2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Analysis of the Subjective Impression on Combinations of Images and Music

Keiko Sato† and Yasue Mitsukura‡

†Department of Psychology, Ritsumeikan University
56-1, Kitamachi, To-jiin, Kita-ku, Kyoto 603-8577 Japan
‡Department of System Design Engineering, Keio University
3-14-1 Hiyoshi, Kohokuku, Yokohama 223-8522 Japan
Email: keikos-a@st.ritsumei.ac.jp, mitsukura@sd.keio.ac.jp

Abstract—Auditory information plays an integral role in AV media because even identical images are perceived differently when they are matched with different music. We now present a few studies in which the changes in subjective impressions of when visual information is matched with auditory information were analyzed on the basis of the physical features. The purpose of this study is to investigate the effects of music on image impressions in terms of physical features of images or music. In this paper, we first elucidate the changes in subjective impressions when the image is presented without music and when it is presented with music. Secondly, to clarify the relationship between the impression and physical features, we compare their respective physical features, which include color information, structural information, and frequency characteristics with subjective impressions. As a result, the color information of an image containing green or saturation and the power of the music were strongly correlated with adjectives expressing activity. Moreover, the entropy of saturation correlated with words expressing spatial extent.

1. Introduction

It is known that auditory information has a significant impact on the perception of visual information because even identical images are perceived differently when they are matched with different music [1]. How different sensation modalities such as visual and auditory interact with each other. It is generally known that the interaction of the two sensory modalities enhances the overall effect. In the field of perceptual psychology, cross-modal sensory interaction is well documented. For example, "brightness" represents both visual and auditory impression. This is known as intermodality phenomenon. Moreover, it is known that auditory "brightness" alters visual "brightness" in the same direction. This phenomenon is called consonance. As previous works, Iwamiya has showed that the various effects created by combination of music and motion picture from the emotional viewpoint [2]. Cohen has observed the direct influence of musical meaning on film meaning [3].

On the other hand, researches of the interaction between the visual and auditory sensory modalities include studies on inter-retrieval systems among different media. For example, a selection method of music corresponding to pictures has studied [4]. However, this doesn't define the relationship between physical features and human feelings. Although we can present some previous studies focused on physical features. Lipscomb et al. have showed that musical soundtracks can influence the meaning of film and the changes in the impression of visual image by variation of tonality or tempo [5]. We now however present a few studies in which the changes in subjective perceptions of when visual information is matched with auditory information were analyzed on the basis of physical features of the perceived items. It is imperative that the relationship between physical features and human emotions be defined.

The purpose of this paper is to elucidate the effects of music on image impression considering physical features. As the first step, general scenery images and piano music are used and we investigate the changes in subjective impression when the image is presented without music and with music. We then compare the different image or music impressions with each other and also compare their respective physical features. In this paper, we first perform the subjective experiment that is based on the semantic differential method. In this experiment, subjects evaluate the image impression in the case where the image is presented without music and where the same image is presented with music in the background. Secondly, we apply the factor analysis technique to the data obtained from the experiment. Finally, in order to reveal the relationship between the impression of an image or music and their physical features, we calculate the correlation between the impression value and the magnitude of the physical features. The physical features of the images considered are color information and structural information. Among the physical features of music, the frequency characteristic is used. For all of these analyses, we find out effects of music on image impression and the relationship between impressions and the physical features.

2. Experiment

We used 8 different stimuli for image and music as shown in Table 1. Fig. 1 shows the sample image presented to subjects in the experiment [6]. These stimuli were selected based on the preliminary questionnaire to prevent the impression bias of stimuli used. In this experiment, the

Table 1: Image and Music

No.	Image	*	Music	*
1	Summer beach	bright	Le nozze di Figaro, K. 492: ”Voi che sapete”/Mozart	birght
2	Animals in the field	bright	Piano Concerto No.1 in E minor, Op.11 1st mov./Chopin	dark
3	Field of flowers	bright	Czardas/Monti	dark
4	Commuter	dark	Symphony No. 5 4th Mov./Mahlar	dark
5	Heavy rain	dark	Piano Concerto No.5 in E-flat major Op.73 ”Emperor” 3rd mov./Beethoven	birght
6	Scene of mountains	dark	Violin Sonata No. 5 in F major, Op. 24, ”Spring” 1st mov./Beethoven	bright
7	Sunshine filtering though foliage	neither	Lieder ohne Worte Op.62-6 ”Fruhlingslied”/Mendelssohn	bright
8	School road in the rain	neither	Albinoni’s Adagio/Albinoni	dark

*Impression of preliminary questionnaire

Figure 1: Sample image (summer beach)

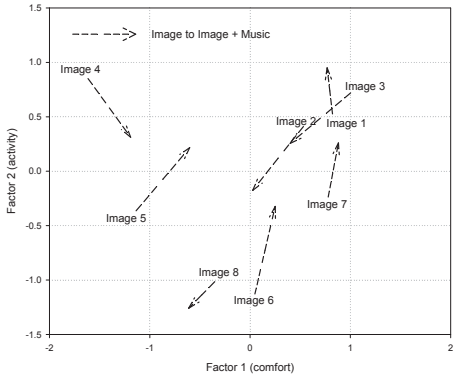

Figure 2: Changes in the image impression

image (600 x 400 pixels) was first presented without music. The subject evaluated the impression of the presented image using impression evaluation words, which are 15 kinds of scales with 5-points bipolar scale method as shown in Table 2. Next, same image was presented with music. Music were played in the combination as shown in Table 1. The subjects were assigned the task of selecting the impression for 19 images in total with another 3 images (dummy images) which are not used for analysis, also randomly ordered. The subjects were 20 students (11 males and 9 females). This experiment took about 20 minutes.

