Journal of Mobile Multimedia, Vol. 8, No.1 (2012) 049-071

MUSIC VISUALIZATION TECHNIQUE OF REPETITIVE STRUCTURE REPRESENTATION TO SUPPORT INTUITIVE ESTIMATION OF MUSIC AFFINITY AND LIGHTNESS

TATIANA ENDRJUKAITE

NTT Communication Science Laboratories, Japan
tatianaendrjukaite@gmail.com

NAOKO KOSUGI

NTT Communication Science Laboratories, Japan
Kosugi.naoko@lab.ntt.co.jp

Received
February 1, 2012
Revised April 15, 2012

This paper proposes a method to identify and visualize repetitive structures in a pairwise representation of music to support people to imagine their affinity for music and the lightness of music intuitively, or in other words without listening to it. Repetitive structures in this paper are fragments that a music piece contains multiple times, and all these fragments may be slightly different but are perceived as very similar. For example, a tune might have little difference in tonality and could be performed by different kinds of musical instruments. We propose an algorithm to identify repetitive structures in a tune by using a self-similarity matrix. Identified structures are visualized on two kinds of images. One is a colored cylinder of varying diameter where colors represent repetitions and the diameter represents volume changes; the other is repetitions lines image, where different pairs of repetitions are shown on the Y-axis and the duration of each repeated pair is shown on the X-axis with a color. We selected eight tunes based on music psychology to evaluate the performance of the identification and visualization technique. Finally, we found that the amount of repetitions is related to the affinity for music, but not to the lightness of music. Volumes in both high-affinity music and high-lightness music change drastically.

Key words: music visualization, repetitions in music, affinity, lightness, Misual, MFCC
Communicated by: D. Taniar & E. Pardede

1 Introduction

Nowadays, people listen to different genres of music, such as classical, popular, jazz, and so on. There are various musical features that influence people as they listen, and after they listen to music. In addition, our impression of a piece fades as we listen to it, and it is hard for us to recall our impression because music is a transient medium. Thus, we are exploring a music visualization technique to help people to roughly imagine how music sounds. Especially, we focus on the repetitive structure and affinity to music, since the number of repetitions, frequency and length may determine whether a piece will be easy to remember or hard to follow. By analyzing repetitions we may discover why a piece of music influences us the way it does. We can also imagine what influence an unknown tune would have on us before we listen to it.

This paper is a part of the Misual project [3,4,5]. In the project, we are trying to overcome the natural limitations related to the traditional way in which we become familiar with a piece of music by listening to it. Misual is a music visualization technology that aims to help people to roughly grasp a tune as a whole at first sight, by depicting multiple musical elements in a music image. It extracts data from acoustic signals to depict length, volume transitions and inner patterns of a tune with colors.

The ultimate goal of Misual research is to visualize the physical quantities of musical elements that affect impressions in images, colors or shapes, which can be understood intuitively. Visualizing music means the translation from transient media to permanent media. This makes it possible to recall one's impression of music, to compare pieces, and so on. Further, if images can be made to represent physical quantities of musical elements that can be grasped intuitively, people will be able to imagine how the music sounds before listening to it.

This paper proposes a music image that represents volume transition and similar repeated component in a piece of music. Information regarding repetition in a piece of music is important since it is related to affinity for the music, though of course the relationship between repetition and affinity varies from person to person.

This paper is organized as follows: Section 2 covers related work. Section 3 describes features extracted from music and techniques for processing them to make music visualization. Target music pieces are analyzed and an evaluation of the algorithm is presented in Section 4. Section 5 presents visualization results for pieces and a discussion based on the images. Conclusions and future work are covered in Section 6.

2 Related Work

At present, there is no complete theory for automatically analyzing music structure. However, there are encouraging researches on detecting the most frequently appearing component in a piece of music based on music structure analysis. Generally speaking, the more repetitions and similar phases there are in a piece of music, the easier it is for people to have affinity for it [3].

There are two motivations behind music visualization research: visualization of music itself and visualization as a tool to show research results on paper [2, 8]. Regarding the second motivation, researches have made "music thumbnails" and "audio summarization" by detecting the most representative part of a piece of music. The most representative part in a piece has been defined as the most frequently repeated component in it. In particular, PCM acoustic data has been used. Some researchers have found repetitions by performing self-similarity calculations with Mel-Frequency Cepstral Coefficients (MFCCs); others have identified them based on approximate transcription results [4].

In the Misual project [3, 4, 5], we want to find information extractable directly from music that, after appropriate processing, might help people to imagine and understand music. The information we seek is mainly discussed in acoustic psychology [1, 7, 9] and as such, its significance is an on-going research topic. Accordingly, this research should be based on music psychology and at the same time this research should contribute to music psychology research.

Previous researches in the Misual project have analyzed the visualizations of structure, volume and speed of music from psychological point of view. The musical information we get and visualize will be able to help people to intuitively understand the music's features in a very short time. Such music images could be used in new styles of music education and new ways of listening to music.

3 Approach

The method of music processing is outlined in figure 1, where boxes show processing steps, text without boxes describe the data that is an input and output. At the top left corner, we can see *music.wav*, which is a music file given as an input. At the bottom there are two outputs: *Misual* and *Repeats image*.

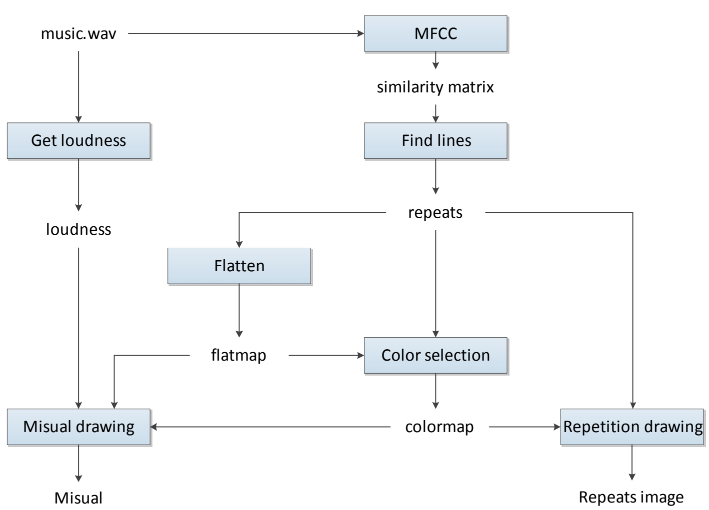


Figure 1 Scheme of processing.

