Affecticon: Emotion-Based Icons for Music Retrieval

Min-Joon Yoo and In-Kwon Lee
Yonsei University

Digital audio is easy to record, play, process, and manage. Its ubiquity means that devices for handling it are cheap, letting more people record and play music and speech. In addition, the Internet has improved access to recorded audio. So, the amount of recorded music that people own has rapidly increased.

Most current audio players compress audio files and store them in internal memory. Because storage costs have consistently declined, the amount of music that can be stored has rapidly increased. A player with 16 Gbytes of memory can hold approximately 3,200 songs if each song is stored in compressed format and occupies 5 Mbytes. Effectively organizing such large volumes of music is difficult. People often listen repeatedly to a small number of favorite songs, while others remain unjustifiably neglected.

We've developed Affecticon, an efficient system for managing music collections (see Figure 1). Affecticon groups pieces of music that convey similar emotions and labels each group with a corresponding icon. These icons let listeners easily select music according to its emotional content. Experiments have demonstrated Affecticon's effectiveness.

Music Emotion Recognition

We represent the emotion in music, and subsequently construct our icons, on the basis of Robert Thayer's 2D arousal and valence (AV) plane (for more on this, see the sidebar). We first evaluated music samples' emotion values through listening tests. Then, we modeled the relationship between these values and the samples' audio features. This model lets us determine the emotional content of an arbitrary piece of recorded music.

Listening Tests

We used music samples from Western (US) and Eastern (Korean and Japanese) popular songs. We chose them according to two criteria:

- The set of samples should be distributed uniformly across the AV plane.
- Each sample should express a certain dominant emotion.

To meet the first criterion, we divided the AV plane into nine regions and assembled 16 samples for each region, using our subjective judgment.

We then conducted the tests using a modified version of the Self-Assessment Manikin, a rating system that uses graphic figures to indicate emotional reactions.[1] Initially, we asked 14 participants to listen to nine songs, one from each region, to give them some idea of the tests' scope. Then, the participants listened to the remaining 135 songs, scoring them on a scale of 0 to 1 for arousal and

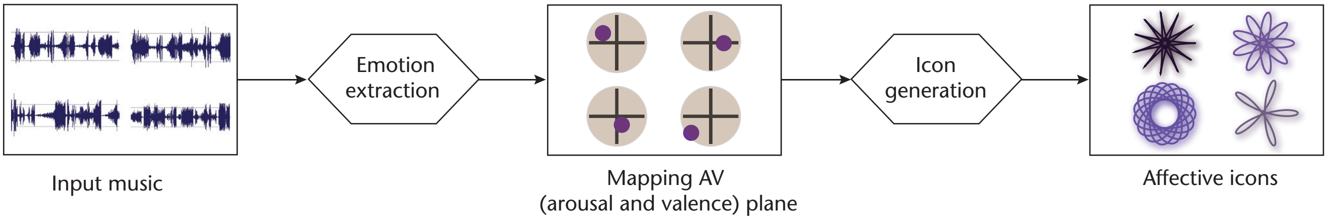

Figure 1. The Affecticon system. Affecticon groups pieces of music that convey similar emotions and labels each group with a corresponding icon. These icons let listeners select music according to its emotional content.

Analyzing Emotion in Music

Many researchers have tried to automatically extract music's emotional content, employing a range of feature extraction and regression methods.[1–3] Most methods have adopted Robert Thayer's 2D emotion model,[4] which represents emotion along two axes: arousal (from calm to energetic) and valence (from happy to anxious). Because of this model's simplicity, researchers have used it extensively in areas such as music information retrieval and computer graphics.[5,6] Because no common or standard methodology exists yet for dealing with the emotion that music evokes, we designed and tested our own emotion recognition model, also based on Thayer's model (see the main article).

