IN-DEPTH ANALYSIS OF MUSIC STRUCTURE AS A SELF-ORGANIZED NETWORK

Ping-Rui Tsai[1], Yen-Ting Chou[1], Nathan-Christopher Wang[2], Hui-Ling Chen[3], Hong-Yue Huang[1], Zih-Jia Luo[4], and Tzay-Ming Hong[1]

[1]Department of Physics, National Tsing Hua University, Hsinchu 30013, Taiwan, R.O.C / [2]College of Pharmacy, University of Michigan, Ann Arbor, MI 48109 U.S.A.

[3]Department of Chinese Literature, National Tsing Hua University, Hsinchu 30013, Taiwan, R.O.C / [4]Advanced Semiconductor Engineering, INC., Kaohsiung 76027628, Taiwan, R.O.C

Words in natural language not only transmit information, but also evolve with the development of civilization and human migration. The same is true for music. To understand the complex structure behind music, we introduced an algorithm called the Essential Element Network (EEN) to encode the music into text. The network behind the music is obtained by calculating the correlations between scales, time, and volume. Optimizing EEN to generate Zipf's law for the frequency and rank of the clustering coefficient enables us to generate and regard the semantic relationships as words. We map these encoded words into the scale-temporal space, which helps us organize systematically the syntax in the deep structure of music. Our algorithm provides deep and precise descriptions of the complex network behind the music. As a result, the experience and properties accumulated through these processes can offer a new approach to the applications of Natural Language Processing (NLP) and a new insight to the nature of music.

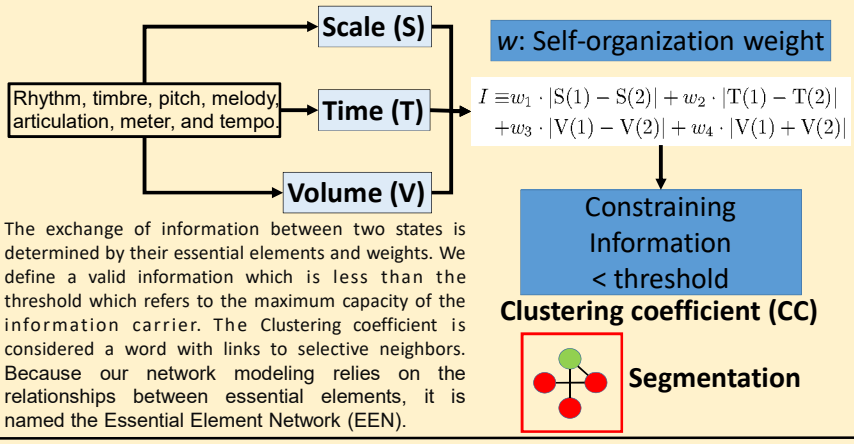
Keyword: Complex network, Music, Deep learning

Goal:

In this project, we will investigate two topics: **1. Encoding music into text, 2. Examining how music has changed across different time periods.** The above two are important problems in natural language processing and music [1]. To achieve the first goal, we must first establish definitions for relevant terms. To do so, we will follow **linguist John Rupert's principle that "a word can be understood by the company it keeps" [2]** and use the relationships between pitch, time, volume, and other factors at each point in space and time to define the amount of information between them, and then use that information to create a network of associations.

Method:

To begin with, we transform the time-frequency representation of an audio file into a scale-time-volume representation. The scale consists of 84 keys, which correspond to those of a piano in equal temperament and cover a frequency range from 1 to 8192 Hz. The time interval is 0.1 second, and the volume is expressed in normalized decibels from 0 to 10 based on the power-time spectrum in order to eliminate differences in recording quality. In information theory, We define the information (I) by comparing the scale (S), time (T), and volume (V) between 2 states in the three dimensional position:



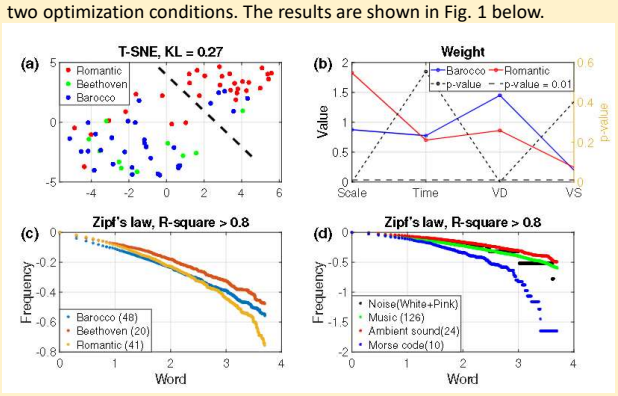
The exchange of information between two states is determined by their essential elements and weights. We define a valid information which is less than the threshold which refers to the maximum capacity of the information carrier. The Clustering coefficient is considered a word with links to selective neighbors. Because our network modeling relies on the relationships between essential elements, it is named the Essential Element Network (EEN).

Result of Self-organization and Zipf's law (Figure 1)

For the development of statistical linguistics, Zipf empirically found that the rank-frequency distribution of corpus and natural language utterances follows the power law. So we check the 4032 group of self-organization weights to achieve two optimization conditions. The results are shown in Fig. 1 below.