Below is a short description of each processing step:

- MFCC – splitting acoustic data into small fragments and comparing all of them to each other using Mel-Frequency Cepstral Coefficients
- Get loudness – computing loudness for every fragment
- Find lines – detecting lines in a self-similarity matrix that show repetitions
- Flatten – showing all repetition pairs on one line so that colors would distinguish them
- Color selection – automatic color choosing so that similar repetition pairs have similar color

- Misual drawing – drawing Misual image
- Repetition drawing – drawing every repeat on a separate vertical level

3.1. Input Data and MFCC

We use PCM data in the mono WAV format with a 44100 Hz sampling and 16-bit quantization as input data. The first step is a paired comparison of small fragments (frames) of music using MFCC.

At this stage, the music is divided into frames of size 0.7 seconds and with a shift size 0.1 second. Each of these frames is converted using MFCC and used to calculate distance to each other. A distance equal to 0 means that the fragments are identical; the greater the value is, the less similar the fragments are. From these values we get a self-similarity matrix [2] showing the results of pair-wise comparisons of all frames. This matrix can be represented in the form of a grayscale image where each number is represented by brightness. Black points in the image shows the value of 0. Further processing is based on the data from this matrix.

3.2. Finding Lines

3.2.1. Self-similarity matrix filtering

In the self-similarity matrix, similar parts in a tune appear in the form of dark lines. To find lines automatically, we should first determine what distinguishes them from other parts of the matrix represented as an image. First, these lines should have a long sequence of diagonally arranged dark dots. Second, on both sides along the lines distance values are much higher, otherwise they would just look like dark spots that also should have diagonal sequences of dark dots inside. Thus, to find the line, we need to find a place in the picture where these two conditions are satisfied. We can take a small window and consider only the values within it. When we take a window with size of 2 by 2 points (figure 2) the distances at the points marked with letter A should be significantly smaller than those at the points marked with letter B. If this is true, considered area would look like a line, so points marked with letter A are a part of a repetition.

B_2	A_2
A_1	B_1

Figure 2 Window size 2x2.

The degree of similarity can be expressed numerically as

$$f = \frac{A_1 + A_2 + \varepsilon}{B_1 + B_2 + \varepsilon} \tag{1}$$

To avoid division by zero, a very small constant ε is added to the numerator and denominator (about 1×10^{-5}). Thus, the lower the value we get using this equation, the more desired the line is. To improve the result, the size of the window should be increased. In this work we used a window size of 3 by 3, for which the degree of similarity is expressed as

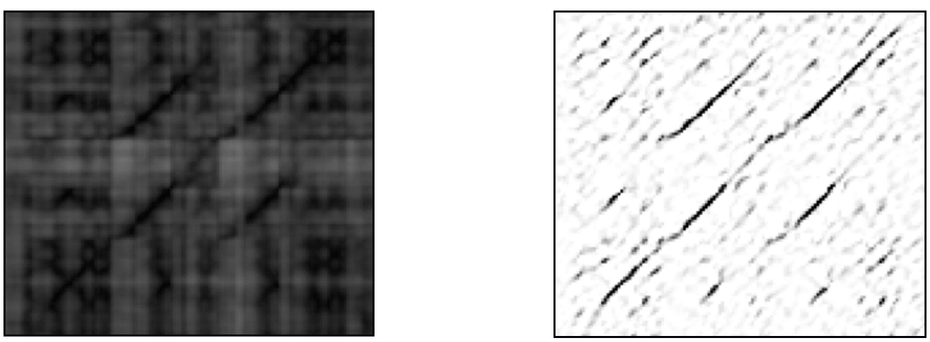
$$f = \frac{A_1 + 2 * A_2 + A_3 + \varepsilon}{B_1 + B_2 + B_3 + B_4 + \varepsilon} \tag{2}$$

	B_4	A_3
B_2	A_2	B_3
A_1	B_1	

Figure 3 Window size 3x3.

According to this formula, if the values at points marked with the letter A (figure 3) are less than the values at points B, the result is less than 1. Thus, if the result is in the range from 0 to 1, it is a potential line. Values closer to 0 indicate a stronger line, and values closer to 1 show a weakly visible line. We are not interested in values greater than 1, so they can simply be replaced by 1, where 1 means the situation when the sum of A-points is approximately equal to the sum of B-points, meaning that considered area points inside the window does not look like a line.

After this process, places with lines are strongly emphasized and it is easy to see them. An example is shown on figure 4b.



a) part of similarity matrix

b) lines detection result

Figure 4 Line parts detection results.

After line detection with a 3 by 3 window, we can better see the lines we are about to detect. These lines vary in brightness (see figure 5). More pronounced bright-dark lines indicate a strong similarity between different musical fragments. Less pronounced gray lines indicate less accurate repetition, which may be because of differences in tone or performance by different musical instruments.

Figure 5 Different similarity level comparison.

When at least inexact repetition as a grey line can be seen in an image, the listener should notice similarity between these segments.

Note that in order to improve the perception of the result we have to "lose" some information. This means that even if not very similar repetitions can be seen in this picture, we are forced to leave only the most similar repetitions to make the result easier to perceive. In other words, it would be wrong to say that a repetition was not found. A repetition could be discarded as not similar enough.

When we select only part of all lines, it is necessary to consider not only the brightness (should be dark) but also the length (should be long enough) because music may contain a lot of short repetitions that are not interesting to us at all as it is shown in figure 6. In this work 5 seconds was selected as the minimal length of repetition to be detected.

Figure 6 Many short repetitions in a music piece.

Actually, before we decide which lines are good enough to visualize, we have to find these lines in the self-similarity matrix which is just an array of numbers. This stage is the most difficult, because the lines can visually look like lines but in fact they may have discontinuities, gaps, different lightness throughout their length and other distortions. This greatly complicates the determination of a single line and some adjacent short lines that are perceived visually as one long line (acoustically as one long repetition).