References

1. L. Lu, D. Liu, and H.J. Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 1, 2006, pp. 5–18.
2. Y.H. Yang et al., "A Regression Approach to Music Emotion Recognition," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 2, 2008, pp. 448–457.
3. Y.E. Kim et al., "Music Emotion Recognition: A State of the Art Review," *Proc. 2010 Int'l Soc. for Music Information Retrieval Conf.* (Ismir 10), Int'l Soc. for Music Information Retrieval, 2010, pp. 255–266; http://ismir2010.ismir.net/proceedings/ismir2010-45.pdf.
4. R.E. Thayer, *The Biopsychology of Mood and Arousal*, Oxford Univ. Press, 1989.
5. A. Hanjalic, "Extracting Moods from Pictures and Sounds," *IEEE Signal Processing*, vol. 23, no. 2, 2006, pp. 90–100.
6. H.L. Wang and L.-F. Cheong, "Affective Understanding in Film," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 6, 2006, pp. 689–704.

valence. For arousal, 0 indicated the calmest reaction and 1 indicated the most exciting reaction. For valence, 0 indicated the most negative reaction and 1 indicated the most positive reaction.

To allow for the variability of emotional responses, we repeated nine songs so that the participants scored them twice. We used this variability as a threshold. We ignored samples with a higher variability than this, to favor songs with a consistent emotional interpretation. This reduced the number of songs to 86.

Feature Extraction

For each song, we extracted 55 features from the audio signal that correspond to emotion. To do this, we used Psysound (http://psysound.wikidot.com), software for analyzing audio features such as the fast Fourier transform spectrum, cepstrum, loudness, and pitch. Then, we computed the correlation between the participants' AV values and the 55 features. We excluded features with a correlation below the threshold, reducing the number of them to 27.

Mapping Music into the AV Plane

We created a linear model F to relate the AV values obtained from the listening tests to the audio features. We set

$$S = MF,$$

where S is an 86×2 matrix of the music samples' AV values, and M is an 86×27 matrix of the extracted features. We can evaluate $F(27 \times 2)$ by pseudoinversion:

$$F = M^{+}S.$$

Having found F, we can use it to analyze any piece of music's emotional content. That is, we can analyze a song by determining the 27 feature measurements using Psysound and multiplying by F to calculate AV values.

Analysis

We ran two experiments to evaluate this model's precision. First, we used a *leave-one-out* method, constructing a new model F' from a set of samples from which we had excluded one song. Then we obtained AV values for the excluded song from the model. We repeated this procedure for all the samples and compared the results with those from the listening test. The average difference between arousal values was 0.063 and between valence values was 0.034.

The second experiment compared the participants' ratings with AV values estimated using F across the whole set of songs. The average error was 0.064 for arousal and 0.119 for valence. These results suggest that the model is self-consistent.

Icon Generation

To generate the icons, we simulate a harmonograph, a mechanical apparatus that employs pendulums to create a range of symmetrical curves. The calculated AV values determine the curves' shape and color. We tested the curves' effectiveness in a series of user trials.

Harmonograph Curves

We can produce a range of similar but simplified curves using these equations:

