GRAPE: A Gradation Based Portable Visual Playlist

Tomomi Uota
Ochanomizu University
Tokyo, Japan
Email: water@itolab.is.ocha.ac.jp

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
Email: itot@is.ocha.ac.jp

***Abstract*—Thanks to recent evolution of portable music players featuring large storage spaces, we tend to carry large number of tunes. It often makes us more bothering to look for tunes which we want to listen to. On the one hand, we usually listen to the tunes on the music players by manually selecting playlists or album names, rather than manually selecting each tune one-by-one. This paper presents GRAPE, a playlist visualization technique being used as user interfaces on the music players. GRAPE presents a set of tunes as a gradation image, by assigning colors to the tunes based on their musical features, and placing them onto a display space by applying Self Organizing Map (SOM). This paper describes the processing flow of GRAPE, and introduces user evaluations to demonstrate the effectiveness of GRAPE.**

***Keywords*-Music visualization, Self organizing map, playlist.**

I. Introduction

We can carry large number of tunes thanks to the evolution of mobile music players featuring large storage spaces. On the other hand, it is often difficult to remember the contents of such large music collections. We had a preliminary questionnaire what kinds of operations are often used to select tunes on the mobile music players. As shown in Table I, many of us usually select "playlists" describing sets of tunes, rather than selecting tunes one-by-one. The top three choices in the result denote that users select the sets of tunes by a single operation. We expect that visualization of playlists will assist the ordinary operations for selecting tunes on the mobile music players in our daily life.

Visualization is an effective approach to quickly understand the contents of such large music collections in a short time. A tutorial material on music visualization [1] introduces recent works on visualization of musical information. However, many of existing works focus on visualization of individual tune, or large sets of tunes, based on artist or genre information. The tutorial material [1] introduced no visualization works which represented playlists.

This paper proposes GRAPE (GRadation Arranged Playlist Environment), a playlist-by-playlist music visualization technique running on personal computers and Android devices. Figure 1 shows a snapshot of the implementation of GRAPE on an Android device. GRAPE displays the playlists as gradation images to represent both the features of playlists themselves and each of tunes simultaneously.

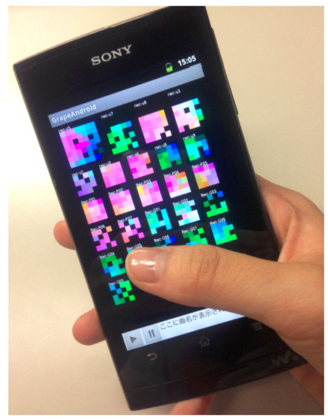

Figure 1. User interface of GRAPE as an Android application.

Table I
QUESTIONNAIRE: OPERATIONS USED TO SELECT TUNES ON THE MOBILE MUSIC PLAYERS.

Choice	Number of agreed answerers (%)
Select names of albums/artists	35
Use shuffle play	22
Select manually created playlists	21
Select tunes one-by-one	16
Never use portable music players	5
Develop useful applications	1

GRAPE calculates musical feature vectors to assign colors to the tunes, and place the tunes on the display space. Consequently, GRAPE represents a playlist as a collection of colored square tiles corresponding to the tunes. GRAPE applies Self Organizing Map (SOM) to calculate the positions of tunes so that similar tunes get closer. Against the typical user interfaces of music players just display titles of playlists and tunes as textual information, GRAPE intuitively represents features of the tunes in a playlist.

Playlist-based music visualization is especially useful under many situations. One situation is while using playlists consisting of automatically collected tunes, not album- or artist-based playlists. Music listeners often use such playlists including collections of recently downloaded tunes, frequently played tunes, or application-recommended tunes.

1550-6037/14 $31.00 © 2014 IEEE
DOI 10.1109/IV.2014.13

It is generally difficult to estimate the contents of the playlists and features of the bundled tunes, just from textual information of the playlists. In this case it is useful to quickly and intuitively understand the contents and features of the playlists by using visual representation. Another situation is while non-owners of music players are operating them. For example, fellow passengers of vehicles may operate the music players in the vehicles, even though the passengers do not know what kinds of tunes are recorded. In this case, it is often difficult for the passengers to understand the contents and features of the playlists just from the names of artists or albums. Again, visual playlist representation should be useful to quickly and intuitively understand the contents and features under such situations.

