Generating Affective Music Icons in the Emotion Plane

Hyun-Ju Kim
Yonsei University
Shinchon Dong 134,
Seodaemungu, Seoul
uyunamu@cs.yonsei.ac.kr

Ji-Yong Kwon
Yonsei University
Shinchon Dong 134,
Seodaemungu, Seoul
mage@cs.yonsei.ac.kr

Min-Joon Yoo
Yonsei University
Shinchon Dong 134,
Seodaemungu, Seoul
debussy@cs.yonsei.ac.kr

In-Kwon Lee
Yonsei University
Shinchon Dong 134,
Seodaemungu, Seoul
iklee@cs.yonsei.ac.kr

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.
ACM 978-1-60558-247-4/09/04.

Abstract
In this paper, we discuss the generation of icons that represent the emotion expressed in music. We use the emotion plane for connecting the music with the icon shape affectively. A model to project arbitrary music on the plane is introduced using the result of a user survey and various features of audio signals. Icon shapes are located on the plane from the result of user survey. The icon shape of the input music is obtained by blending neighbor icon shapes of the point of the music on the emotion plane. Using this method, one can easily guess the emotion of music from the corresponding icon shape and find the music he or she wants.

Keywords
emotion plane, arousal-valence, music icons, affective icon

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
For a long time, music has been with human history. People have expressed their feelings through music. Recognition of emotions expressed in music is important to understand the human perception of

music. Recently, there have been several attempts to do this automatically [15,7,10]. Meanwhile, the amount of music data has increased as it has long been with human history and some people even own hundreds or thousands of pieces of music. Thus, it becomes a serious problem to manage music efficiently. It is also difficult to select music suitable for a person's current emotions because recalling the ambience of all the music collected by the person is very hard.

In this paper, we suggest a novel music representation method based on affection implicated in music. We generate icons representing the emotion expressed in the music so that people can easily guess the emotion of music just by seeing the icons of music. Thus, users can select the music which is suited for their mood more easily.

Recently, Lewis [9] suggested fast navigation method of files in his VisualID study. He generated icons which are similar with slight mutations based on the different filenames. Icons in our system have different shapes by the different emotions of music instead of the text data such as filenames.

In a study of music created icons, Kolhoff [6] generated flower-shaped icons from MFCC data of music signals. In this study, although the icons are generated from the music signal, the icons are only specialized in personal choice and don't express the emotion of the music. We generate the icons representing the general human emotion expressed in the music.

Icons as Human-Computer Interaction

It is general fact that showing real shape is much more effective to recognize an object than explaining it in long words. Pictures have been used to deliver information through time and cultures. They are effective not only to deliver information but to understand it. Icon is a way to apply the features of pictures on a machine. The icons in human computer interaction have the special features [11]: Guessability, Learnability and Experienced User Performace.

With the boom of a home computer marketplace, the needs of practical and functional icon design technique have increased. By these needs, there has been a great deal of research about the artistic side of icon design [1,5]. While the previous studies focused only on the artistic side of icon design, Chen[2] suggested scientific-oriented methodology for graphical icon designs.

Emotion Plane

Recently, there have been various studies on the automatic extraction of emotions from multimedia. Especially, arousal-valence model [14] was used in the works related to music [15,7,10] and the works related to images[3,4]. This model represents human emotions on a two dimensional emotion plane having an arousal and valence axis. The arousal axis represents the exciting rate of human emotion and the valence axis represents the positive rate of human affection. We use this emotion model in our research. We represent the emotion of music and icons on the emotion plane and match them with each other on the plane.

System Overview

Our system generates icons that reflects the emotion of music for searching and finding music more efficiently. The emotions of music and the icons are represented in the emotion plane (see figure 1).