$$x(t) = A_1 \sin(tf_1 + \pi/2) + A_2 \sin(tf_2 + \pi/2),$$

$$y(t) = A_1 \sin(tf_1) + A_2 \sin(tf_2).$$

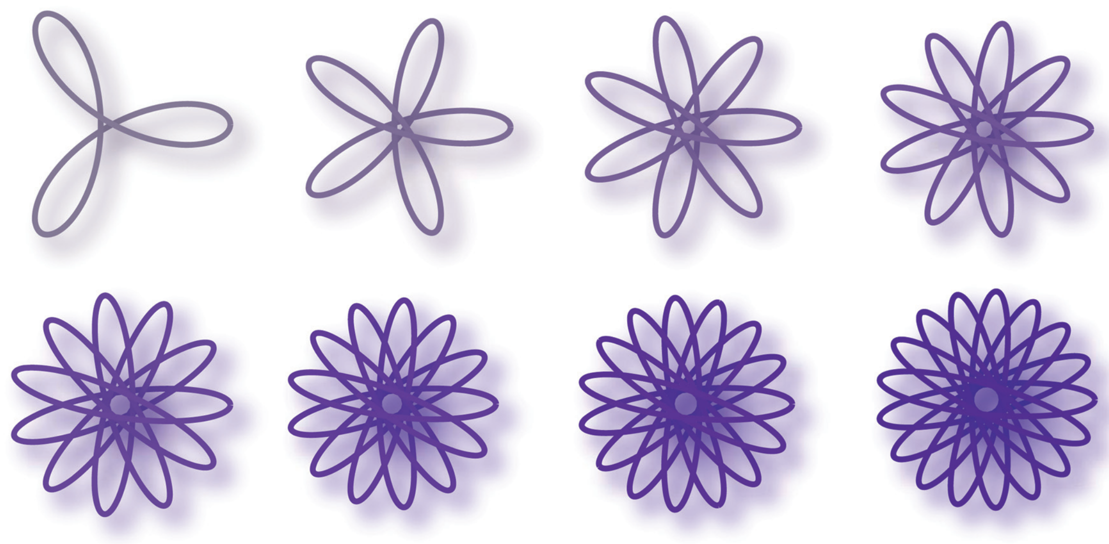

Figure 2. Images generated by successively reducing the ratio between frequencies f_1 and f_2 in the harmonograph equations. We used the number of ellipses to indicate arousal.

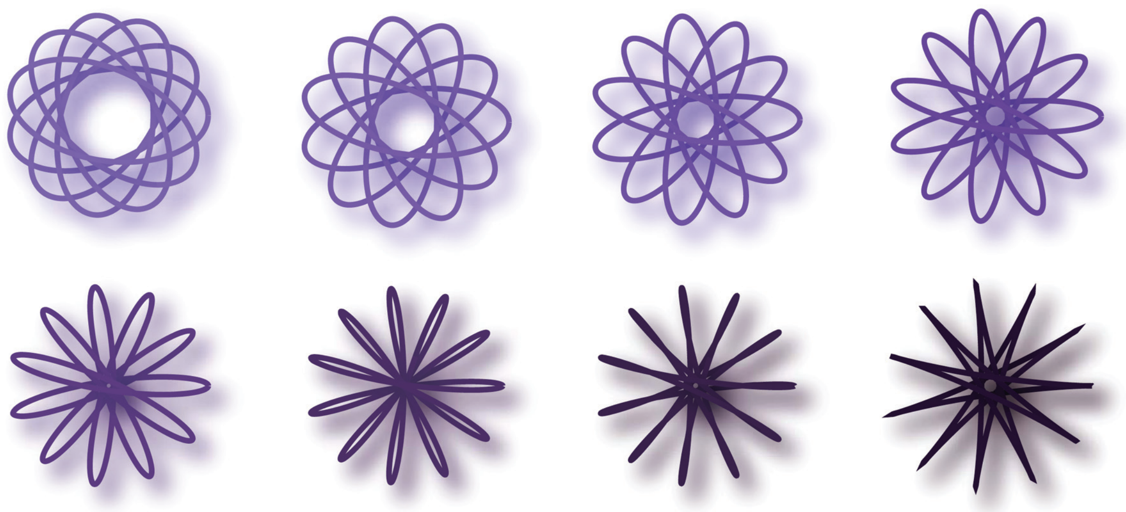

Figure 3. Images generated by successively reducing the ratio between amplitudes A_1 and A_2 in the harmonograph equations. We used the different degrees of roundness to indicate valence.

The values of the amplitudes A_1 and A_2 and the frequencies f_1 and f_2 control the results. This process produces symmetrical closed curves consisting of many ellipses, with almost circular convex hulls.

Mapping Arousal

If we adjust the ratio between f_1 and f_2, the number of ellipses in a curve changes (see Figure 2). Researchers have correlated high arousal with high pitch, loud volume, rapid movement, fast tempo, and fast speech.[2] So, we used the number of ellipses to indicate arousal.

