# Meeting Notes

THEORY407, Fall 2022

# Contents



## 2022-09-23

**Humdrum Ch.1** It seems like a tool used for analyzing music similar to music21. With multiple funtionalities, Humdrum can be used for counting specific notes, showing certain voice in a piece, and sometime even some specfic analysis on harmonic/tonal funtions. It is more like a library of tools on my idea of searching how similar two pieces are. We can conduct research with the same approch, but with different methods. Starting with these simple metrics:

- **Note Sequence**: How many notes are the same, how many are different?
  - Might be useful for comparing two pieces in the same work, such as *Gymnopedie*. But will yield a very low score if the key changed.
- **Note Sequence with Key**: How many notes are the same, given pieces in different key but represented in a $\mathbb{Z}_7$ format.
  - Can address some very basic isssues such as exact transposition, but not effective if the phrase changes.
  - Can use the technique mentioned in the proposal to solve the above problem. But what if the phrase was changed a lot?

We then shall do the analysis with more and more complex methods such as counting leading tones, and chords using current avaliable tools. In this way I believe we can come up with various ways to compare two pieces, but not being too impractical to do. The ultimate goal is also simple: output a "similarity score" for two pieces.

**Will Read:**

- Temperley, *A Bayesian Approach to Key Finding*
- Lerdahl, *Tonal Pitch Space* **Chapters ***
- Tymoczko, *Geometry of Musical Chords* **Chapters ***

## 2022-11-04

**Test on modal music** The Krumhansl-Schmuckler was tested on Bartok's *14 Bagatelles Op.47 No.1*. The algorithm regonized the right hand part as in E major/C# minor, but ignored the entire left hand in C Locrian/G Lydian. This is expected, since the algorithm only outputs one key given all notes in the piece and it seems like notes in the left hand sums up to a less duration than the right hand. So the KS algorithm put more weights in right hand than the left.
Link to a playable score:`https://musescore.com/user/4887176/scores/6403822`

**Potential Solution** We might do twice for the treble and bass clef seperately(limited to piano), and then see if the two keys are the same. If they are, then we can say the piece is in one key.

**Question** How do we determine whether we should talk about keys seperately?

**Set-Class Similarity and Fourier Transform by Tymoczko** Fourier transform assign two-dimensional vector whose components are:

$$V_{p,n} = (\cos(2\pi pn/12), \sin(2\pi pn/12))$$

Where for integer $n$ from $0$ to $6$ and $p$ in $\{0...11\}$ is the pitches in a chord. Each fourier component is the sum of all such component:

$$n\text{th Fourier Component} = \sum_{p \in v} V_{p,n}$$

- Voice leading and set-class similarity. Steps to find the minimal Euclidean voice leading between two n-note multiset-classes $A$ and $B$:
  - Choose a representative (prime form?) of $A$ calculate the sum of its pitch classes.

- Find the $n$ (12/n semitones for each) transpositions of $B$ with the same sum.
- For each of the transposition, calculate the $L_2$ norm of $A$ and the vector. Do the same for inversions.
- Take the minimum of these $2n^2$ numbers and output the result.

- Fourier Magnitude
  - In a set class space constructed by pitches of some perfectly even $n$-note chord, $n \in \{1...6\}$. Note the $n$-note chord means the chord even seperate the $12$ tone equal temperament pitches.
  - Given a $x$-note set-class space, the $n$th Fourier component of a chord will decrease as pitches move away from the subset of pitches in $n$ notes chord. Illustrated below:

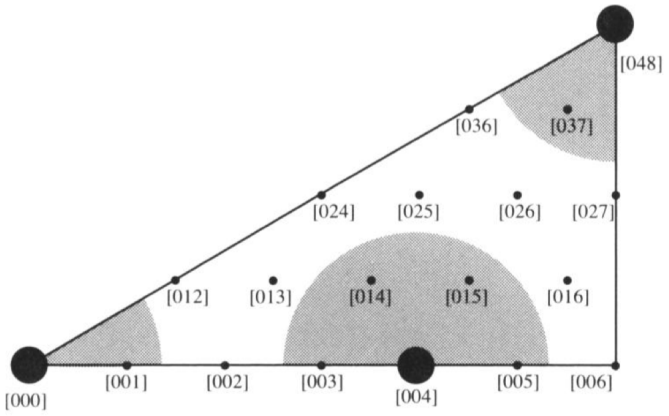



**Figure 2. Set-classes in the shaded region will have a large third Fourier component, since they are near doubled subsets of {0, 4, 8}. Those in the unshaded region will have a smaller third Fourier component.**

