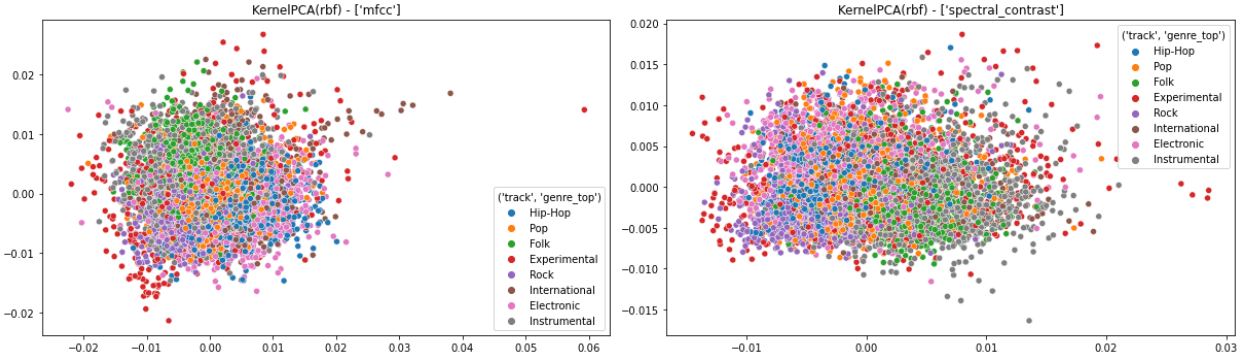
The feature space is high-dimensional, with 518 total features in all. MFCC alone has 140 features, while spectral rolloff, ZCR, and RMS each contribute only 7.

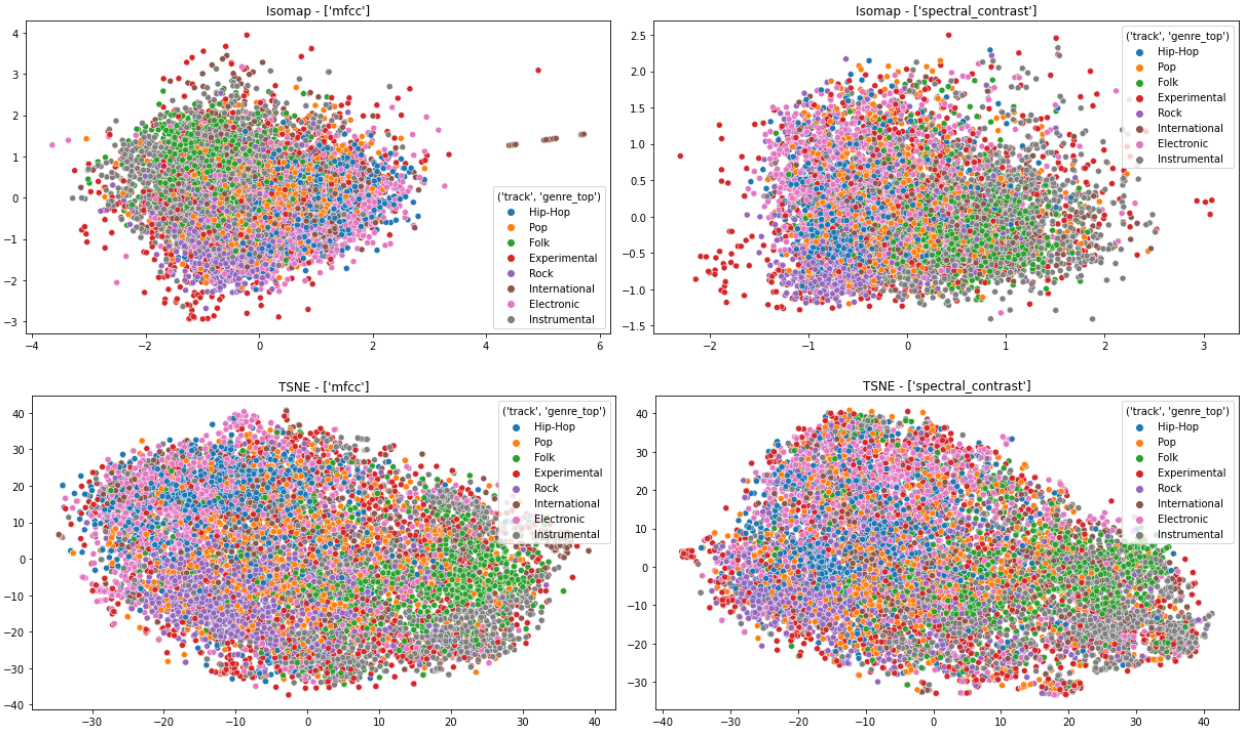
Applying a 2-component PCA to each feature shows some clustering of genres but does not definitively separate them. MFCC, spectral_contast, spectral rolloff, and tonnetz showed slightly better results:



Non-linear Techniques - Kernel PCA, Isomap, TSNE

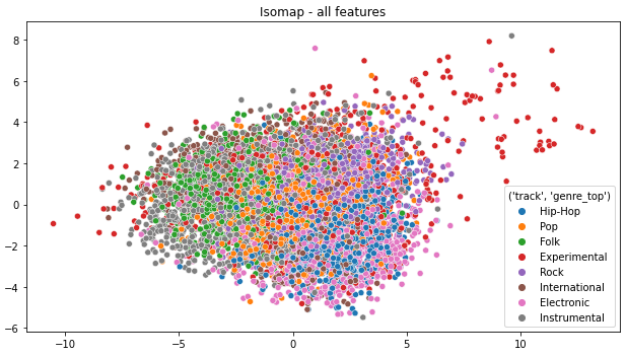
Non-linear approaches produced little benefit. Each approach was tried with several iterations of tuning parameters without marked improvement:





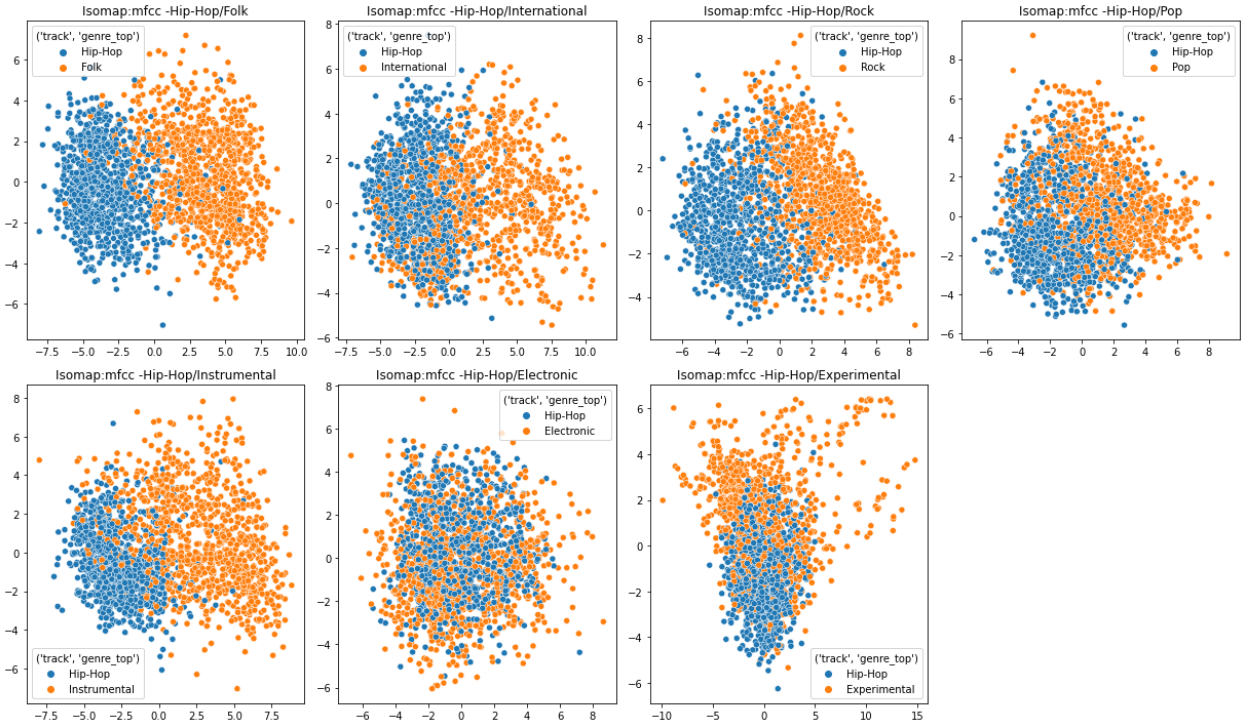
TSNE and Isomap perhaps show slightly better clustering than PCA, but still don't discern more structure.