Mapping Valence

The ratio between A_1 and A_2 changes the ellipse's shape and hence its area and the sharpness of its ends (see Figure 3).

Geoffery Collier's research on emotion-shape associations suggests that triangles evoke anger, pentagons and squares evoke embarrassment, circles evoke cheerfulness, and ellipses are soothing.[3] Jocelyn Scheirer and Rosalind Picard suggested that circles evoke high-valence emotions, whereas diagonal lines produce low-valence emotions.[2] They also posited that people view "smooth" as more positive than "rough."

We deduce that sharpness affects valence: thinner or sharper ellipses should make someone feel more negative. So, we used different degrees of roundness to indicate valence.

An Arousal Experiment

We showed 12 participants 10 pairs of images; each image in a pair had a different arousal value

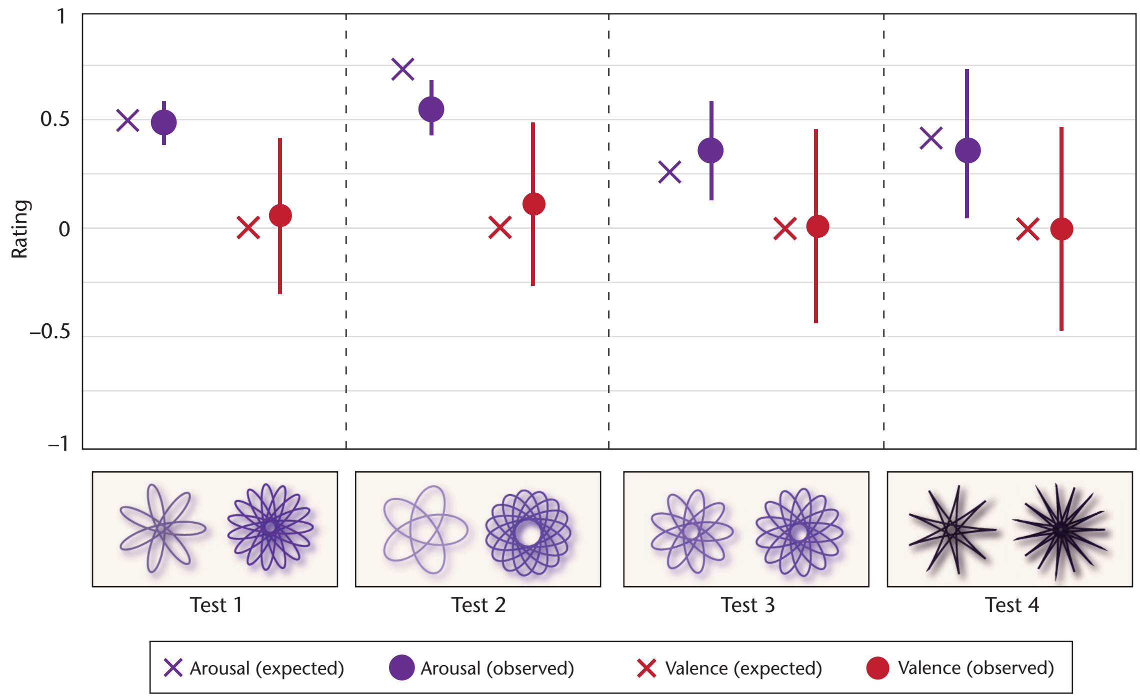

Figure 4. Results for four pairs of images from the arousal experiment. Most participants found many more differences in arousal than in valence. A rating of 1 indicated the participant thought the image on the right had the highest arousal or valence value; –1 indicated the participant thought the left image had the highest value.

but the same valence value. We randomly placed the image with the intended higher arousal on the left or right. We then asked the participants whether they felt a difference in arousal or valence between the left and right images in each pair and, if so, how great it was. The ratings ranged from 1 (the right image had the highest arousal or valence value) to –1 (the left image had the highest arousal or valence value).