  - We can use a linear equation to estimate the $n$-th Fourier component of chord given the *minimal voice leading* to the nearest doubled subset(VL). This means there is a close relationship between FC and the minimal Euclidean distance between two chords.
  - 2 analysis procedure below:
  - The analogy of minimal voice leading on a pitch class circle is: as the Fourier Component increases (which is the sum of all vectors), the minal voice leading decreases.
  - FC6 is the difference absolute value of difference between the number of a chord's notes in one whole tone scale and the nuber of its notes in the other ($[0, 2, 4, 6, 8, 10], [1, 3, 5, 7, 9, 11]$).

*a*) Calculating the third Fourier component ($FC_3$) of {0, 2, 5}.

Step 1: assign the vector $(\cos 2\pi pn/12, \sin 2\pi pn/12)$ to each pitch class $p$.

$0 \to (\cos(2\pi(0 \times 3)/12), \sin(2\pi(0 \times 3)/12)) = (1, 0)$

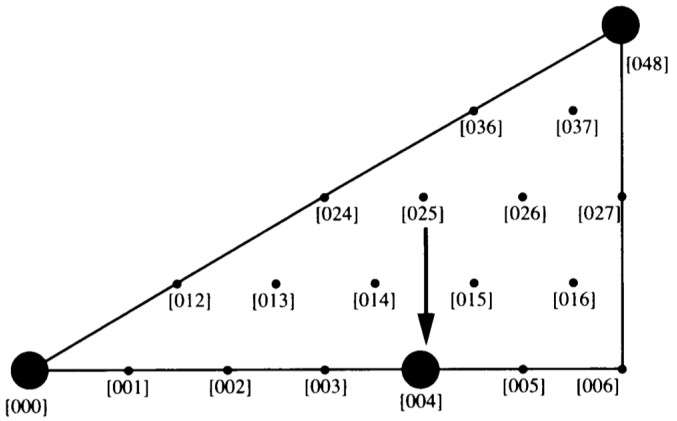
$2 \to (\cos(2\pi(2 \times 3)/12), \sin(2\pi(2 \times 3)/12)) = (-1, 0)$

$5 \to (\cos(2\pi(5 \times 3)/12), \sin(2\pi(5 \times 3)/12)) = (0, 1)$

Step 2: add these vectors: $(1, 0) + (-1, 0) + (0, 1) = (0, 1)$

Step 3: determine the length of this sum: $\|(0, 1)\| = \sqrt{0^2 + 1^2} = 1$

*b*) Determining the distance to the nearest doubled subset of {0, 4, 8}.




Following the procedure in Section I, we learn that $(0, 2, 5) \to (1, 1, 5)$ is the minimal voice leading, moving two voices by one semitone each, and with a Euclidean size of $\sqrt{1^2 + 1^2} \cong 1.41$

**Reflectiond and Questions**

1. Both Temperley and Tymoczko are trying to find the similarity in terms of "fitness" to a certain pitch set class. In Temperley's case, it is the fitness to major and minor scale, while in Tymoczko's case, it is the fitness to a certainn evenly distributed $12/n$-notes chord.

2. What is the purpose of doing this? For Tymoczko, maybe it is demonstrating certian piece confirms such minimal voice leading motion?
`http://dmitri.mycpanel.princeton.edu/ChordGeometries.html`

3. So it seems like the notion of "motion"(no puns intended) here is important, how should we show such motion and make analysis upon it?

