Music Beyond Major and Minor

Aleksandr Timofeev (aleksandr.timofeev@epfl.ch), Andrei Afonin (andrei.afonin@epfl.ch), Dubravka Kutlesic (dubravka.kutlesic@epfl.ch), supervisor: Robert Lieck (robert.lieck@epfl.ch), Digital and Cognitive Musicology Lab, *EPFL, Switzerland*

***Abstract*—Interpreting structures and discoveries beyond major and minor keys in a piece can help understand music space, confirm existent theories and derive new findings. We employ the Dirichlet Mixture Model, the Isomap method, and visualization techniques to find answers to the aforementioned questions. Our analysis found that the space of pitch scapes possesses a torus-wave structure with many attractive properties cohered to musical theory.**

I. Introduction

In general, Western music is usually thought of as being either in a major or minor key. In reality, there are more sophisticated characteristics in the tonality of the piece. The aim of the project is to discover novel relevant keys, the ones that are beyond major and minor, which can provide fascinating insights.

Different statistical approaches represent powerful tools to explain and confirm the findings from the music theory. However, the numbers or labels that represent music pieces are not easy to retrieve. Here, we use a model of novel pitch scape representation [1], which is applicable to different music eras.

In our work, we focus on visualization techniques that can help in explaining the structure of the space of pitch-class distributions. Auxiliary tools for this are Dirichlet mixture model (DMM) and Isomap method [2]. The former is designed in similar to Gaussian mixture model manner [3] and inspired by [1]. The latter is the dimensional reduction method that allows us to interpret results obtained with 12-dimensional data. In addition, we inject in our model transposition invariance, the underlying property of pitch classes. We show the following space features. First, pitch scapes are distributed over a torus forming two intertwined wave-like structures that correspond to major and minor keys. Second, points in both keys do not lie uniformly. Some regions are denser than others. Third, outliers that do not follow the torus structure correspond to pitch scapes taken from a small-time interval. Finally, we conclude that Bach followed patterns that other composers did not. All of these findings are confirmed by musical theory and described in our report.

In section II, we introduce musical concepts necessary to comprehend obtained results. Next, in section III, we describe models and methods that we employ. Training details and the dataset description are given in section IV. Our main results and its discussion are presented in section V.

II. Musical background

A. Pitch Scapes

Despite the natural origin of music, it has a rigorous structure. Thus, musical theory operates with many terms that can be represented from statisticians' points of view.

We begin with the concept of a pitch. Pitches reflect discrete tones with individual frequencies or simply notes. However, it is inconvenient to use it for building mathematical theories. The pitch C of the third octave is not the same as one of the fifth octave. That is, they do not possess octave equivalence. That is why musical theory introduces a new concept - neural pitch classes. They have two pleasant properties. The first is the octave equivalence; pitches from different octaves are considered as one. The second is related to enharmonic equivalence. This implies that notes, intervals, or chords that sound the same but are spelled differently belong to one pitch class [4].

Next, we move to the most crucial concept that we employ throughout our work - pitch scapes. Distribution of pitch classes changes varying observed time intervals. Thus, it is convenient to have its probabilistic representation. For this purpose, pitch scapes are originally defined in [1]. They allow taking into account the hierarchical structure of Western classical music obtained in both musical form and harmonic syntax. We define pitch scapes following [1] :

Definition 1 (Pitch Scape): A pitch scape $\mathfrak{S}$ is a function that maps each proper time interval $[t_s, t_e]$ $(t_s < t_e)$ to pitch class distribution:

$$\mathfrak{S} : \mathbb{R} \times \mathbb{R} \to [0,1]^{12} \ s.t. \sum_{\pi=0}^{12} \mathfrak{S}(\pi|t_s, t_e) = 1,$$

where π denotes pitch classes.