Figure 4 shows the results; in it, the higher-arousal images are on the right to provide a clearer comparison. Most participants found many more differences in arousal than in valence. Also, the difference in arousal ratings across the participants was much lower than the difference in valence ratings. These results validate our theory on how the icon shape can indicate the degree of arousal.

A Valence Experiment

The participants also rated differences in arousal or valence after seeing 10 image pairs with different valence values but the same arousal values.

Figure 5 shows the results. Although the differences in arousal aren't as large as the differences in valence, they're considerable. This suggests that arousal and valence might be not fully independent, even though they form two axes in the 2D emotion plane we're using. The participants' valence ratings varied significantly, suggesting that precisely evaluating valence is more difficult than evaluating arousal.

Emotion and Color

Color is known to affect emotion, and a large body of literature exists about the psychology of color. So, we color our icons to reinforce their intended significance.

According to Patricia Valdez and Albert Mehrabian, saturation is more related to arousal, whereas brightness has more influence on valence.[4] Emotional reactions to hue (pure color) are more controversial. Nevertheless, many researchers have represented emotion by hue (for instance, red can connote energy, danger, anger, passion, desire, or love), suggesting an established interpretation of hue. This might be because culture and education largely determine the emotions associated with hue.

Therefore, we generated icons with our subjective choice of hues (h = 270 degrees) but with saturation and brightness corresponding to arousal and valence.

Testing Affecticon

Figure 6 shows Affecticon's interface. It resembles a standard file explorer system, with our affective music icon replacing the generic icon used for a music file. To create each icon, Affecticon determines AV values for the music. From these values, it generates a colored harmonograph curve. Users can sort the icons by arousal or valence. They can create a set of files by dragging them from the file-viewing window to a playlist.