II. Related Work

There have been many studies on visualization of musical contents and features. MusicIcons [3], MusCat [4], and MusicThumbnailer [10] are typical techniques which represents a tune as an image. MusicIcons generates tunes as glyphs from their acoustic features. MusCat represents musical features as abstract pictures, while MusicThumbnailer arranges images so that users can estimate the genres of tunes. There are many other studies on tune visualization which converts musical features into visual properties. Meanwhile, other musical elements such as lyrics or genres are visualized in several works. For example, Lyricon [5] visualizes pop songs by assigning multiple icons based on the story of lyrics. However, there have been few studies on playlist-based music visualization [1], even though many of music player users select tunes playlist-by-playlist.

Meanwhile, several music visualization techniques represent distribution of tunes or artists, rather than representation of individual tunes. Islands of Music [7] places a group of tunes onto a 2D display space based on their musical similarity. MusicRainbow [9] radially places artists based on the similarity of their own tunes. There are several similar works on visualization of large music collections [6] [8]. These techniques can assist users to understand the distribution of stored tunes. However, they do not visually represent detailed musical contents or features of independent tunes, but just display metadata of manually pointed tunes.

III. Processing Flow

This section describes the processing flow and user interface design of GRAPE. It consists of three preprocessing steps for the gradation image generation. It firstly calculates the musical feature vectors of the given tunes in a playlist. The vectors are applied to calculate colors of the tunes, amd consumed by Self Organizing Map (SOM) to calculate their positions. Consequently, the tiles corresponding to the tunes are colored and placed to form a gradation image.

Table II
QUESTIONNAIRE: MUSICAL ELEMENTS BRINGING ASSOCIATION OF COLORS WHILE THE ANSWERERS LISTEN TO THE MUSIC.

Musical elements	Number of agreed answerers (%)
Harmony and tonality	84
Vocal sound	78
Tempo and rhythm	75
Story of lyrics	71
Instruments	62
Genre	60
Fashion	25
Loudness	24

A. Musical Feature Extraction

GRAPE supposes that musical feature values of each tune are calculated as a preprocessing. Our implementation applies MIRtoolbox [1] to calculate the following three musical feature values: ”Tempo”, ”RMS energy” which is a root mean square of acoustic energy, and ”Brightness” which is a percentage of high frequency elements. Here, our implementation uses the normalized feature values. We specified maximum and minimum values of the musical features from the tunes introduced by RWC Music Database [2]. This database contains various genres of tunes, and therefore we suppose it is appropriate to use for the specification of minimum and maximum feature values.

We selected the above three musical features based on the questionnaire result shown in Table II. We asked the answerers which musical elements bring association of colors while they listen to the music, and considered the elements which more than 50% of answerers agreed in the result.

“Harmony and tonality” was the most important in the result, and actually, MIRtoolbox can calculate the possibility of tonality and ratio of major and minor chords. We did not apply these musical features, because we did not agree that musical feature values related to harmony and tonality well represent the impression of the tunes in our preliminary experiments. “Vocal sound” and “Story of lyrics” were the second and fourth important elements in the result. We did not apply the musical features related to the vocal sound and story of lyrics, because many of tunes we used in our experiments did not contain vocal parts.

“Tempo and rhythm” was the third important element, and actually we applied the musical feature “Tempo” calculated by MIRtoolbox. “Instruments” and “Genre” were moderately important in the results, and we concluded that the musical features “RMSEnergy” and “Brightness” were quite related to these elements. “RMSEnergy” is the root mean square of the loudness, which tends to larger at the pop, rock, or electric tunes, because they are compressed to obtain the flat loudness, while “RMSEnergy” of tunes played by acoustic instruments is often smaller. “Brightness” is the

[1] http://www.jyu.fihum/laitokset/musiikki/en/research/coe/materials/mirtoolbox
[2] http://staff.aist.go.jp/m.goto/RWC-MDB/

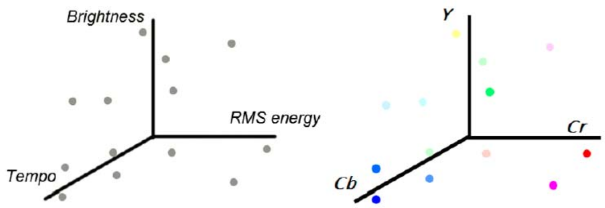

Figure 2. Mapping three features into YCbCr color space.

ratio of high tone (higher than 1,500Hz) mainly contained as harmonic overtones of the instruments. Sound of particular instruments contain rich overtones, and therefore the value of "Brightness" may bring estimation of arrangements.