However, it is not clear still how this distribution can be estimated. Before, we need to formalize mentioned above distribution of pitch classes:

Definition 2 (Pitch Class Density): The pitch class density $\delta(\pi|t)$ for pitch class π at time t corresponds to the normalised pitch class counts over all tones that sound at time t:

$$\delta(\pi|t) := \frac{1}{\max\{1, |T_t|\}} \sum_{\tau \in T_t} \mathbb{1}[\tau \bmod 12 = \pi],$$

where T_t is the multi set of all tones represented as integers sounding at time t; the max operation avoids division by zero for silent parts where $T_t = \varnothing$.

Leveraging the definition of pitch class density we can describe estimation of pitch scapes:

Definition 3 (Pitch Scape Estimate): The posterior estimate of the pitch scape $\mathfrak{S}(\pi|t_s, t_e)$ for pitch class π and time interval $[t_s, t_e]$ is:

$$\mathfrak{S}(\pi|t_s, t_e) := \frac{1}{t_e - t_s + 12c}\left[c + \int_{t_s}^{t_e} \delta(\pi|t)dt\right],$$

where the integral over the pitch class density computes the overall pitch class counts, $c \geq 0$ specifies the prior counts, and the leading term ensures proper normalization.

B. Transposition invariance (musical view)

In music, the term transposition refers to shifting all pitches by a constant interval. However, as it is already mentioned, pitch classes have octave equivalence. It results in one important property that we must take into consideration for building models - transposition invariance. It raises naturally from the human perception that perceives and memorizes musical compositions in terms of pitch intervals. Therefore, transposition along pitch dimension does not change perception.

C. Circle of fifths

The next important concept, a very powerful tool for a musician, is the circle of fifths. It is applied to find key signatures, build scale, build chords, and many others.

The circle of fifths is built as a sequence of pitches located in a circle such that every next pitch (clock-wise) is found seven semitones (a fifth) higher (Fig. 10 in Appendix). Furthermore, it consists of two concentric circles representing major and minor keys outside and inside, respectively. It is designed such that pitches on the top (C or Am) do not have any accidentals, i.e., sharps or flats. Then, moving clock-wise, every new key adds one sharp while moving counter-wise it increases the number of flats. Notably, the circle, due to its structure, accounts for the enharmonic equivalence of pitch classes.

D. Tonnetz

The Tonnetz is a generalization of the circle of fifths. It can describe more harmonic relations between chords. The Tonnetz is often used to analyze chord progressions, find relative major and minor keys, and transpose chord sequences.

The first line of the Tonnetz is obtained by unrolling the circle of fifths horizontally and giving it a periodic structure (Fig. 12 in Appendix). The same line is repeated limitless times vertically and staggered to place the major third and sixth above and minor thirds and sixths below each tone, which corresponds to three and four steps on the circle of fifths. Also, Tonnetz diagram is known to be topologically equivalent to a torus. To verify this, one needs to fulfill the following transformations. First, the Tonnetz is folded such that enharmonic equivalent pitches place on top of each other, which gives a tube. Second, this tube is folded along the circle of fifths, such that multiple occurrences of the same tone place on top of each other. Eventually, we observe a torus as in Fig.11 (Appendix).

III. Models and methods

A. Dirichlet Mixture Model

We use Dirichlet Mixture Model (DMM) [5] for clustering given points, which is similar to the Gaussian Mixture Model [3], where instead of the Gaussian distribution Dirichlet distribution [5] is used. The motivation of using Dirichlet distribution is that data points lie on a simplex, a support set of this distribution. We maximize the log-likelihood function [5] by adapting the parameters of our model. Assume that we have M classes, then classes indices vary in $\{1, .., M\}$. Then we set mixture weights as:

$$W_j = \frac{\exp(w_j)}{\sum_k \exp(w_k)},\ \forall j \in \{1, .., M\}$$

And for a given point $x \in \mathbb{R}^D$, we define likelihood as:

$$\text{likelihood}(x) = \sum_{j=1}^{M} W_j \cdot \frac{1}{B(\alpha)} \prod_{i=1}^{D} x_i^{\alpha_i - 1}$$

Hence, this model has a set of parameters $w \in \mathbb{R}^M$, which are mixture weights and a set of parameters $\alpha \in \mathbb{R}^{M \times D}$ which are parameters of all Dirichlet distributions included in the mixture.