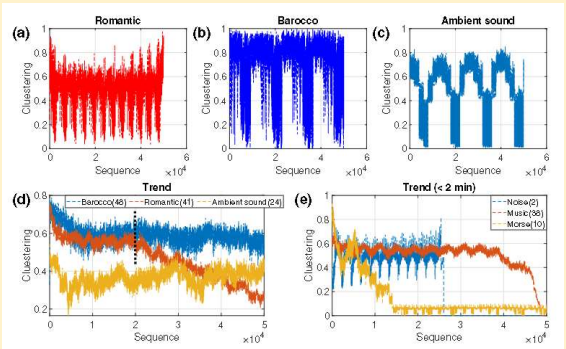
4032 group of Self-organization weight

1. The distribution of Zipf's law for CC in the music score must exhibit an R-square value exceeding 0.8 after deleting the first rank and plotting the frequency vs ranking in full logarithm.
2. The largest type of CC should be selected as the optimization condition in order to extract the maximum number of word types to maintain diversity.



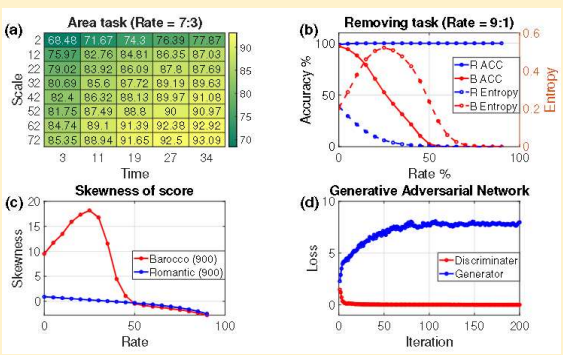
(a) A T-SNE mapping of the four weights and threshold value onto the eigenspace. The dash line is to highlight the existence of two clusters. (b) A T-test is conducted to assess the statistical significance of the weight selection. Blue and red lines denote the weight value on the left y-axis, while the black dotted line is for p-value on the right. (c) This full logarithmic plot shows the Zipf distributions in different periods, as indicated by the statistical population in parentheses. Except for Morse code, the other three audios are shown to obey Zipf's law in (d) where the source of ambient sound includes bird, river, and city traffic.

We can project words with temporal and spatial information onto their corresponding pitches and times to form a text. By concatenating the words on all pitch points whole time, we obtain a one-dimensional periodic distribution, as shown in Fig. 2. We also use a convolutional neural network (CNN) [3] with image recognition capabilities to perform two tasks on the two-dimensional text: an Area task and a Removing task. The Area task aims to determine how much resolution is needed in the two-dimensional text to distinguish between which marks the beginning and end of the common era, and Romantic music. The Removing task involves randomly removing a certain proportion of words to test the accuracy with which the original music genre is determined. We also try to generate new text by using a generative adversarial network. These tasks are shown in Fig. 3. In the Area task, we use a training-validation rate of 7:3 regardless of the resolution, and each label has 30,000 samples. The training ratio for the Removing task is 9:1, and each file has 12,000 samples.



Result of 1-dimension structure (Figure 2)

(a-c) Sampled three audio files' EEN distributions in one dimension, all of which can be found to have unique periodic characteristics. (d) Analysis of the difference between the Baroque and Romantic periods of the Common practice era and the background sound shows that there are trends of difference and diversity in compositional styles between the two. (e) the distribution of different types of audio files in one dimension.



Result of 2-dimension structure (Figure 3)

(a) difference between baroque and romantic characteristics under different sizes of areas. (b) difference in structural integrity before and after the era of mutual understanding - from pursuing preciseness to collapse. The Shannon entropy and judgment in the two periods show opposite trends with the eliminating rate of Zipf's word. (c) different skewness in score spread is shown as a key feature of grad CAM in the exploration of the two periods. (d) the learning curve of the GAN.

Discussion:

Our results demonstrate the significant differences in the arrangement of structure between Figs. 2 & 3. By projecting the essence of Zipf's word inside the Baroque music text into one- and two-dimensional structures using the Essential Element Network (EEN), we get the same result showing that the composition structure is regular. This is because the form emphasizes repetition of the same type, such as Fugue and Johann Pachelbel's Canon[4-5]. Bach also placed importance on rationality and mathematical thinking in his compositions [6]. In contrast, the Romantic period is characterized by pluralism. For example, Chopin's nocturne, a representative figure of the era, was seen as rebellious at the time.

Reference:

[1] Chuan, C. H., Agres, K., & Herremans, D. (2020). From context to concept: exploring semantic relationships in music with word2vec. *Neural Computing and Applications*, *32*(4), 1023-1036.

[2] https://www.tenpoint7.com/2020/05/three-machine-learning-lessons-learned/.

[3] Gu, Jiuxiang *et al.* (2018) "Recent advances in convolutional neural networks." *Pattern recognition* 77: 354-377.

[4] Follet, R. (2005). The Thematic Catalogue of the Music Works of Johann Pachelbel.

[5] Shedlock, J. S. (1897). The Evolution of Fugue. *Proceedings of the Musical Association*, *24*, 109-123.