B. Coloring on YCbCr Color System

The next step assigns colors to the tunes from their musical feature values. GRAPE directly converts the three feature values to the three components of the YCbCr color system as shown in Figure 2.

The color assignment implemented for GRAPE is based on the psychology regarding the colors [11]. Our implementation assigns "Brightness" to the Y-axis which denotes the intensity, because bright color may associate bright sounds. It assigns "RMS energy" to the Cr-axis, because the red is psychologically suggestive to activeness and energy. It assigns "Tempo" to the Cb-axis because the blue is psychologically suggestive to speed and briskness. Consequently, colors of loud tunes are close to red, and colors of speedy tunes are close to blue. On the other hand, colors of quiet and slow tunes are close to green, because both Cb and Cr values are small. This is also intuitive because the green is psychologically suggestive to gentleness and calmness.

We tested various calculation schemes to translate musical feature values to colors, applying various color systems such as the RGB and HSB color systems as well as the YCbCr color system. We experimentally and subjectively selected the YCbCr color system, because tunes are well distributed in the color space while applying the YCbCr color system.

C. Layout by Self Organizing Map (SOM)

Musical feature vectors are also applied to calculate the positions of the tunes on the display space. GRAPE places tunes which have similar feature vectors closer, to generate a gradation image from a playlist. Human tends to psychologically want to touch objects painted by gradation colors [11]. Such psychological tendency is effective to develop touchable user interfaces of visually playlists.

GRAPE generates a rectangular gradation image by arranging tunes in a playlist as square tiles based on the result of SOM. We selected this design because we would like to evenly display the tunes as equally-sized squares, while SOM has a good property to evenly place the data items.

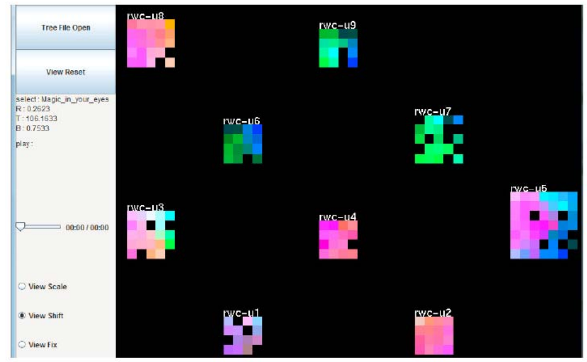

Figure 3. User interface as a PC application.

This design also has a good property that square is not easy to collapse while drawing in the small displays. We suppose to use GRAPE on the small devices such as cellphones or mobile players, and therefore this property is very important.

D. Display and User Interface

We implemented GRAPE as applications on the personal computers and Android devices. Our implementation displays the set of playlists as the set of gradation images, as shown in Figures 1 and 3.

1) Application on the personal computers: Figure 3 shows the user interface we implemented for personal computers with Java Development Kit (JDK) 1.6. This application features the following user interfaces:

- Shift and scaling of images by mouse drag.
- Display of tune titles by cursor pointing.
- Start and stop of playing tunes by mouse click.

They enable to firstly overview the many playlists, show the details of interested playlists on demand, and finally play the tunes of preferred gradation images.

2) Application on the Android devices: Figure 1 shows the user interface we implemented on an Android device with JDK 1.6, Android Software Development Kit (SDK) 2.3.3, and Android Development Tools (ADT) Plugin. This implementation also features shift and scaling of images, display of tune titles on demand, and start/stop operation.

IV. Example

Figure 4 shows the examples of gradation image generation for three playlist A, B, and C. The playlists contain 36, 28, and 29 tunes respectively. Here, white squares denote the blank regions which tunes are not assigned.