We introduce two thresholds. The first threshold allows us to find the presence of the line and start looking for its boundary; the second defines the threshold at which brightness the line is considered as ended. With this approach we detect less noise but we find lines boundaries more carefully. Here we used thresholds of 0.45 and 0.49 respectively. As a result, we get a matrix with only two possible values 0 and 1 – an image with only black and white points.

Finally, to detect lines we iterate through all closed sets of black dots and search for coordinates of the bottom left point and the top right point within every set. These are the boundaries of lines – the result of the line search in the image. Suppose we have found a line with coordinates (x1, y1)-(x2, y2). This would mean that fragment (x1 - x2) is similar to fragment (y1 - y2) in this piece of music.

After selecting the most significant lines from all found lines, we get a picture such as shown in figure 7.

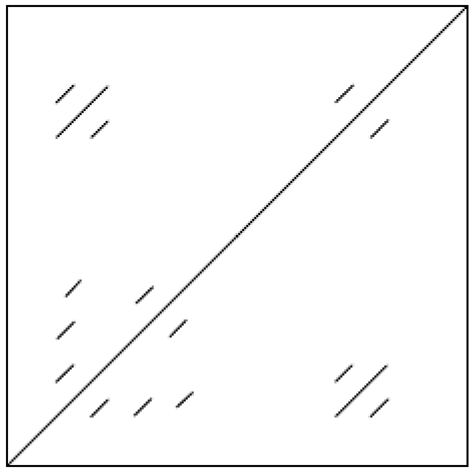


Figure 7 Example of most significant repetitions found.

3.2.2. Multiple repetitions

In the approach described here we can find repetitions by pairs, but music may contain some phrases that can be repeated three or more times. There are two ways to handle this.

One way is to look for repetitions of identical parts among all found repetition pairs and combine them into one big multiple repetition that contains more than one repetition of a fragment [3]. The merit of combining repetitions is that the result is more understandable. Without combining repetitions, we may get a lot of repetition pairs when there are a lot of identical repetitions. For example, if we have five occurrences of the fragment, we will have ten pairs of repetitions.

Another way is to show similarity of repetitions in some other manner, such as color. In this work, it was decided not to combine but to show repetitions pairwise with appropriate colors because combining repetitions cannot be done without loss of information and accuracy deterioration. Repetitions may be similar in a chain when every new repetition is similar to previous one, but the first and the last repetition could be very different. In pairwise representation this information is visible.

3.3. Results Visualization

Figure 8 shows repetitions painted in different colors for better visual distinction. Different pairs of repetitions are shown on different vertical levels. This is necessary because repetitions may overlap.

Figure 8 Repetition diagram example.

Having several levels (vertically) complicates the perception, especially when there are many levels. Therefore, to facilitate the perception, all of these colored lines can be shown on one axis. This way of musical visualization is called Misual and is described in papers [3, 4].

To improve the readability of the result, another version of visualization has been proposed. Misual image at the top is a main visualization subject and few axes with representation of each separate repetition below. The upper part gives a general picture of the tune, and the lower part determines the locations of repetitions, which often cannot be understood on one axis of Misual due to the overlapping of repetitions. See Section 5 for examples.

3.3.1. Color selection

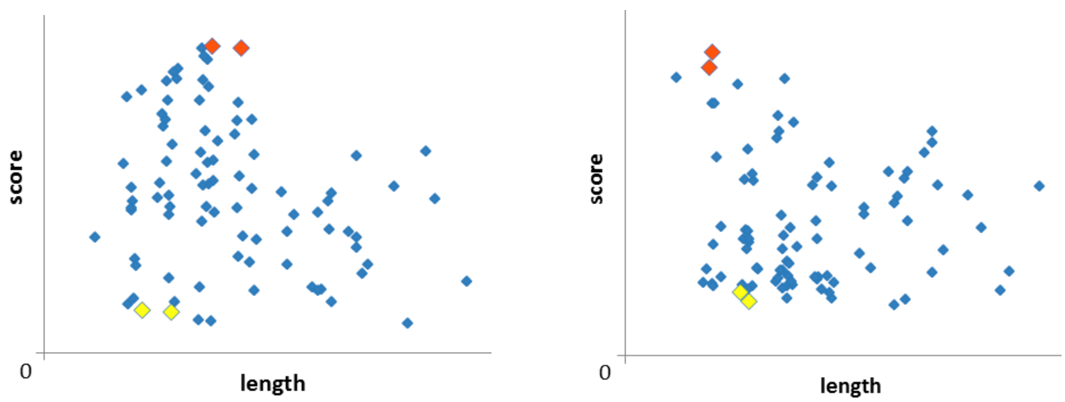
Once we have found a set of repetitions in a piece we should choose colors so that it would make the result more understandable and easier to perceive. When a phrase is repeated many times, we need to choose similar colors for every such repetition pair since the repetitions are similar. To do this, first of all, we must determine how all of the repetitions are similar to each other. Repetition similarity is calculated as maximal similarity of a shorter fragment inside a longer fragment using SSD (The Sum of Squared Differences) [10]. Comparing repetitions to each other, we get a distance matrix that shows distances of repetitions. Using this matrix we can find the best disposition of some abstract points on the plane so that their distances are similar to the defined in the matrix. This could be solved using any of the dimensionality reduction approaches that are widely used in statistics and data mining [11, 12].

In this work we used an iterative approach. Each iteration performs a slight move of all points in the direction of the desired location according to the distance matrix. Finally, the system comes to a stationary state when all points stop moving. This is considered to a result of points' disposition. After that the plane with the points can be represented as a part of the HSI space (Hue, Saturation, Intensity) [13], actually in the form of a plane Hue – Intensity with Saturation equal to maximum color saturation.

Figure 9 Map of colors.

With this choice of colors, similar repetitions (represented as points) will be close to each other on the plane, and the colors for the repetitions will therefore be similar. If several pairs of repetitions are somehow similar, then after selecting the colors, we will see them as similar because the colors will be similar. This way of selecting colors may not work very well if too many repetitions that are just similar but not equal are detected. Another disadvantage of this approach is that it is heuristic and may give different results as many times as we start it. Despite these drawbacks, automatic color selection makes the results much more readable, but we need to be careful with it since a difference in colors does not definitely mean a difference in musical fragments.

4 Target Music Analysis and Experimental Results of Repetition Identification



a) Affinity b) Lightness

Figure 10 Score distribution for tunes selection.