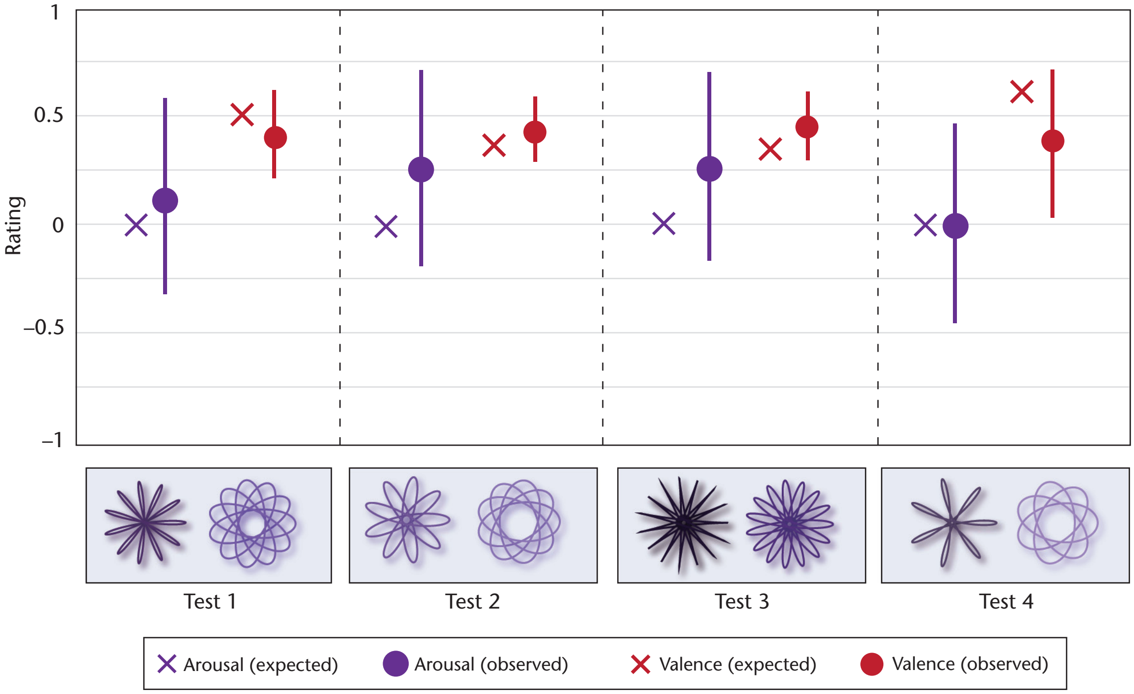

Figure 5. Results for four image pairs from the valence experiment. The participants' valence ratings varied significantly, suggesting that precisely evaluating valence is more difficult than evaluating arousal. A rating of 1 indicated the participant thought the image on the right had the highest arousal or valence value; –1 indicated the participant thought the left image had the highest value.

The Affective versus the Regular Icon

Our experiment involved 35 participants (ages 21 to 35) who were familiar with the standard file explorer system and had no extreme musical preferences. We prepared two music collections containing well-known songs with which most participants would be familiar. We asked them to generate two playlists from each collection, using the generic icons to generate one playlist and the affective icons to generate the other. We told them each playlist should contain 10 songs they wanted to listen to while driving to a picnic with their girlfriend or boyfriend. We told them there was a way to sort the affective icons but didn't explain how.

As Figure 7 shows, the participants generally spent less time generating the playlist with the affective icons. However, some participants took longer because they needed time to understand the sorting function and adapt to it.

After a few days, we conducted a similar experiment with the same participants, using other music collections. Figure 7 also shows these results. The participants chose playlists more quickly, presumably because they were more familiar with the interface.

We also compared the AV values of the songs selected with the generic icons with the values of the songs selected with the affective icons (see Figure 8). As we expected, all the selected songs had high arousal values, indicating that the participants preferred exciting songs. The AV values for each type of icon seem quite similar.

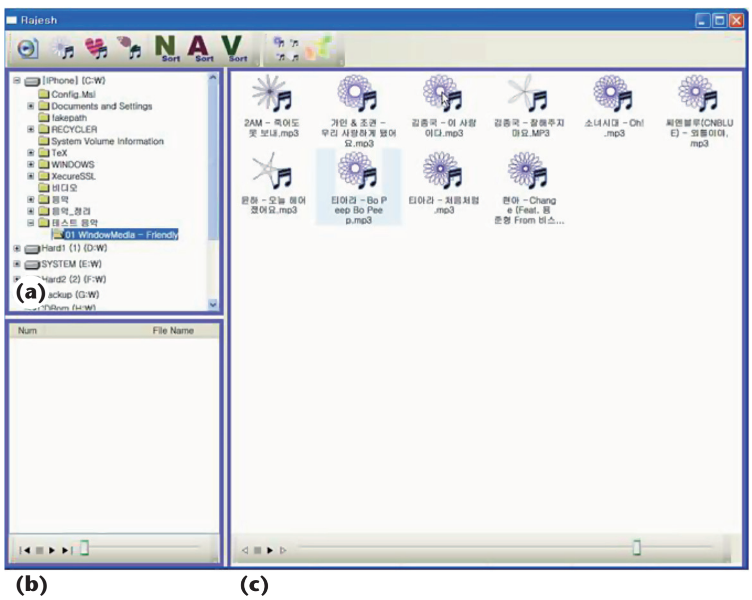

Figure 6. Affecticon's prototype interface. (a) The folder-viewing window. (b) The playlist window. (c) The file-viewing window. To create each icon, Affecticon determines arousal and valence values for the related music. From these values, it generates a colored harmonograph curve.