Isomap across all features also appears marginally more tightly clustered:

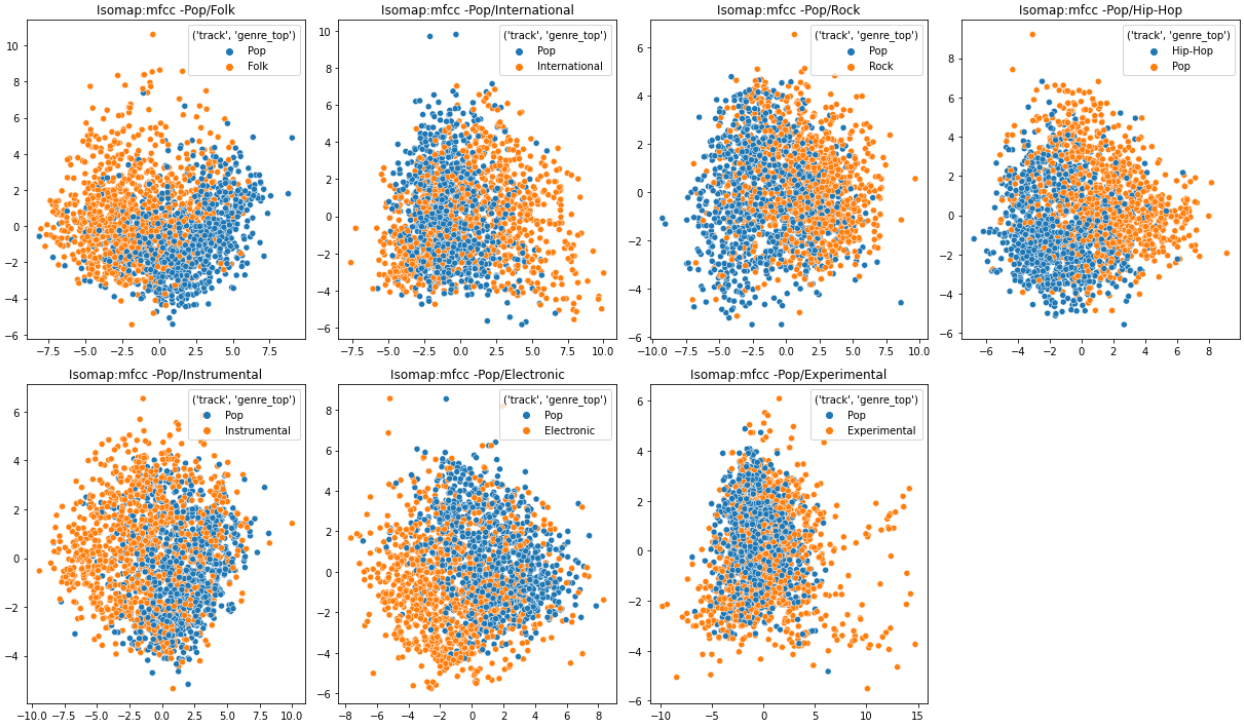


Isomap on Genre Pairs

Although dimension reduction of the audio feature space is not able to resolve genres distinctly at low dimension, it does appear that some genres discriminate better than others. Isomap on each genre pair shows good discrimination between some pairs and poor or no discrimination between others. For example, Hip-Hop shows good separation against most genres except Electronic:



Pop, on the other hand, only discriminates well vs. Folk and Hip-Hop:



Results

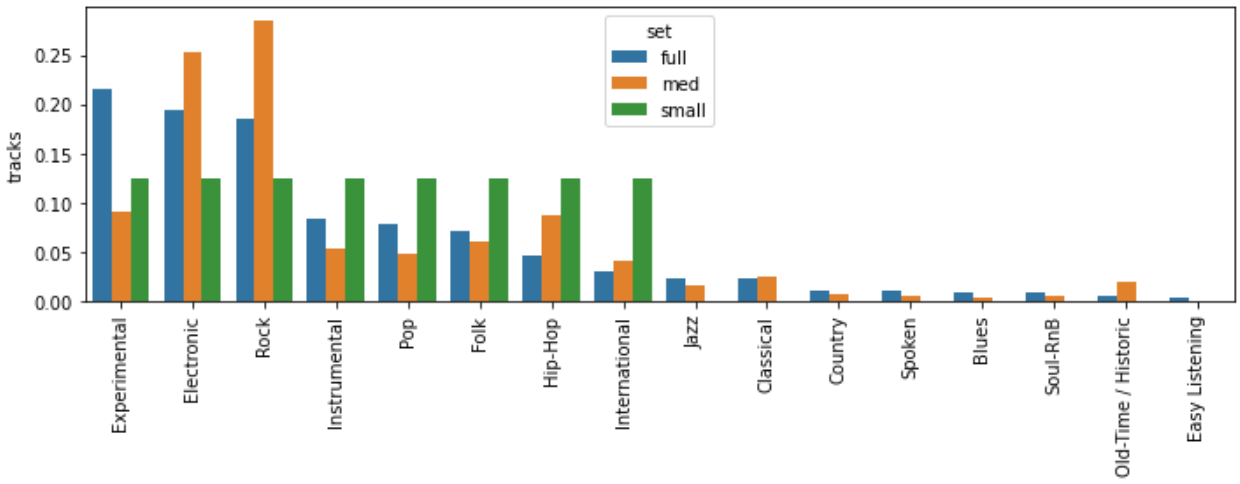
Dimension reduction was unable to extract clear structure from the feature space. Several possibilities come to mind as to why this is the case. Perhaps the selected features do not discriminate well enough. They are aggregate statistics of extracted audio features, not the features themselves. As noted before, the aggregation discards the time domain and also the temporal structure of the music. Alternatively, perhaps there is some structure in the high-dimension feature space, and these dimension reduction techniques cannot adequately project it to lower dimensions. Finally, maybe the selected audio features are simply unable to discriminate between genres, and additional features are required.

In addition, an examination of the last plot reveals certain genres cluster better than others. Hip-Hop and Folk seem to be tightly clustered, while Experimental seems to be far more dispersed. This could be because some genres are more closely related than others, and therefore harder to discriminate. It could also indicate a lack of precision in the FMA genre hierarchy and in the assignment of genres to tracks.

Genre Balance and Relationships

The preceding analysis was restricted to the small subset of the FMA dataset, a set of tracks with a single root genre, and balanced across 8 genres. Given the difficulty in discriminating between certain genre pairs, a further examination of the genre balance and relationships of the full dataset is instructive.

Examining the genre counts of the full, medium, and small subsets, it is clear that the subsets each have very different genre balances. The full dataset is highly biased towards Experimental, Electronic, and Rock, with very poor support for half the root genres. The medium dataset is even more strongly biased towards Rock, with similarly poor support for minor genres.