We have been discussing the relationship between affinity to music and the amount of repetitions in it [3]. In this paper, we selected target music on the basis of the research on music psychology by Taniguchi [1], who selected 90 tunes and evaluated them with respect to six features: affinity, lightness, a sense of exaltation, strength, majesty and likeness.

Table 1 Music pieces used in the paper.

Short name	Type	Description
BA	High affinity	Air on G string Orchestral Suite No3 in D major by Bach, [14]
MMT		Meditation for Thais by Massenet, [15]
C2511	Low affinity	Etude Op. 25 No. 11 by Chopin, [16, 17]
VS3		Four Seasons: Summer 3rd movement by Vivaldi, [18, 19]
AP	High Lightness	Plink Plank Plunk by Anderson, [20]
ABH		Bugler's Holiday by Anderson, [21]
SG3	Low Lightness	Gnossienne 3 by Satie, [22, 23]
SG4		Gnossienne 4 by Satie, [22, 24]

The evaluation was performed by 50 people. We focused on two features: affinity and lightness. Figure 10 shows the score distribution for affinity and lightness of each tune.

For each feature we chose two tunes from the top, with the highest scores, and two tunes from the bottom, with lower scores and not long durations of the tunes. Figure 10 also shows the selection results of eight target tunes. Red squares mean that these tunes display pieces of music with high scores and yellow squares relevant to tunes with low scores according to Taniguchi. Table 1 shows the selected tunes and their short names. We use the short names hereafter in this paper. Eight pieces of music were visualized with the method described in Section 3.

Detection result for BA

Table 2 shows the repetitions in the music score and their performances time in CD.

The number of bars and the performance times of each part are as follows.

1. From the 1st to 6th bar and from the 7th to 12th bar are the same. Performed from 0 to 43.2 seconds and from 47.2 to 90.1 seconds accordingly.
2. From the 13th to 24th bar and from the 25th to 36th bar are the same. Performed from 94.6 to 185.0 seconds and from 186.5 to 278.3 seconds accordingly.

Table 2d shows the repetition detection results, where letters correspond to those in Table 2. For example, we can see that part 1A in Table 2 is detected as a part that begins from 0.2 second and ends at 43.1 seconds.

Hereafter, starting from Table 2 and ending with Table 9d detected repetitions and ground truth repetitions information are represented in the same format.

Table 2: Ground truth repetitions in BA

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	1 – 6	0 - 43.2
	B	7 – 12	47.2 - 90.1
2	C	13 -24	94.6 - 185.0
	D	25 -36	186.5 - 278.3

Table 2d: Repetition identification result for BA

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	0.2 - 43.1	B	47.0 - 90.0	ok
2	C	92.7 - 182.1	D	184.9 - 274.3	ok

Detection result for MMT

Table 3: Ground truth repetitions in MMT

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	3 - 10	11.3 – 47.1
	B	40 - 47	174.5 - 211.1
2	C	15 - 20.5	63.9 - 96.7
	D	52 - 57.5	229.9 - 258.8
3	E	3 - 4	11.3 – 19.0
	F	11 – 12	47.8 – 56.4
	G	40 - 41	175.1 – 182.1

Table 3d: Repetition identification result for MMT

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	E	11 - 19.2	F	47.7 - 56.5	ok
2	in A	19.4 - 48.6	in B	184 - 213.2	ok
3	F	48 - 56.6	G	175.7 - 182.6	ok
4	C	64.9 - 97.0	D	229.7 - 260.9	ok
5	-	252.4 - 260.9	-	275.5 - 284	not in GT
6	-	301.6 - 308.7	-	328.7 - 335.8	not in GT
7	-	303.4 - 315.8	-	322.2 - 334.6	not in GT

Detection result for C2511

Table 4: Ground truth repetitions in C2511

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	5 - 7.5	23.9 - 30.5
	B	13 - 15.5	38.6 - 46.5
	C	69 - 71.5	151.6 - 159.1
	D	77 - 79.5	167.8 - 176.0
	E	23 - 25.5	58.5 - 66.5
	F	31 - 33.5	75.0 - 83.0
2	G	5 - 15.5	23.9 - 46.5
	H	69 - 79.5	151.6 - 176.0

Table 4d: Repetition identification result for C2511

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	22.4 - 30.5	B	38.5 - 46.5	ok
2	A	23.0 - 30.9	E	58.6 - 66.5	ok
3	-	26.9 - 33.8	-	78.2 - 85.7	not in GT *
4	G	22.6 - 46.3	H	151.5 - 175.2	ok
5	B	38.7 - 46.4	C	151.5 - 159.2	ok
6	E	59.4 - 67.3	F	75.0 - 82.7	ok
7	C	151.5 - 159.7	D	167.8 - 175.9	ok
8	A	22.4 - 30.7	D	167.6 - 175.9	ok

* - explanation is in figure 11.

Detection result for VS3

Table 5: Ground truth repetitions in VS3

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	10 - 17	12.5 - 22.0
	B	101 – 108	122.0 - 132.0

Table 5d: Repetition identification results in VS3

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	12.7 - 22.3	B	121.8 - 131.4	ok

Detection result for ABH

Table 6: Ground truth repetitions in ABH

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	9 - 22.5	5.7 - 16.1
	B	35 - 38.5	16.4 - 28
2	C	59 – 74	43.0 - 54.6
	D	141 – 156	104.0 - 116
3	E	75 – 88	54.8 - 64.8
	F	157 – 170	116.3 - 126.3
4	G	97 – 104	70.1 - 75.7
	H	113 – 120	83.0 - 87.9

Table 6d: Repetition identification result for ABH

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	5.2 - 16.0	F	115.4 - 126.5	ok
2	A	5.6 - 15.4	-	11.2 - 20.6	not in GT
3	A	5.2 - 16.2	B	17.0 - 28.0	ok
4	AB	9.2 - 28.9	CE	46.4 - 66.2	ok
5	CE	39.5 - 65.0	DF	100.8 - 126.5	ok
6	A	5.2 - 16.1	E	53.9 - 65.2	ok
7	G	69.6 - 76.8	H	81.9 - 88.9	ok
8	AB	9.2 - 27.8	DF	108.2 - 126.5	ok