B. Transposition invariance (algorithmic view)

It is already mentioned that pitch classes have octave equivalence and possess transposition invariance (TI) property along the pitch dimension. Since this needs to be taken into account, we inject transposition invariance property into our model. As a result, for a given point $x \in \mathbb{R}^D$ we define transposition invariant DMM with the likelihood function:

$$\text{likelihood}(x) = \sum_{j=1}^{M} W_j \cdot \frac{1}{D} \sum_{k=1}^{D} \frac{1}{B(\alpha^k)} \prod_{i=1}^{D} x_i^{\alpha_i^k - 1},$$

where α^k means the k-th transposition of α. Here, we have the same number of model parameters as in the case of regular DMM.

C. Isomap

We apply the non-linear dimension reduction method Isomap [2] to visualize the results of clustering. Color of the point $x \in \mathbb{R}^D$ corresponds to the given cluster label. By applying this method, we map points from high-dimensional space to two dimensional or three dimensional one to find patterns and better understand the given data structure. In our case, the Isomap algorithm leverages five neighbors for each point and Euclidean distance to measure points closeness.

IV. Training

A. Dataset

The results are obtained on 1910 pieces from 77 composers who lived in various epochs and countries. Therefore, results do not rely on one particular style or time. Musical pieces are extracted from XML files using library [1]. First, a piece is divided into 10 equally sized sub-sections with no overlapping

Type	Cluster 1	Cluster 2
Major	4933	35955
Minor	36286	8776

TABLE I: The distribution of major/minor pieces over the found clusters

time intervals. A pitch-class distribution is estimated from each sub piece. Next, partially-overlapping sub pieces are combined to longer sub pieces giving $9, 8, 7..2$ pitch-class distributions and finally, the distribution representing the complete piece. Therefore, for each piece, we have 45 data points, giving in total 85950 pitch-class distributions. The distribution represents the input for the Dirichlet mixture model. Furthermore, additional metadata about pieces is provided, such as composer names, opuses, years etc.

B. Models and Training details

We train DMM (w/o TI) and DMM (with TI) with a different number of clusters. The models are trained by maximizing corresponding log-likelihood function and adapting their weights using Adam optimizer [6] with learning rate $= 0.05$ (other parameters are default to PyTorch [7]) for 1000 epochs. Finally, we choose a cluster label for each point according to the maximum over likelihoods in the mixture of the trained model.

V. Experiments

A. DMM experiments

1) Major and Minor clustering: For each data point, we have an estimate of whether the given data point corresponds to a major or minor key, based on a template-based key-finding algorithm. Hence we can test an assumption that our clustering procedure can divide a given dataset into major and minor pieces. To test this assumption, we run DMM with transposition invariance with 2 output clusters, that is, $M = 2$. The experiment results are given in Table I. Each value is the number of data points in the dataset corresponding to the given cluster and given key. This result shows a strong correlation between output clusters of the model and real major/minor attachment. Unfortunately, it is hard to make some deeper conclusion from this result.

2) Clustering results: This section shows the result of clustering by regular DMM (w/o TI) with 3 clusters and clustering by DMM with transposition invariance property (with TI) with four clusters. For each point, we find 2 dominant tonics, and Fig. 1,2 show corresponding distributions of points over these dominant pairs on the circle of fifths. The sectors in these pie charts are colored according to cluster prediction and the size of the piece corresponds to a number of points in this sector. Also, Fig. 3 and Fig. 4 show corresponding Isomap mappings into two-dimensional space, which slightly remind us of a torus structure. Mappings are colored according to cluster prediction for both models. One can notice similarities between the Isomap plot and the circle of fifths plot for both models. Indeed, Fig. 1 and Fig. 3 show us that clusters