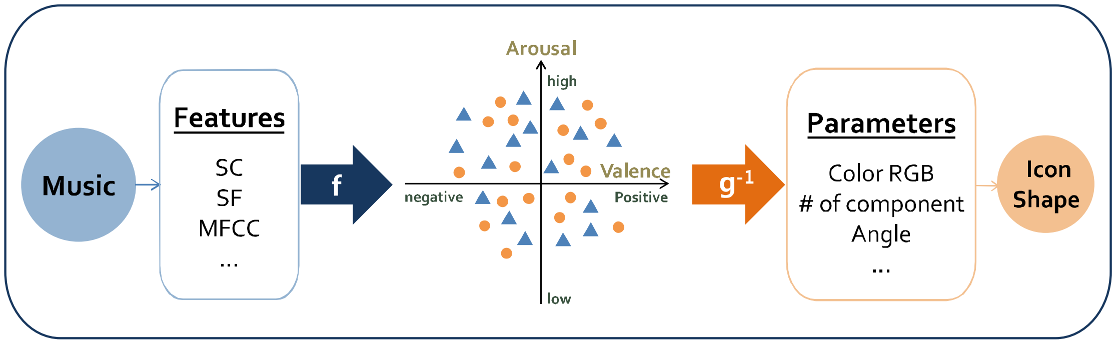

figure 1. The generation procedure of the icon shapes from arbitrary music. After the features of the input music are extracted, the value in the emotion plane (AV value) is determined through the model representing the relation between the features of music and their emotion. Triangle points are AV value of the example music and circle points are AV value of the example icon shapes. The result icon shape of the input music is generated by the shape parameters obtained from the AV value of the input music. To figure out the multidimensional icon shape parameters from the two dimensional AV plane, we blend the shape parameters of the shapes projected to the two dimensional space.

First, we conduct a user survey about emotion, and define the model representing the relation between the music features and the values on the emotional plane (AV values) in order to locate the music in the emotion plane. The emotions expressed in various icon shapes are also determined from the user survey. The icon shapes are located in the emotion plane through the user survey data. We blended the parameters of several icon shapes on a two dimensional plane for the purpose of finding out the multidimensional icon shape parameters that represent the music emotion projected on a two-dimensional plane.

Experiment Design

First of all, we determined the AV values of some training music samples from the user survey to make the model that extracts the emotion expressed in music. The songs in the training music set are collected from popular songs in Korean, Western and Japanese music. Two criteria are applied to the music selection [15]: First, the training music set should be distributed uniformly in the emotion plane. Second, each training music sample should express a certain dominant

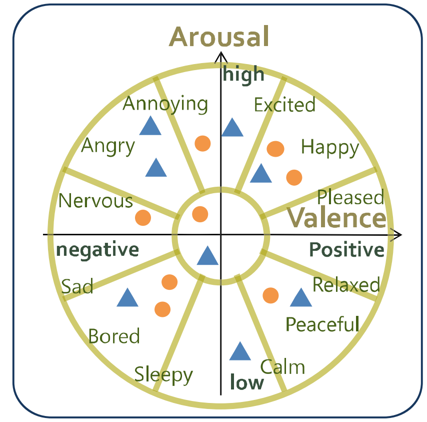

figure 2. We divide the emotion plane into nine regions.

emotion. That is, the emotion expressed in the music should be similar to each person. We divide the emotion plane into the nine regions to satisfy the first criterion (see figure2). Then we make up the set of 135 training music samples which may be included in each region uniformly. (The initial distribution is executed by the authors.)

We conducted a survey about the emotion expressed in the music using a modified Lang [8]'s user survey method. Before the real survey, the users listen to nine exercise music samples, each of which are selected from each of the nine regions in figure 2, to increase to raise the precision of the user survey. The real training music set made up of 135 songs: 15 songs in each region. Users listen to the music and score the emotion of the music: the arousal and the valence value. The range of the arousal-valence value is 1~9. The number of participants sampled was 14.

According to the second criterion for making up the correct training music set, we selected music expressing a certain dominant emotion. We excluded music samples which had high variant scores than a threshold from the training music set.

We assume that human emotion expressed in the music has some variance. We include nine pairs of the same music in the training music set and make the users score the same music twice. Then we use this variance obtained from the scores as a threshold. As a result, the number of training music samples decreased from 135 to 86 after excluding the outliers.

Feature Extraction

Next, we extract various features from the audio signal to generate the model that may convey the emotion expressed in the music. We extract 55 feature vectors (FFT Spectrum, Hilbert, Cepstrum, Auto-Correlation, and Loudness et al.) of each song in the training music set using Psysound [12]. Then, we select the more relevant features for improving the precision of the model.

We compute the correlation between the users' AV values mentioned above and the 55 features of each training music set, and we exclude the features which have the low correlation than a threshold. Threshold is determined by the authors after observing the correlation. As a result, the number of the audio features decreased from 55 to 27.