**A proposed struture to show such motions** This is only a speculative model to show such motions in terms of intervals through time. Fix certain time untis and a pitch class space ($k$-euqal temperament) we can represent a piece of music as intervals between notes in the same voice through time. For each voice, if the piece alst for $m$ time units, there would be a 1D array of size $m - 1$ where the each entry of the array is the interval between the pitches $n_i, n_{i+1}$ on time $i, i + 1$. We can also counts the intervals between all voices on the same time (Lewin's IFUNC) and utilize his Fourier test for further analysis. Given a proper time unit, we can also represent a "segment" of notes by finding the

closest "evenly distributed pitch class set"(Tymoczko's) and then do analysis in terms of these summarized segments in correspondence with the notion of "notes, phrase, section, passage".

**Questions**

1. How do we determine a proper time unit?
2. What kind of analysis shall we use for the horizontal interval vectors?
3. This is most suitable for piano/counterpoint, what about other instruments?

**Will Read:**

- Tymoczko's Geometry of Chord Spaces
- Issacson's The Interval Angle
- Yust's Set Theory

## 2022-11-27

**Yust's Set Theory** Introduced Fourier componenets and phase spaces.

- **n Phase Space**: A circle evenly divided by $n$ interval(in the context of 12-equal temperament).
- **Fourier Component**: See summarization for Tymoczko's paper.

The process is similar to DFT, which can be seen as a matrix multiplication written as:

$$F_k = \frac{1}{n}\sum_{j=0}^{n-1} e^{-2\pi i \frac{jk}{n}}$$

Where $F_n$ is the $n$-th Fourier component, $k$ is the pitch class, and $n = 12$. Or, more generally: $\hat{f}_k = [[\mathsf{DFT}]] \cdot f_k$. Think of the process as matrix multiplication. Different phase space reveals different properties of music. $Ph_5$ indicates the a set's affinity toward diatonic collection. Each $6$ phase space approximates:

1. chromaticism
2. quartal harmony?(not sure what this means)

3. hexatonicity
4. octatonicity
5. diatonicity
6. whole-tone quality.

*Scale Theory? Eveness? Might need more explanation.* We can also use DFT to calculate the common tones of two sets.

$$\frac{1}{12}\sum_{n=0}^{11}|f_n(A)||f_n(B)|cos(\phi_n(A)-\phi_n(B))$$

.Intuitively, common tones will show harmonic closeness. Another aspect in the common tone analysis is the "wormhole" effect. Which is simply apply analysis in two phase spaces.

**Questions**

- What can I do for DFT in the interval space?
- Since DFT can be seen as a matrix operation, what does the eigen value of a DFT matrix show?
- Structure for my writing toward the end of calss?

**Will Read:** Isaacson's The Interval Angle

# Segmentation task

Segmentation is a cognitive process that "selects structually significant intervallic relations" according to Hasty. The process itself involes with decisions based on multiple aspect/"domains" of music properties such as dynamic, timbre, set class etc. There are two ways to do segmentation task using computers:

1. **Bottom-up**: Start from the smallest unit of each domain and write scripts for them. But difficulties occurs in defining herustics for each domain such as searching for "strongest" set class in given phrase.
2. **Machine Learning**: Build a machine learning model that can learn from the data and make segmentation decisions. But it is difficult to find a proper training set and the model itself is hard to define since the task of segmentation can be hard to define.

## Bottom-up

We can follow what was written in the Hasty's *Segmentation and Process in Post-Tonal Music* and start the segmentation process with a list of properties of different **domains** such as timbre, dynamic, intervalllic associations (*set class*) etc. and then develope a logic that puts these properties in a kind of hierarchy and then make decisions based on the hierarchy. For example, we can first we can list out the properties of each note in the music waiting for segmentation, and group them up according to each property. Then gradually change the grouping by searching for notes that have share more than 1 common **domain**

An advantage of this method is the whole logic requires no prior data since we can test out the logic with only a few midi sequence transcribed by human. The ***domains*** can be modulized into functions that outputs each note's specific property. However, the logic itself can be hard to define and there are a lot of scrutinies needed for the logic itself since it is a fixed set of logics that mimic a cognitive process.

## Machine Learning Model

The whole thing can be viewed as a black box that do the same thing written in the bottom-up method – grouping notes to reveal hidden structures. The advantage of this method is that the complex decision making process is done by a black box, which can be a neural network or a decision tree. This saves a lot of search into modulation of the logic. It can also handle a larger sequence/music.