3. Analysis I

We applied the factor analysis to the data obtained from the experiment. The data include the impression value of

Table 2: Result of the factor analysis

Impression word	Factor	
	1	2
pleasant – unpleasant	0.87	0.21
loose – tight	0.86	-0.12
beautiful – ugly	0.82	0.15
warm – cold	0.81	0.25
bright – dark	0.75	0.54
extended – narrow	0.74	0.17
soft – hard	0.74	0.19
cheerful – gloomy	0.74	0.51
active – inactive	0.70	0.47
light – heavy	0.66	0.51
wet – dry	0.42	0.15
fancy – plain	0.46	0.62
new – old	0.23	0.59
calm – excited	0.43	-0.57
clear – cloudy	0.04	0.23
Contribution rate(%)	43.81	15.78

when the image was presented without music and that of when it was presented with music. Factor analysis using principal components extraction with varimax rotation was performed on the variables of 15 items. Table 2 presents the result of the 2-factor extraction. This table shows that Factor 1 includes some adjectives expressing “comfort” and Factor 2 represents “activity”.

Fig. 2 shows the distribution of mean factor scores of 20 subjects. In this figure, a starting point of an arrow represents the score of when the image was presented without music, and an end point of an arrow means the score of when same image was presented with music. Results obtained from the score of when only the image was presented lead us to interpret that images 1, 2, 3, and 7 have positive score of Factor 1 expressing “comfort”, that is, their impressions are pleasant, loose, beautiful, and warm. Whereas images 4, 5, and 8 have negative score of Factor 1. These images have the opposite impression. Moreover, impressions of images 1, 2, 3, and 4 are fancy and new, and that of images 6 and 8 are plain and old.

Next, in the results obtained from the score of when same image was presented with music, we can understand that both score of images 5, 6, and 7 with music are higher. On the other hand, both score of images 2, 3, and 8 with music are lower. Music which were played in images 2, 3,

Table 3: Physical features of images

Mean value of each color sapce	Entropy of each color space	Direction feature (Gabor filter)
red	r-entropy	45°
green	g-entropy	90°
blue	b-entropy	135°
hue	h-entropy	180°
saturation	s-entropy	
intensity	i-entropy	

Table 4: Physical features of music

Mean value of power (mean)
Standard deviation of power (SD)
Mean value of low frequency power (MLFP)
Mean value of high frequency power (MHFP)
Maximum value of power (max)
Frequency at the maximum value of power (frequency)

Table 5: Result of the correlation analysis (factor)

Factor	Physical features of images
Factor 1	direction feature†(45°, 135°, 180°), s-entropy†
Factor 2	saturation†
	Physical features of music
Factor 1	SD†, max∗
Factor 2	mean∗, MLFP†, MHFP∗

† : $p < 0.1$, ∗ : $p < 0.05$

4, and 8 have a dark impression, and music in images 1, 5, 6, and 7 have a bright impression from the result of the preliminary questionnaire. It was shown that "bright" music improved the score of image impression, whereas "dark" music lowered it. we can however see the opposite for the results of image 1 and 4. Music with image 1 lowered the "comfort" score, whereas music with image 4 improved it.

4. Analysis II

In this section, we describe the analysis conducted in order to reveal the relationship between the impression of an image or music and their physical features.

4.1. Physical Features of the Images

The physical features of the images considered in this study are color information and structural information as shown Table 3. Color information is expressed by each average value and entropy in RGB and HSI color spaces. Another color information, entropy is measurement on a color distribution that gives a value on the amount of information the distribution contains. Entropy is then expressed as follows:

$$H = -\sum_{i=1}^{n} P(i) log_2 P(i) \tag{1}$$

where the P, are the relative rates of occurrence of each of the n (in the case of R-value, $n = 256$) in the picture. Characteristically, the pixel histogram is broad and multimodal, with large entropy.