Mapping Music into Emotion Plane

We make the linear model in order to find out the relation between the AV values obtained from the user survey and the audio features:

$$S = F \cdot M,$$

Where S is a matrix (86×2) of the AV values from the user survey result from the 86 training music samples. M is a matrix (86×27) of the 27 audio features extracted from the training music samples. We can find out the linear model F through the pseudo inverse:

$$F = S \cdot M^{\dagger},$$

Using this model F and the audio features of arbitrary music, we can project any arbitrary music on the emotion plane.

Analysis

We ran two experiments to evaluate the precision of the model. First, we use 'leave one out' method; We make up a model F with the temporary training music set in which one music sample is excluded from the whole training music set. We input the music sample excluded before to the model, and we obtain the estimated score. We repeat this procedure one by one to other music samples in the whole training music sample set. After 'leave one out' to all of the samples, we compare the scores of the user survey (ground truth) with the estimated scores. The expectation value of the difference between the ground truth and the estimated score E(|ground truth – estimated scores|) of

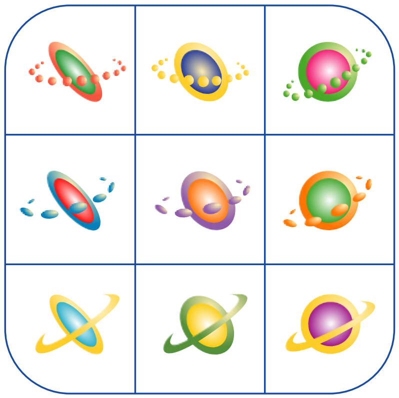

figure 4. An example of icons created by using some rules in figure 3.

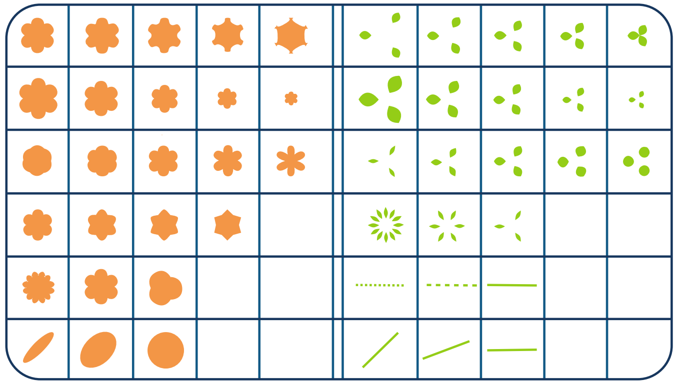

figure 3. Some examples of icon shapes made by parameters. Shapes are changed from various parameters such as scale, orientation and length. We conducted a survey about user affection of the shape changed by these parameters. Now, we are making more shape parameters.

the arousal value was 0.8161, of the valence value was 1.4857. The worst-case value of the difference between the ground truth and the estimated score is 8, because the rating value in the user survey is 1 to 9. From this analysis, we can see the linear model in this paper is considerably reliable.

As the second test, we compare the ground truth from the user survey with the estimated AV values computed using linear model about the whole training music set. Expectation value of the differences between the ground truth and the estimated scores is also used to calculate the error. As for the test result, an error of the arousal and valence values were 0.5093 and 0.9518, respectively. We can say it is a highly reliable model, because the possible range of the difference is also [0, 8].

Future Work

We introduced the procedure that arbitrary music can be projected on the emotion plane. Now, our study is on the making icon shapes and projecting the icons on emotion plane.

We made some shapes which can be used to represent human emotion, and defined the shape parameters. Figure 3 shows the simple shape rules. Figure 4 shows some icon examples. We can find out the AV values of the icon shapes from the user survey of the example icon shapes. We plan to make more rules about the shape parameters. We refer to the emotion attribution of Pavlova [13] for making the shapes expressing the human emotion.

The generation procedure of icon shapes we are planning is as following: If we obtain an AV value from an arbitrary input music, we can find some of the example icon shapes laid around the AV value of the input music in the emotion plane. We blend the parameters of the icon shapes, and obtain the new shape parameters of the input music (see Figure 5). Through the blended parameters, we can generate a new icon shape expresseding the emotion of the input music.

Acknowledgements

This research is the result of the promotion project for culture contents technology research center supported by Korea Culture & Content Agency (KOCCA).