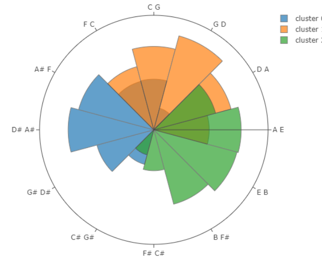

Fig. 1: Distribution over notes (DMM w/o TI, 3 clusters)

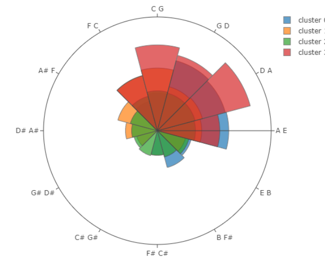

Fig. 2: Distribution over notes (DMM w TI, 4 clusters)

Fig. 3: Isomap mapping(DMM w/o TI, 3 clusters)

Fig. 4: Isomap mapping(DMM w TI, 4 clusters)

predicted by the DMM (w/o TI) try to occupy their sector of a circle of fifths with small overlap on the boundaries. One can notice similarity comparing Fig. 2 and Fig. 4 as well. In Fig. 2 one can notice bias in points distribution towards the top right part of a circle. This result tells us that composers used to write music using C, G, D, A tonics. It is quite a common phenomenon because there are no flats and sharps, which makes the musical composition more complex.

B. Exploratory analysis

1) Torus-wave structure: This section is devoted to structure analysis of a given data with the help of the circle of fifths and Tonnetz. We show plots of mapped points in three-dimensional space using the Isomap method [2] and color them according to their key or transposition. We expect that keys of pitches that are close on the circle of fifths (Fig.10) lie close in the embedding space. Moreover, they should repeat a circle structure. It can be seen, for example, in Fig.5, where points belong to the major key and colored corresponding to its key estimate. The minor key also possesses this structure. However, some points tend to lie close to the center (Fig.6). This happens because the minor key is more diverse. Unlike the major key, it has two possibilities to resolve a harmonic. For

Fig. 5: Top view major

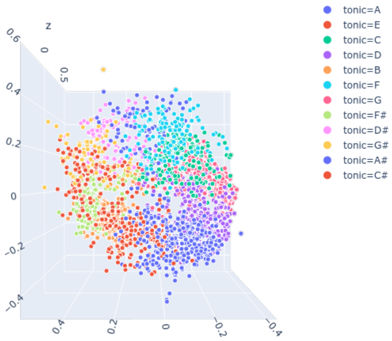

Fig. 6: Top view minor

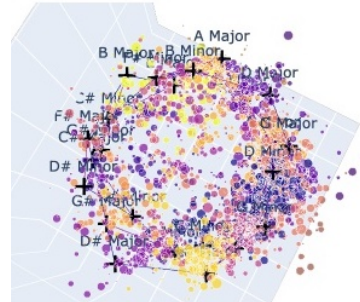

Fig. 7: The whole corpus

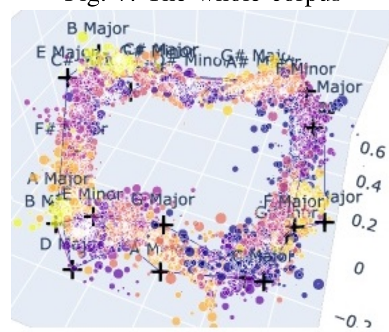

Fig. 8: Bach

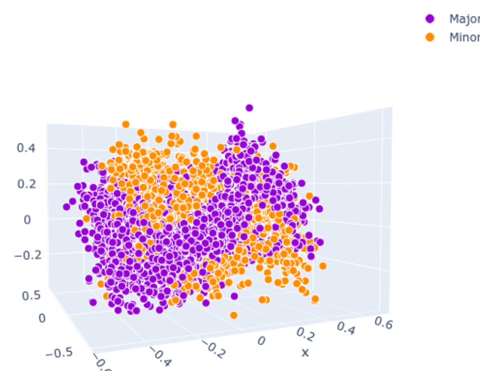

Fig. 9: Side view for both keys

example, composers may want to convey various perceptions or tension by changing the standard minor scale, i.e., playing $G\sharp$ instead of G. This moves the minor compositions to the uniform distribution.