Detection result for AP

Table 7: Ground truth repetitions in AP

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	4.5 - 19 f	3.1 - 15.4
	B	20.5 - 19 p	15.9 - 27.2
	C	37.5 – 52	40.1 - 50.7
	D	37.5 – 52	62.4 - 72.9
	E	75.5 – 90	99.6 - 110.5
	F	107.5 - 122	122.0 - 133.4
	G	123.5 - 122	133.5 - 143.1
2	H	21 – 36	27.4 - 38.9
	I	53 – 36	51.0 – 62.0
	J	91 – 106	110.9 - 121.7
3	K	55 - 70 f	74.0 – 85.0
	L	55 - 70 p	85.2 - 96.4

Table 7d: Repetition identification result for AP

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	1.6 - 15.2	E	97.7 - 110.5	ok
2	A	2.7 - 15.5	D	61.4 - 73.8	ok
3	A	2.8 - 15.2	C	38.5 - 50.5	ok
4	A	3.3 - 14.3	G	133.5 - 143.8	ok
5	A	3.3 - 14.8	B	15.5 - 26.9	ok
6	B	15.5 - 26.8	F	121.9 - 132.9	ok
7	B	15.5 - 26.9	D	62.1 - 73.1	ok
8	AHC	15.5 - 50.3	CID	39.0 - 73.2	ok
9	C	39.0 - 49.7	G	133.5 - 143.9	ok
10	D	61.7 - 73.1	E	99.0 - 110.4	ok
11	D	62.1 - 72.5	G	133.5 - 143.9	ok
12	K	74.1 - 85.2	L	85.5 - 96.6	ok
13	E	98.9 - 109.7	G	133.3 - 143.8	ok
14	E	99.0 - 110.6	F	121.6 - 133.3	ok
15	AHC	15.4 - 50.4	EJF	99.1 - 133.2	ok
16	AB	2.7 - 26.3	FG	121.2 - 143.9	ok
17	F	121.9 - 132.3	G	133.5 - 143.8	ok
18	CID	38.7 - 73.7	EJF	99.0 - 133.5	ok

Detection result for SG3

Table 8: Ground truth repetitions in SG3

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	1 – 3	0 - 9.5
	B	5 – 7	11.5 - 22.4
	C	46 – 48	148.1 - 154.9
	D	50 – 52	163.4 - 170.9
2	E	1 – 7	11.7 - 22.5
	F	46 – 52	144.7 – 155.0
3	G	16 – 18	49.8 - 59.5
	H	20 -22	62.2 - 71.7
4	I	24 – 28	75.4 – 91.0
	J	30- 34	92.5 – 107.0
5	K	39 – 41	119.2 - 132.8
	L	43 – 45	132.4 - 145.2

Table 8d: Repetition identification result for SG3

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	0.1 - 9.6	B	11.6 - 22.3	ok
2	AB	0.1 - 22.6	CD	144.9 - 169.9	ok
3	E	11.6 - 22.4	F	144.8 – 155	ok
4	-	27.3 - 34.8	-	138.0 - 145.2	not in GT
5	G	49.9 - 59.4	H	62.1 - 71.5	ok
6	I	75.4 - 90.9	J	92.7 – 107	ok
7	K	119.0 - 132.7	L	132.5 - 145.2	ok
8	C	148.2 - 154.9	D	163.2 - 170.8	ok

Detection result for SG4

Table 9: Repetitions in SG4

ID	GT Mark	Repetition (bar)	Performance (second)
1	A	1	0.0 - 5.9
	B	2	6.0 - 11.8
	C	19	102.5 - 107.8
	D	31	180.0 - 187.9
2	E	3 – 4	13.4 - 23.0
	F	6 – 7	30.5 - 40.4
	G	22- 23	120.0 - 133.3
3	H	8 – 9	40.8 - 51.0
	I	10 – 11	51.4 - 62.6

Table 9 continue: Repetitions in SG4

ID	GT Mark	Repetition (bar)	Performance (second)
4	J	11 – 12	58.1 - 68.6
	K	24 – 25	133.4 - 146.7
5	L	13	69.2 - 74.5
	M	15	80.6 - 85.7
6	N	26 – 27	147.4 - 159.2
	O	28 – 29	160.0 - 172.7
7	P	19 – 20	102.5 - 114.4
	Q	31 – 32	180.1 - 195.4
8	R	14	74.5 - 79.7
	S	17	91.3 - 96.8
9	T	18-20	97.1 - 113.2
	U	30-32	173.6 - 195.1

Table 9d: Repetition identification result for SG4

ID	Source part		Detected similar parts		Evaluation
	GT Mark	from/to (sec.)	GT Mark	from/to (sec.)	
1	A	1.4 - 9.2	L	64.8 - 72.3	partially ok
2	AB	2.9 - 11.7	T	99.3 - 108.1	partially ok
3	AB	2.7 - 11.9	U	176.4 - 186.9	partially ok
4	BE	7.7 - 25.4	G	115.8 - 133.2	partially ok
5	E	16.0 - 24.3	F	33.3 - 41.6	partially ok
6	F	29.3 - 41.5	G	118.4 - 132.2	partially ok
7	H	39.6 - 50.5	I	50.2 – 62.0	partially ok
8	J	57.4 - 70.9	K	132.9 - 146.5	partially ok
9	R	71.9 - 80.4	S	87.7 - 96.4	partially ok
10	ST	92.2 - 106.0	O	167.6 - 185.1	partially ok

4.1. Experimental Results Discussion

For piece C2511 in Table 4d with detected repetitions there is a repetition number 3 that is not in Table 4 with ground truth data. This is rather common situation, and we will give an explanantion as an example only for this fragment. To find out how similar those fragments actually are, we show scores of detected repetition in figure 11. In the figure we can see two fragments of the score these are detected as similar fragments and the stripe between them shows how similar the scores are. The green parts of the stripe show that corresponding score parts are similar enough; the red parts shows different parts. Looking at the whole stripe we can say that disregarding the two small different fragments, this repetition is similar enough. So despite the repetition absence in ground truth data it is a good point to show the repetition in the result.