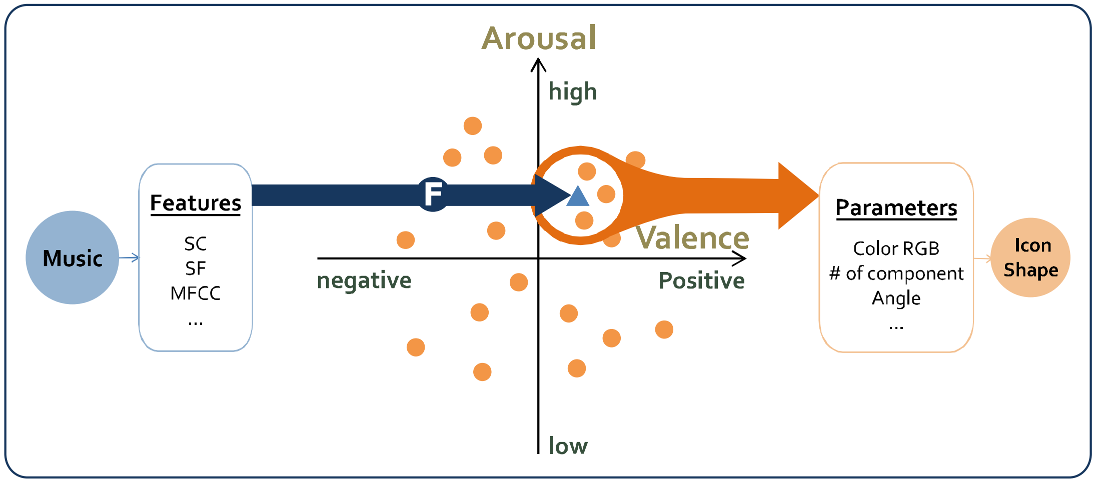

figure 5. The generation procedure of icon shapes from arbitrary music. We get the AV values from the features of input arbitrary music. In the figure, the triangle point is the AV value of an input music, and the circle points are the AV values of the example icon shapes. There are some shape points nearby one music points. We blend the parameters of some nearest neighbor icon shapes and then generate a new icon shape.

References

[1] Calpin, S., Icon Design: Graphic Icons in Computer. Interface Design, Watson-Guptill Publications (2001).

[2] Chen, P.P., Toward a Methodology of Graphical Icon Design. Proceedings of the IEEE Fifth International Symposium on Multimedia Software Engineering (2003).

[3] Hanjalic, A., and Xu, L-Q. Affective Video Content Representation and Modeling. IEEE Transactions on Multimedia, 7, 1 (2005), 143-154.

[4] Hanjalic, A. Extracting Moods from Pictures and Sounds. IEEE Signal Processing Magazine, March, (2006), 90-100.

[5] Horton, S., The Icon Book: Visual Symbols for Computer Systems and Documentation. John Wiley & Sons (1994).

[6] Kolhoff, P., Preuss, J., Loviscach, J. Music Icons: Procedural Glyphs for Audio Files. Computer Graphics and Image Processing, SIBGRAPI (2006), 289 – 296.

[7] Korhonen, M.D. Clausi, D.A. Jernigan, M.E. Modeling Emotional Content of Music Using System Identification. IEEE Transactions on Systems, Man, and Cybernetics_Part B: Cybernetics (2005), 588-599.

[8] Lang P.J., Bradley M.M., Cuthbert B.N. International Affective Picture System (IAPS): Instruction Manual and Affective Ratings. NIMH Center for the Study of Emotion and Attention (2007).

[9] Lewis, J.P., Rosenholtz, R., Fong, N., Neumann, U. VisualIDs: automatic distinctive icons for desktop interfaces. ACM Trans. Graph. (2004), 416-423.

[10] Lu, L., Liu, D., and Zhang, H-J. Automatic Mood Detection and Tracking of Music Audio Signals. IEEE Transactions on Audio, Speech, and Language Processing 14, 1, (2006), 5-18.

[11] Moyes, J., Jordan, P.W. Icon design and its effect on guessability, learnability, and experienced user performance. People and Computers VIII. Cambridge University Press, Cambridge (1993), 49-59.

[12] Psysound. http://psysound.wikidot.com.

[13] Pavlova M. Perceived dynamics of static images enables emotional attribution. Perception (2005), 1107-1116.

[14] Thayer, R.E. The Biopsychology of Mood and Arousal. New York: Oxford Univ. Press (1989).

[15] Yang, Y.H., Lin, Y.C., Su, Y.F., and Chen, H.H. A Regression Approach to Music Emotion Recognition. IEEE Trans. on Audio, Speech, and Language Processing (2008), 448-457.