The same argument can be applied to the Tonnetz diagram, which is known to be topologically equivalent to a torus. That explains a torus in Fig.5 and 6. One can notice that points of both keys form wave-like structures, which can be seen from a side point of view. The latter observation is motivated by the fact that the Tonnetz adds major and minor third relations, which correspond to four steps for major and three steps for the minor. Hence, placing such related points closer to each other in embedding space can result in the wave-like structure. This argument relates to C major and E major, and C major and $E\flat$ major. Our plot of embedding (Fig.5) with coloring according to transpositions clarifies the above explanations.

2) Why was Bach special?: In this section, we compare the whole corpus and pieces composed by Bach. Figures 7 and 8 depict Isomap embedding for the entire dataset (2963 points are sampled) and on pitch-class distributions from Bach. Twelve major and twelve minor profiles are added to the plots as crosses. They are connected according to the circle of fifths. Colors are transpositions. We can see that points that belong to Bach almost entirely follow the torus structure in the space. However, on the whole corpus, we can not conclude that this structure exists. Therefore, after the Bach period, music became a lot more diverse. Additionally, the size of a point corresponds to the relative time duration of a scape. It is calculated as $\frac{t_e - t_s}{T}$, where T represents the duration of the piece. The prior counts for both figures is 10^{-5}. It can be seen that smaller points scatter more broadly and tend to be further away from the torus center.

VI. Conclusion

In our work, we presented various approaches that can tackle the problem of finding beneficial structures beyond major and minor from different angles. The first crucial finding is that the structure of musical space can be explained with known musical concepts. We showed that major and minor keys form a torus-wave structure in a three dimensional space. Moreover, eliminating points with small window size result in a denoised torus-wave structure. Next, pitch scapes composed by Bach almost perfectly follows the torus structure. Finally, Dirichlet Mixture Model with transposition invariance allowed us to reveal that composers tend to create "simple" pieces close to C-major and A-minor, that is, the pieces do not have flats and sharps.

References

[1] R. Lieck and M. Rohrmeier, "Modelling hierarchical key structure with pitch scapes," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 2020.

[2] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 01 2000.

[3] D. A. Reynolds, "Gaussian mixture models.," *Encyclopedia of biometrics*, vol. 741, 2009.

[4] J. Straus, *Introduction to Post-tonal Theory*. Prentice Hall, 1990.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing ed., 2007.

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[7] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), pp. 8024–8035, Curran Associates, Inc., 2019.

Appendix

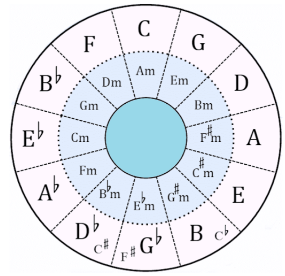

Fig. 10: The circle of fifths. The image is downloaded from http://fretjam.com/3-practical-uses-for-the-circle-of-fifths.html

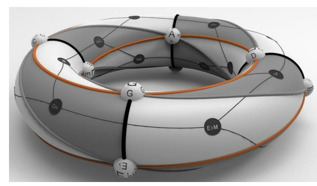

Fig. 11: Tonnetz mapped in torus. The image is downloaded from https://www.epfl.ch/labs/dcml/the-dcml-blog/

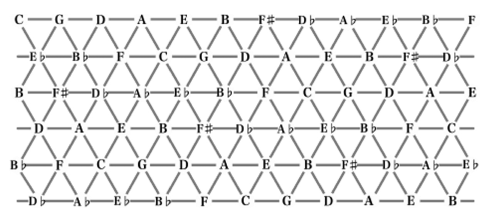

Fig. 12: The Tonnetz. The image is downloaded from https://link.springer.com/chapter/10.1007/978-3-030-50417-5_31