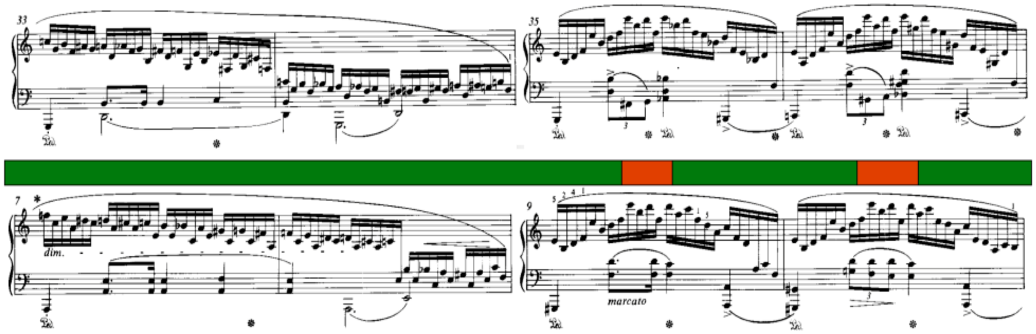


Figure 11 Detected repetition that was not in Table 4. Green – similar, Red – different.

Now let's consider the SG4 piece. It is a little bit difficult to compare data in Table 9d and Table 9 with detected repetitions and ground truth data accordingly. Figure 12 shows a graphical comparison of ground truth data (white lines below) with detected repetitions (color lines above). In the figure 12 we can actually see that repetition boundaries are sometimes not very precise but in most cases repetitions are detected correctly. This piece of music is performed very slowly and quietly. It has many nested repetitions. Quiet moments are usually detected as similar and this makes the repetition boundaries inaccurate in the result.

Figure 12 Comparison of ground truth data (below) and detected result (above) for SG4 piece.

Overall, based on these experimental results we can say that we have achieved very good results for most tunes. Almost all repetitions were detected by the provided algorithm. Exceptions were only repetitions with rather big difference in performance for the same scores. Moreover, the algorithm also detected as similar those repetitions that have inessential changes in scores or in performance and those that have tonality shifts. This ability is very important because repetitions often have small changes that usually are not perceived by a listener. Thereby, low sensitivity to small changes makes the algorithm more robust.

5 Visualization Results and Discussion Based On The Images

Figures in this section show visualization results for the pieces described in Section 4. The X-axis represents time and the Y-axis represents the volume in the upper images. Black intervals mean that those fragments have no similar parts within the piece.

5.1. Affinity Results Discustion

5.1.1. High affinity score

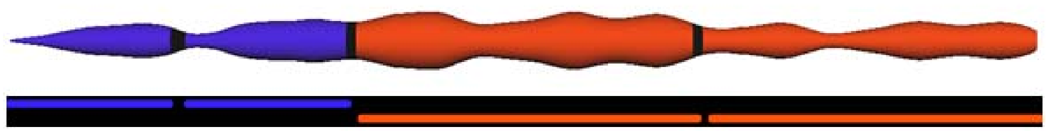


Figure 13 Visualization of BA.

Figure 13 represents the visualization of BA. This piece has got the higher score in the Affinity table in the Taniguchi book [1]. It's Misual shows us that it starts at low volume and the volume increases in time but very smoothly. In regard to repetition identification, in the Misual image we can see two kinds of repeated parts, since there are two colors: blue and red. Also, these repeated parts are spread-eagled through the piece which means that BA consists of repetitions only. Thus, someone might imagine that it would be easy to memorize the piece and also be affected by it.

Based on analysis result, the blue parts are the 1st repetition. We can see the 1st repetition is performed twice, since there are two blue parts and their shapes are similar. Moreover, we can see two red parts in the Misual image. The red parts are the 2nd repetition. From their shapes we can see differences in volume during performance. The first red part looks like a doubled second red part.

From the visualization, we can intuitively imagine that the first red part is the most powerful and emotional part in the musical composition. We also can notice that the blue parts are two times shorter in time than the red parts.

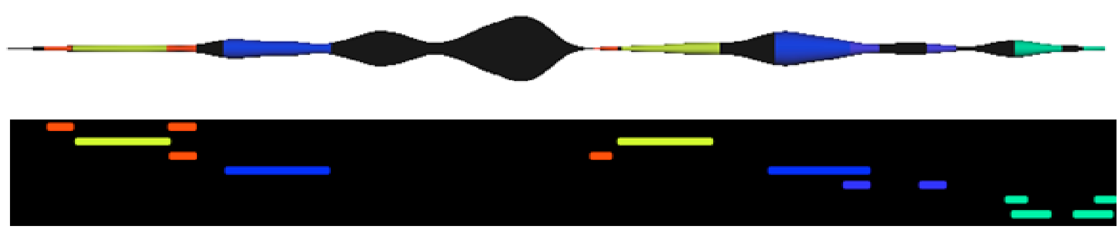


Figure 14 Visualization of MMT.

Figure 14 represents the visualization of MMT. It's Misual shows us that the volume of this tune changes dynamically. In regard to repetition identification, in the Misual image we can see two kinds of long repeated parts, yellow and blue. We can see that yellow and blue repeated parts are played almost at the same volume. Also, the order of appearance is the same: first yellow and then blue.

Actually, this piece contains long repetition red-yellow-red-blue at the beginning with slightly different red-yellow-nothing-blue in the middle of piece. Because of the difference, this big repetition has been found as the couple of smaller repetitions. If we assume that this big repetition would be found, we could say that this piece consists of long repetitions.

What is common for BA and MMT – high affinity score

Pieces of music with a high affinity score mainly consist of long repetitions. From the visualizations, we can intuitively imagine that in tunes with a high affinity score, there are long and simple repetitions without overlapping and nesting. Thus, someone might quickly recall their emotions while listening to the piece and even recall the melody, because changes are infrequent.

5.1.1. Low affinity score

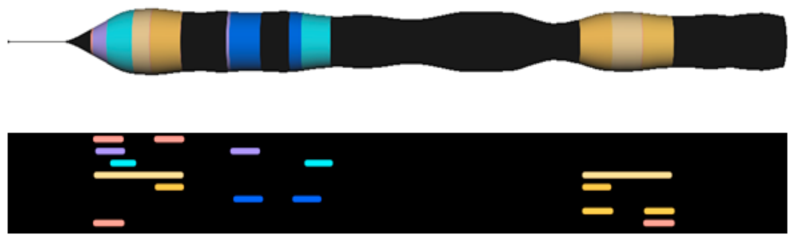


Figure 15 Visualization of C2511.