Familiar versus Unfamiliar Music

It's reasonable to assume that our interface will

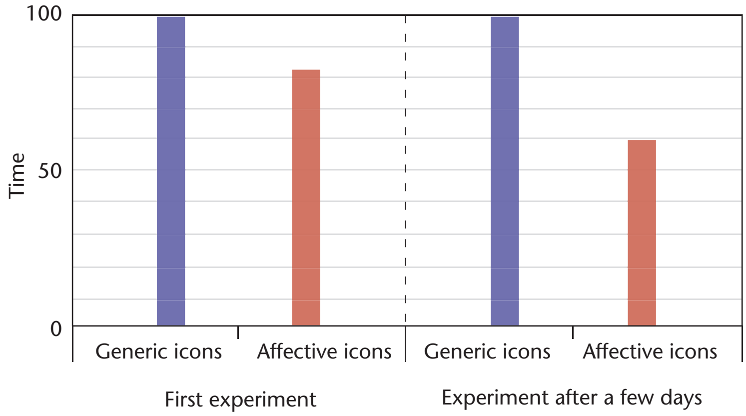

Figure 7. Relative times for generating a playlist using the regular and affective icons. Playlist generation was generally quicker with the affective icons.

be more useful in selecting music from collections of unfamiliar songs because participants already know the emotional associations of familiar songs.

To test this hypothesis, we created two more music collections, one with well-known songs and the other with more obscure songs (for example, ones by indie bands). We conducted tests similar to those we just described. Figure 9 shows the results. As we expected, the ratio of the time to select songs using the generic icons to the time for the affective icons increased for the unfamiliar songs. That is, the affective icons let the participants more quickly select those songs.

Figure 10 compares the AV values for the two types of icons. Compared with the first two experiments (see Figure 8), the distribution is less biased toward high arousal. We think this is because participants had more trouble choosing a playlist from unfamiliar songs, so they selected middle-of-the-road songs without listening to all the songs. For the affective icons, the arousal values were higher. After the experiments, the participants said that the affective icons helped them produce a better playlist. So, we deduce that affective icons can speed up selection of unfamiliar songs.

We plan to extend Affecticon by designing and analyzing a more comprehensive set of music features. Because feature extraction is the system's most expensive part, it needs to be more effective to be valuable for commercial use. We also hope other elaborated methods such as nonlinear modeling can improve Affecticon's accuracy. In addition, we're considering other icon shapes that can represent emotional content intuitively. Finally, we're exploring using various hue values for icons because hue values are also important for representing emotional content.

Acknowledgments

This work was supported by the Industrial Strategic Technology Development Program (grant 10035538, Developing a Tool Converting a Video to Nonphotorealistic Animations), funded by the Korean Ministry of Knowledge Economy and Mid-career Researcher Program through National Research Foundation grant 20100000389, funded by the Korean Ministry of Education, Science, and Technology.

References

1. P.J. Lang, M.M. Bradley, and B.N. Cuthbert, *International Affective Picture System (IAPS): Instruction Manual and Affective Ratings*, tech. report A-6, Center for Research in Psychophysiology, Univ. of Florida, 2005.
2. J. Scheirer and R.W. Picard, *Affective Objects*, tech. report 524, MIT Media Lab, 1999.
3. G.L. Collier, "Affective Synesthesia: Extracting Emotion Space from Simple Perceptual Stimuli," *Motivation and Emotion*, vol. 20, no. 1, 1996, pp. 1–32.