However, there are two major challange for this method:

1. **Training Set**: The training set can be very hard to find since we require the information of all domains defined by us is embeded in them, let alone it will require a certain amount of data to start the training process.

2. **Model**: The model itself is hard to define since the task of segmentation is hard to define. We can only use a small set of music that is segmented by human as the training set.

## Questions

1. In the last paragraph of the paper, Hasty concluded that "Until we know eough about the nature of the structural formation in this music to be able to take more for granted , there seems to be no way of avoiding this difficulty." is there any new development in this field?

2. One of the biggest chllange for now is wirtting a *set-class finder* that can find the set class of a given set of pitches. Any paper on formulizing this process?

## 2023-09-25

# Representation and Evaluation Metrics

## Representation

The current representation for Symbolic Music Generation deep leanring models involve with two major types [1]: **piano-roll** [2] and **Event-based** [3]. Piano-roll is a relatively easy way to represent symbolic score (specifically for piano) since it is just a 2D matrix with dimension p*t where p is the number of pitches and t is the number of time steps. Event based representation contains more information in general due to MIDI's nature. A *token* can be a combination of information including pitch, duration, velocity, bar, tempo, and even instrument. However, all of them do not contain **interval information and things that can be infered from it.** Chuan et al. drew inspiration from tonnetz [4] allowing pitch from $C_0$ to $C_8$ to be represented in a 12x24 matrix, with pithes in rows in a circle of fifth relation. However, it lacks the the chordal quality in the triangle representation for triads. Ayzenberg et al. proposed a *chordal and pitch embedding* method drawing the mathmatical relation between pitch and chord in the original tonnetz and mapped pitches in a vector space where the distance between "strongly" related pitches closer to that of "weakly" related.

### Quetions

1. How to put both vertical and horizontal interval information into the representation?
2. Quick ways to crudely segment the input, mark repetitions and other important structural information?

## Evaluation Metrics

Most of the evaluation metric for generative deep learning model is subjective evalution with human test and taggings. But these are too time consuming and not scalable. So some has also developed objective metrics that only utilize math and statistical model for analysis. The main idea of the objective measurement is to compare the generated results with the ground truth. As for music generation, the current popular evaluation metrics can be divided into two categoies [1]: **rule-based** and **distribution-based**. Rule-based metrics are based on some music-relavent rules such as *harmony progression* and *melodic contour*. Yang et al. [5] propsoed a similarity comparison between generated results and training dataset. The metric first select some simple musical elements such as *pitch count, pitch class transition matrix* as "features", extrating from both datasets. Then compute the overlapping area and Kullback-Leibler divergence between two distribution. The metric is simple and computationally efficient, but it is not very musically meaningful since the features are too simple to show any musical meaning. Another limitation of the metric

is that it requires requires full access of the training dataset, which is not usually very accessible due to quality of the dataset and permission issues from the author. So here are two main issue to target for the evaluation metrics:

**Question**

1. How to improve features?
2. How to make the current metric less dataset dependent?
3. Besides instantaneous and continous interval analysis, structural analysis can also be very important, is there a rule-based metric that can perform such automatical analysis?

# Bibliography


[1] Shulei Ji, Xinyu Yang, and Jing Luo. A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges. *ACM Comput. Surv.*, 56(1), aug 2023.

[2] Luca Angioloni, Tijn Borghuis, Lorenzo Brusci, and Paolo Frasconi. Conlon: A pseudo-song generator based on a new pianoroll, wasserstein autoencoders, and optimal interpolations. 10 2020.

[3] MusicBERT: Symbolic Music Understanding with Large-Scale Pre-Training — arxiv.org. `https://arxiv.org/abs/2106.05630`. [Accessed 25-09-2023].

[4] Ching-Hua Chuan and Dorien Herremans. Modeling temporal tonal relations in polyphonic music through deep networks with a novel image-based representation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018.

[5] Li-Chia Yang and Alexander Lerch. On the evaluation of generative models in music. *Neural Comput. Appl.*, 32(9):4773–4784, may 2020.