Figure 15 represents the visualization of C2511. It's Misual shows that it starts at low volume but after that, the volume gradually increases and it is kept almost the same.

This piece has some short repetitions, but most part of it has no repetitions. We can see a lot of black on its Misual. Most repetitions appear in the beginning part.

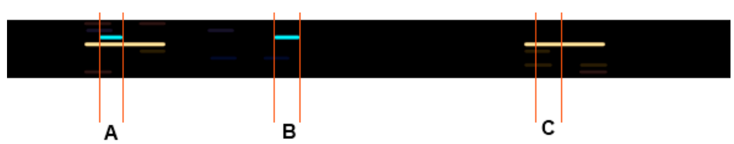


Figure 16 Transitiveness does not always works for similar repetitions.

In figure 16 we can see the following:

- Sky blue repetition shows that A ~ B (A is similar to B)
- Yellow repetition shows that A ~ C (A is similar to C)

So we have B ~ A and A ~ C, but this does not mean that B is similar to C (B ~ C). Transitivity does not work here since we show similarities, not equal parts. B is not similar enough to C, though it was not shown as pair repetition.

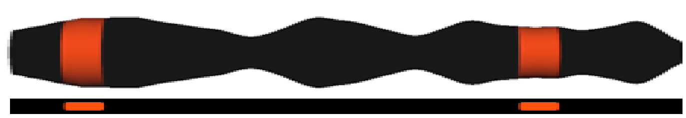


Figure 17 Visualization of VS3.

Figure 17 shows the visualization result for VS3. In regard to the volume transition, it is clear that this piece starts loud. Some listeners may feel an impact at the beginning. There are two huge volume transitions in the music. However, the entire final part is performed at a loud volume. Thus, some listeners might keep feeling a strong effect after the performance.

In regard to repetition identification, we can see two red strips on its Misual. Thus, listeners understand that red parts are the sole repetitions repeated twice in the piece, once at the beginning and again almost at the end of the tune. We can see that there are no more repetitions inside the tune. It seems that it may therefore be a little difficult for people to recall the piece from memory.

What is common for C2511 and VS3 – low affinity score

Pieces with a low affinity score may be distinguished by a lot of black parts in theirs Misuals. In other words such pieces of music would mostly contain parts with no repetitions. This may be one of the reasons that people feel a low affinity for the tunes.

5.2. Lightness Results Discustion

5.2.1. High lightness score

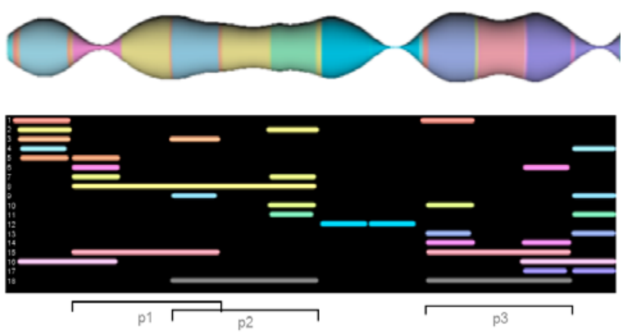


Figure 18 Visualization of AP.

Figure 18 shows the visualization result for AP. In regard to the volume transition, it is clear that this piece starts at a large volume and the volume powerfully changes three times during the performance.

Repetitions number 8, 15 and 18 in Table 7d (repetition numbers 1 – 18 are shown on the left) shows that these parts are similar:

- in repetition number 8 we can see that parts p1 and p2 are similar (note that these parts are overlapped);
- in repetition number 15 we can see that parts p1 and p3 are similar;
- in repetition number 18 we can see that parts p2 and p3 are similar.

The lines figure for AP shows that the piece mainly consists of similar repetitions: one fragment is repeated seven times, another one is repeated three times (repetitions 8, 15, 18), and the third one is repeated two times (repetition number 12). Alphabetically we can express this as AABABACCABAA. These repetitions make up almost the whole piece.

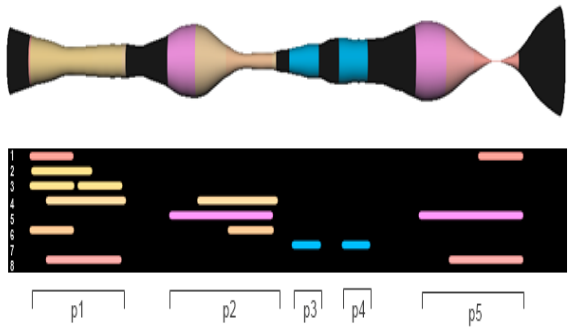


Figure 19 Visualization of ABH.

Figure 19 shows the visualization result for ABH. In regard to the volume transition, it is clear that this piece also starts at a large volume and the volume changes dynamically during the performance. The entire final part is performed at a very loud volume. Thus, some listeners might keep an uplifted mood after the performance.

In the ABH Misual it is clear that there are five parts, p1 – p5, separated by black parts. Parts p1, p2 and p5 are similar as we can see in the lines image. Repetitions number 4, 5 and 8 (repetition numbers 1 – 8 are shown on the left) shows that these parts are similar:

- in repetition number 4 we can see that parts p1 and p2 are similar
- in repetition number 5 we can see that parts p2 and p5 are similar
- in repetition number 8 we can see that parts p1 and p5 are similar

Alphabetically we can express this as AABBA (B represents blue repeated parts that are much shorter than A parts). These repetitions make up more than for half of the whole piece.

What is common for ABH and AP – high lightness score

We can see that these two pieces have most of the performance in similar repetitions. These similar repetitions are most representative parts in pieces. Thus, it makes it easy to perceive. Moreover, in their Misuals we almost do not see black parts, so that someone might feel familiarity with the pieces much faster.

5.2.1. Low lightness score

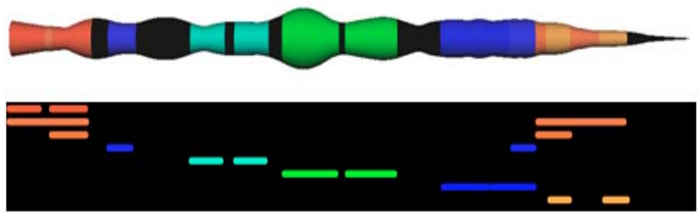


Figure 20 Visualization of SG3.