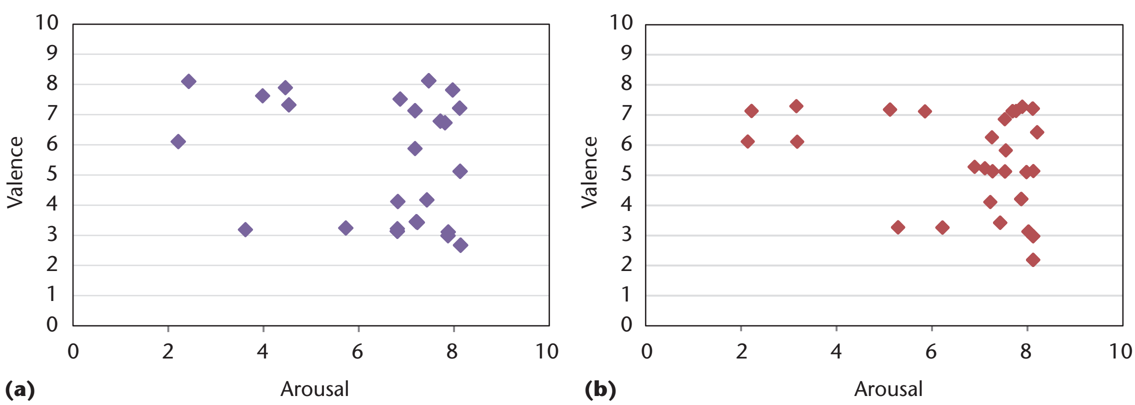

Figure 8. Arousal and valence values of songs selected using the (a) regular and (b) affective music icons. The values for each type of icon seem quite similar.

4. P. Valdez and A. Mehrabian, "Effects of Color on Emotions," *J. Experimental Psychology: General*, vol. 123, no. 4, 1994, pp. 394–409.

***Min-Joon Yoo** is a researcher at Yonsei University's Institute of Culture Technology. Contact him at debussy@cs.yonsei.ac.kr.*

***In-Kwon Lee** is an associate professor in Yonsei University's Department of Computer Science. Contact him at iklee@yonsei.ac.kr.*

Contact department editor Mike Potel at potel@wildcrest.com.

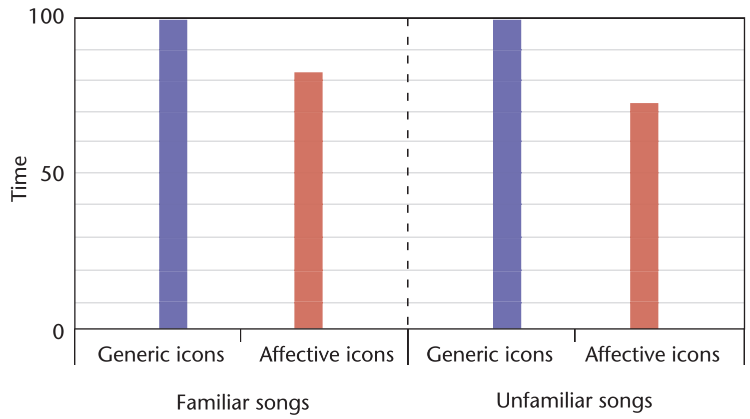

Figure 9. Relative times for generating playlists with the generic and affective icons (normalized to the generic icons). The affective icons let the participants more quickly select unfamiliar songs.

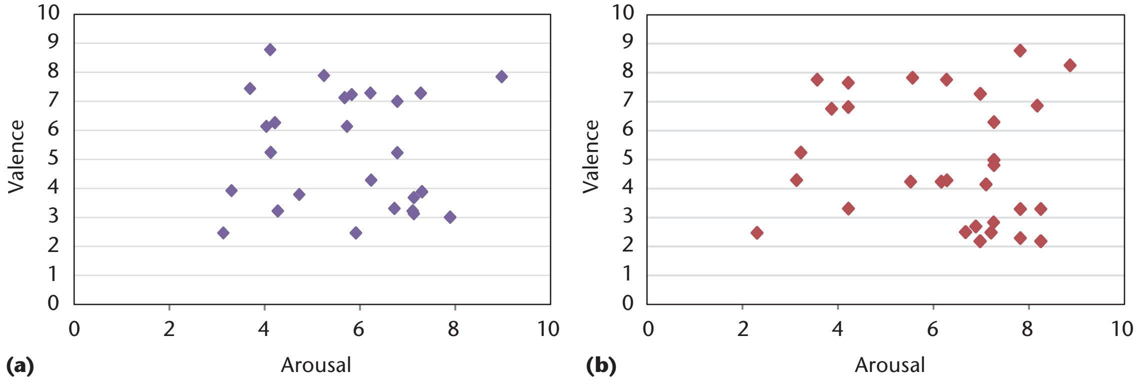

Figure 10. Arousal and valence values of songs selected from an unfamiliar music collection, using the (a) generic and (b) affective icons. The participants reported that the affective icons helped them produce a better playlist.