Next, we consider structural information as a direction characteristic. The use of a Gabor filter is required to extract the direction feature from the image. Given an image, we convert an RGB color image to grayscale and pass it through a bank of the Gabor filter shown below. The impulse response of the Gabor filters $F(x, y)$ is given by

$$F(x,y) = exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) cos\left(\frac{2\pi x'}{\lambda}\right) \tag{2}$$

where,

$$x' = xcos\theta + ysin\theta, y' = -xsin\theta + ycos\theta$$

where γ denotes the aspect ratio of the filter, θ represents the rotation angle, λ is the radial frequency, and σ means the spatial frequency. In this paper, we applied only 4 rotations ($\theta = 45°, 90°, 135°, 180°$) to 8 images. The parameters are as follows: $\sigma = 5$, $\lambda = 4.2$, $\gamma = 1$

4.2. Physical Features of the Music

Among physical features of music, the frequency characteristic obtained by FFT (Fast Fourier Transform) is used. In this paper, we transformed 8 music (30 second music clip at the beginning) into the spectrograms with a window size of 1024 samples at 44.1kHz. A hamming window with 50% overlap was used and FFT was calculated in each window. We next calculated the value of mean power at all frequencies (mean), the standard derivation (SD), the mean value at low frequency (MLFP), the mean value at high frequency (MHFP), the maximum value of power (max), and the frequency at the maximum value of power (frequency). Table 4 shows the physical features of music considered in this paper.

4.3. Correlation Analysis

After that, we calculated the correlation coefficient between two factors and the mentioned physical features. the correlation coefficient which is expressed the strength of the relationship between factor score (mean of 20 subjects) and feature value of the image is as shown below.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{3}$$

We did the same with music feature. Where, the impression value of music was used for the difference value between that of when the image was presented without music and that of it was presented with music. Finally, we tested the statistical significance of each correlation coefficient calculated. The statistical significance of r was tested using a

Table 6: Result of the correlation analysis (impression word)

Impression word	Physical features of images	Physical features of music
pleasant – unpleasant		MHFP, max
loose – tight	s-entropy	
beautiful – ugly		
warm – cold	saturation	SD, MHFP, max
bright – dark	saturation, green	max
extended – narrow	s-entropy	
soft – hard	saturation, direction feature (45°, 135°, 180°)	max
cheerful – gloomy	green	
active – inactive	saturation	mean, MLFP, MHFP, max
light – heavy	saturation	mean, MLFP, MHFP, max
wet – dry	blue, g-entropy, h-entropy	
fancy – plain	green, saturation, intensity	mean, MLFP, MHFP, max
new – old		
calm – excited		frequency
clear – cloudy		mean, SD, MLFP, MHFP, max

t-test. Test static t_0 is expressed by

$$t_0 = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \tag{4}$$

where r and n are the correlation coefficient and the number of data, respectively. t_0 follows a t distribution with $n-2$ degrees of freedom. In this paper, we performed the two-tailed test at the 5% significance level. Moreover we calculated the correlation coefficient between each impression value (15 words) and the physical features and tested the statistical significance.

4.4. Result and Disucussion

Tables 5 and 6 list the results of the correlation analysis. Table 5 shows the physical features which achieve the significance at the 5% level or marginal significance in connection with factors. Factor 1, or "comfort" is correlated with direction feature, saturation-entropy, and SD and maximum value of power. Whereas Factor 2, namely "activity" is correlated with saturation, and mean value of power, low and high frequency power.

Table 6 presents the result of the correlation with each impression words. We can see that saturation is correlated with the most impression words. These words include many adjectives expressing "activity", so this is in agreement with the result of the correlation analysis with factors. Others, saturation-entropy links to "loose – tight" and "extended – narrow", and green element is correlated with "bright – dark", "cheerful – gloomy", and "fancy – plain". Moreover, the power of the music are strongly correlated with adjectives expressing "activity".

5. Conclusion

In this paper, we investigated the effects of music on image impression considering physical features. As the first step, general scenery image as visual information and piano music as auditory information were used and we elucidated the changes of subjective impression when the image was presented without music and with music. Moreover, in order to reveal the relationship between the impression of image or music and the physical features, we calculated the correlation between the impression value and the magnitude of the physical features. The physical features of images and music considered were color information, structural information, and frequency characteristic. The main findings were as follows: The color information of an image containing green or saturation colors and the power of the music were strongly correlated with adjectives expressing "activity". Moreover, the entropy of saturation correlated with words expressing spatial extent.

References

[1] T. Oyama, S. Imai, and T. Wake (Eds.), "Handbook of Sensation and Perception -revised edition-", Seishin-Shobo, 1994. (in Japanese)

[2] S. Iwamiya, "Interactions between auditory and visual processing when listening to music in an audio visual context: 1. Matching 2. Audio quality", *Psychomusicology*, vol.13, pp.133–154, 1994.

[3] A.J. Cohen, "Associationism and musical soundtrack phenomena", *Contemporary Music Review*, vol.9, part 1 & 2, pp.163–178, 1993.

[4] H. Koga, Y. Shimoshio, and Y. Oyama, "Selection method of music suitable for motion pictures with kansei information", *Technical report of IEICE. HCS*, vol.99, no.320, pp.25–32, 1999. (in Japanese)

[5] S.D. Lipscomb and R.A. Kendall, "Perceptual judgement of the relationship between musical and visual components in film", *Psychomusicology*, vol.13, pp.60–98, 1994.

[6] Healing & Relaxation Photo Image Database, http://iyashi.midb.jp