Figure 20 represents the visualization of SG3. In regard to the volume transition, it is clear that this piece starts loud and the end of the tune is very quiet.

Here we can see that in this piece there are four different parts with repetitions:

- red repetitions at the beginning and in the end of the piece
- two light-blue repetitions
- two green repetitions
- and two long dark-blue repetitions

These four parts are different from each other since there are no lines in the figure that show similarity. Alphabetically we can express the structure as AABCCDDBBAA. In this piece there is no dominant repetition. The whole piece is uniformly divided into those four different repetitions.

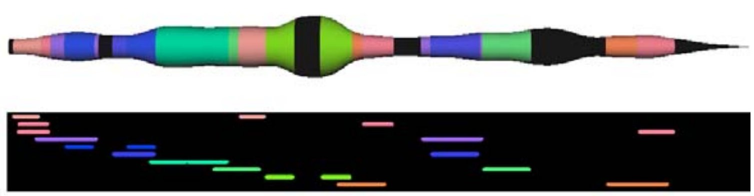


Figure 21 Visualization of SG4.

Figure 21 represents the visualization of SG4. In regard to the volume transition, it is clear that this piece starts not loud and the end of the tune is very quiet. There are few changes in volume.

In regard to the repetition, many repetitions are nested and overlapped. Thus, this tune seems to be more difficult to grasp, since listeners may be confused as to whether some parts are repeated or not. Alphabetically we can express the structure of the piece as ABBCCADDABA, which means that inside this piece there is no dominant repetition. The whole piece of music is uniformly divided into different repetitions.

What is common for SG3 and SG4 – low lightness score

Both pieces have many and short different repetitions within. Also, for example, SG4 has overlapped and nested repetitions during the whole performance. Moreover, there is no the most representative part in these tunes. That's probably why they are located at the lower step in the Taniguchi book [1].

6 Conclusions and Future Work

This paper described a new method of identifying and visualizing repeated patterns in music. The method is based on self-similarity matrix processing. A self-Similarity matrix contains similar repeated fragments as recognizable diagonal lines that are pairwise repetitions. These repetitions are automatically detected with the proposed algorithm and used to display a visualization of music by means of repetitions lines images and by means of Misual. The advantages of pairwise repetitions are fullness and high accuracy compared to visualization with multiple repetitions. Repetition colors are used to provide information about repetitions similarity. Misual produces images of music based on the length of the tune, volume transitions, and repetitions in one image. Thus it helps to understand duration of the piece. Further, represented volume transition pattern data in Misual images makes it easy to perceive information about fragments in a tune and in the whole piece of music. For each repeat component color is allocated and painted on Misual.

For improvement in detecting and visualizing images, color selection for each repetition should also take into account the level of similarity for detected repetition pair, since the pairs are similar but not the same. All the detected repetitions pairs could sometimes be grouped into one multiple repetition of a fragment, which could make the result more understandable and easy to perceive.

Acknowledgements

This research is granted by the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for Next Generation World-Leading Researchers (NEXT Program)", initiated by the Council of Science and Technology Policy (CSTP).

References

1. T. Taniguchi. Ongaku to Kanjoh (Music and Emotion). Kitaohji-Shobo, 1998, a book in Japanese.
2. J. Foot. Visualizing music and audio using Self-Similarity. In Proc. of ACM Multimedia '99, pages 77-80, November 1999.
3. N. Kosugi. Misual – Music Visualization based on acoustic data. In Proc. of the 12th International Conference on Information Integration and Web-based Applications & Services (iiWAS2010 Best Paper Award), pages 607 – 614, 2010.
4. B. Puzon, N. Kosugi. Extraction and Visualization of the Repetitive Structure of Music in Acoustic Data –Misual Project–. Proc. of the 13th International Conference on Information Integration and Web-based Application & Services, (Dec. 2011).
5. A. Gumilia, B. Puzon, and N. Kosugi. Music Visualization –Misual Project–. To appear in ACM Multimedia 2011.
6. B. Logan. Mel Frequency Cepstral Coefficients for Music Modeling. In International Symposium on Music Information Retrieval, 2000.
7. P. Lindsay, N. Donald. Human Information Processing: Introduction to Psychology. Academic Press Inc; 2nd edition, 1977, a book in English.

8. M. Cooper and J. Foot. Automatic music summarization via Similarity Analysis. In Proc. of ISMIR 2002, pages 81 – 85, 2002.
9. D. Levitin. This Is Your Brain on Music: The Science of a Human Obsession. Dutton Adult, a book in English.
10. J. Modersitzki. Numerical methods for image registration. Oxford University Press, 2004, a book in English.
11. T. Hill and P. Lewicki. Statistics: Methods and Applications. StatSoft Inc.; 1 edition, 2005, a book in English.
12. J. Han and M. Kamber. Data mining: Concepts and Techniques. Morgan Kaufmann, 2001, a book in English.
13. G. Joblove and D. Greenberg. Color spaces for computer graphics. Computer Graphics. In Proc. of SIGGRAPH '78, 12(3):20-25, 1978.
14. Bach, Johann Sebastian, Air On the G String (Aria), Licence: Art Libre, music score.
15. MASSENET ,Meditatiion de "THAIS", NIPPON GAKUFU SHUPPANNSHA, music score.
16. Chopin 12 Etudes, Performance by Pollini Maurizio, piano, Universal Music Classics, CD number: UCCG – 7071.
17. Chopin, Etude Opus 25 No. 11 – Allegro con brio, in A Minor, Berlin: Bote&Bock, music score.
18. Vivaldi four seasons, Pina carmirelli – violin, PHILIPS, CD number: PHCP – 9001.
19. Vivaldi, Four seasons, Summer III.Presto, music score.
20. Leroy Anderson, Plink, Plank, Plunk!, Alfred Publishing Co., Inc, music score.
21. Leroy Anderson, Burghley's Holiday, Alfred Publishing Co., Inc, music score.
22. Erick Satie, Performance by Roger Pascal, piano, DECCA-RECORDER, CD number: POCL – 4697.
23. Satie, Gnossienne No. 3, Lerolle &Cie. music score.
24. Satie, Gnossienne No. 3, Rouart, Lerolle &Cie. music score.