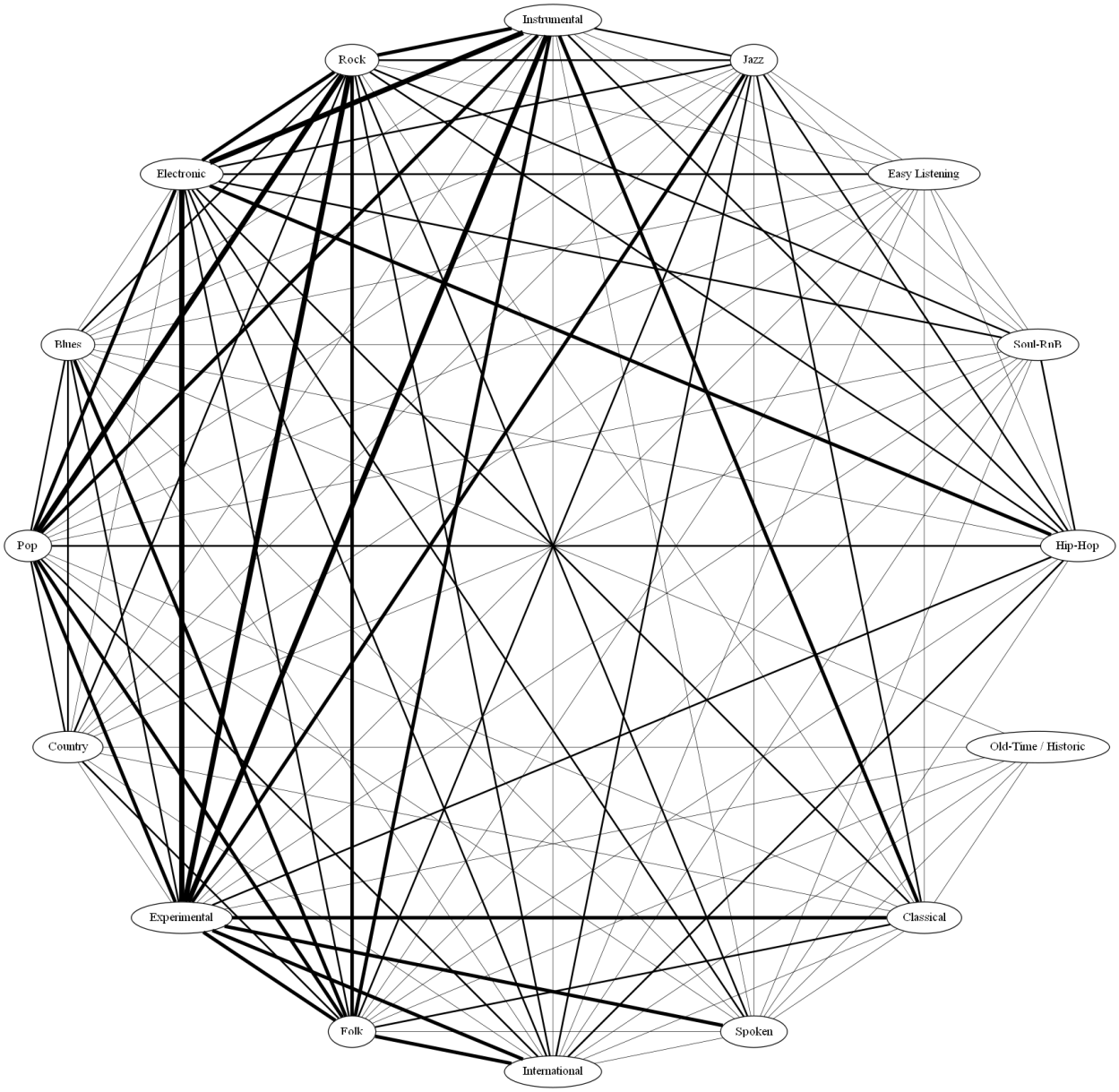


This leads to several observations:

- Despite its size, the full FMA dataset is probably an unrepresentative corpus.
- The subsets are not representative of the full dataset.

- Tracks with a single root genre may be unrepresentative of the corpus; given that the full dataset is majority multi-genre, perhaps the single root genre tracks are incorrectly labeled.

Examining the multi-genre set of audio-tracks, the frequency of genre combinations can be embedded in a graph, and the relative frequency of combinations captured as edge weights:



The full dataset clearly has a lot of multi-genre structure, and therefore it is likely that the single root genre subsets are not ‘pure’ examples of a single genre.

Conclusion

The FMA dataset is a useful contribution to MIR research due to its size, freely available audio tracks, extensive metadata, and additional enrichment with audio features. The selected audio features show some ability to discriminate genre, but both linear and non-linear dimension reduction techniques were not completely able to resolve all genres in low dimension:

- The techniques applied may not be sufficiently powerful.
- The audio features may be insufficient; in particular, the lack of temporal structure seems important.

Further exploration of the genres highlights that the full dataset is biased towards a handful of genres and is strongly multi-genre. As a result, single genre analysis with the dataset might be difficult:

- Subsets may be unrepresentative given the different genre balances.
- Tracks with a single root genre may not be ‘pure’ representations of the genre.

References

[1] Rosner, Aldona, Marcin Michalak, and Bożena Kostek. "A study on influence of normalization methods on music genre classification results employing kNN algorithm." *Proceedings 9th National Conference on Databases: Applications and Systems*. 2013, [PDF] pg.gda.pl.

[2] Shao, Xi, Changsheng Xu, and Mohan S. Kankanhalli. "Unsupervised classification of music genre using hidden markov model." *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*. Vol. 3. IEEE, 2004, [PDF] psu.edu.

[3] Xu, Changsheng, et al. "Musical genre classification using support vector machines." *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)*.. Vol. 5. IEEE, 2003, [PDF] researchgate.net.

[4] Dong, Mingwen. "Convolutional neural network achieves human-level accuracy in music genre classification." *arXiv preprint arXiv:1802.09697* (2018). [PDF] arxiv.org.

[5] Tang, Chun Pui, et al. "Music genre classification using a hierarchical long short term memory (LSTM) model." *Third International Workshop on Pattern Recognition*. Vol. 10828. International Society for Optics and Photonics, 2018, [PDF] cuhk.hk.

[6] "About." *Free Music Archive: About*, freemusicarchive.org.

[7] Defferrard, Michaël, et al. "Fma: A dataset for music analysis." *arXiv preprint arXiv:1612.01840* (2016), [PDF] arxiv.org, [GitHub].