These result well represents the features and impressions of the playlist. Image A has respectively larger number of bright red or pink squares, while the corresponding playlist contains larger number of bright energetic pop/rock tunes. Image C has respectively larger number of dark green or blues squares, while the corresponding playlist contains larger number of non-electric and simple tunes. Image B has

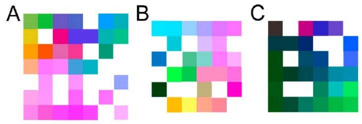

Figure 4. Result with three playlists.

relatively variety of colors, while the corresponding playlist contains more variety of tunes. These results denote that features and characteristics of playlists are well represented by the gradation images generated by GRAPE.

V. Evaluation

This section shows three user evaluations with gradation images generated by GRAPE.

A. Evaluation (1): consistency between visual and acoustic impressions

We showed three gradation images shown in Figure 4, and 16 keywords including 8 adjectives and 8 terms related to music genres, to 14 subjects. We then asked them to select keywords associated from each of the three images. Table III shows the selection by the subjects. The result denotes that many subjects associated particular common keywords from each of the gradation images. It suggests that many subjects had common impressions for the gradation images.

Most of subjects selected particular keywords such as “Bright”, “Bustling”, and “Pops” for the playlist A. Actually the playlist A was the collection of enjoyable pop songs. We supposed that the impression was brought from larger Y and Cr values of the tiles, since many tunes in the playlist had larger “Brightness” and “RMSEnergy” values.

More than half of subjects selected “Classic” and “Ballad” for the playlist C. We supposed that they appropriately imagined the contents of the playlist C, because actually the playlist C was the collection of quiet type of classical music. Also, most of them selected “Dark” and “Slow” for the playlist C. We supposed that the dark impression was brought from smaller Y values, and it is appropriate because many tunes actually had smaller Brightness values. On the other hand, the impression “Slow” was not appropriate. We suppose one reason is that Tempo is somewhat difficult to calculate for some kinds of classical music, because their tempo is not constant, and many of classical music do not have beat instruments such as snare drums or bass drums.

Selection of keywords for the playlist B was relatively split. We suppose it is a good result, because the playlist B contained variety of genres of tunes.

B. Evaluation (2): comparison with non-feature-based visual representation

Next, we had another evaluation with another set of playlists 1, 2, 3, and 4. We generated the following three types of images for each of the playlists:

Table III
Result of the evaluation (1).

Word	A(%)	B(%)	C(%)
Bright	86	57	0
Dark	0	0	100
Fast	14	29	0
Slow	0	14	93
Bustling	79	14	0
Quiet	0	21	21
Glorious	14	0	64
Delicate	7	43	7
Rock	36	14	0
Pops	79	36	0
Jazz	0	21	36
Classic	0	21	57
Ballad	7	21	64
R&B	7	0	7
No-vocal	0	43	14
Techno	7	21	0

(A) gradation images generated by GRAPE,
(B) images generated by arranging icons of genres, and
(C) images generated by arranging CD jacket images,
as shown in Figure 5(Left). We showed the images and asked subjects to select the most preferable image for each of the playlists. We gathered answers from 138 subjects. Figure 5(Right) shows the statistics. We also asked the subjects to write any comments or suggestions regarding the selection of images. Following are the typical comments:

- Comment (a): I listen to the music only album-to-album. CD jacket image is sufficient for me.
- Comment (b): I never listen to the tunes which I do not know the contents. Situations which GRAPE supposes are out of my daily life.
- Comment (c): I would like to select images while listening to the tunes for a short time. Authors should conduct the evaluation again after preparing sound files.
- Comment (d): Gradation image should be useful while sharing the tunes with friends or family.
- Comment (e): Gradation image should be useful when we would like to listen to something, but we do not have any particular tunes to listen to.
- Comment (f): Gradation image should be useful while music creation and mastering.

We suppose GRAPE will not be effective for subjects who gave the comments (a) and (b), while other subjects would be interested in the concept and goal of GRAPE. Also, the comment (c) suggests that we need to have additional user experiments preparing tunes. On the other hand, the comments (d), (e), and (f) suggest GRAPE will be interested by various music listeners who have various situations in addition to the situations mentioned in Section 1.

C. Evaluation (3): satisfaction of playlist selection results

Finally, we conducted the third user evaluation to determine if the gradation image is useful as a user interface for playlist selection. We showed the user interface of

Figure 5. Pictures provided for the user evaluation (2).

Table IV
RESULT OF THE EVALUATION (3).

	GRAPE		Random	
Subject	Playlist	Preference	Playlist	Preference
A	7	4	8	2
B	2	3	5	2
C	5	3	8	2
D	7	2	1	1
E	6	4	7	3
F	5	3	1	2
G	2	4	9	3
H	7	4	3	2
I	2	3	6	2

GRAPE running on the PC to the subjects. We showed the nine playlists displayed as gradation images in Figure 3, and asked the subjects to select the preferred or interested gradation image. We also asked to listen to the tunes in the playlists corresponding to the selected gradation image, and evaluate how the tunes were close to the music they wanted to listen to at that time. We also asked the subjects to listen to the tunes in the randomly selected playlist, and similarly evaluate how close to the music they wanted to listen to. The evaluation was 4-level, where 4 was the best, and 1 was the worst. The playlist was created from the tunes bundled by RWC Music Database. All the subjects have never listened to the tunes, and therefore they needed to select playlists only from the impression of gradation images.

Table IV shows the evaluation of the 9 subjects. All the subjects evaluated that the playlist selected by looking at gradation images were closer to the music they wanted to listen to. The result suggests that gradation images generated by GRAPE provide meaningful impression for the playlist selection on the music player software.

VI. Conclusion and Future Work

This paper presented GRAPE, a visualization technique representing features of playlists as gradation images. The paper also introduced results and experiments demonstrating the effectiveness of GRAPE.

As a future work, we would like to apply more variety of musical features to GRAPE and have experiments to clarify what kinds of features are more important. Also, we would like to implement more features and variation of visualization and user interfaces of GRAPE. For example, gradation images of our current implementation do not represent the order of tunes in a playlist, and therefore users may prefer to sometimes use an additional mode which visually represents the order of tunes. Or, we heard that several subjects preferred to fill the blank part of the gradation images by the colors of adjacent tunes. Such kinds of variation of representation in the gradation images may make users more satisfactory.

Finally, we would like to develop open APIs to easily generate the gradation images and share among friends or families. We expect such kinds of environments may make more enjoyable to share tunes among them.

References

[1] J. Donaldson, P. Lamere, Using visualizations for music discovery, International Conference on Music Information Retrieval (ISMIR), 2009.

[2] T. Kohonen, The self-organizing map, Proceedings of the IEEE, 78, 1464-1480, 1990.

[3] P. Kolhoff, J. Preub, J. Loviscach: Music icons: procedural glyphs for audio files, Brazilian Symposium on Computer Graphics and Image Processing, 289-296, 2006.

[4] K. Kusama, T. Itoh, Muscat: a music browser featuring abstract pictures and zooming user interface, ACM Symposium on Applied Computing (SAC'11), 1227-1233, 2011.

[5] W. Machida, T. Itoh, Lyricon: A Visual Music Selection Interface Featuring Multiple Icons, 15th International Conference on Information Visualisation (IV2011), 145-150, 2011.

[6] F. Morchen, A. Ultsch, M, Nocker, C. Stamm, Databionic Visualization of Music Collections According to Perceptual Distance, International Conference on Music Information Retrieval (ISMIR), 396-403, 2005.

[7] E. Pampalk, Islands of music: Analysis, organization, and visualization of music archives, Master's thesis, Vienna University of Technology, 2001.

[8] E. Pampalk, A. Rauber, D. Merkl, Content-based organization and visualization of music archives, ACM International Conference on Multimedia, 570-579, 2002.

[9] E. Pampalk, M. Goto, MusicRainbow: A new user interface to discover artists using audio-based similarity and web-based labeling, International Conference on Music Information Retrieval (ISMIR), 367-370, 2006.

[10] K. Yoshii, M. Goto, Visualizing musical pieces in thumbnail images based on acoustic features, International Conference on Music Information Retrieval (ISMIR), 211-216, 2007.

[11] S. Yoshihara, Textbook of Colors, Yosensha, ISBN-13-978-4862487483, 2011.