

Consideration of Expression Method of the Entropy Concept

Correlation between the entropy for a given binary format image and the psychological quantity from verbal expression

Tomoaki Sato¹, Mutsumi Suganuma²

¹ Kanagawa Institute of Technology, Japan, satotomo@me.kanagawa-it.ac.jp

² Waseda University, Japan, mutsumi@aoni.waseda.jp

Abstract: The concept of entropy which was firstly introduced as a thermodynamic property has been extended to the economics theory, information theory, and others. Even though entropy is used in many fields of science and technology, it is not easy to understand and represent the concept of entropy, as well as describing it by words/terms. We assume that one of the causes of such difficulty is the gap between psychological quantity caused by words/terms used in explaining the concept of entropy and actual entropy value. We have been investigating the correlation between the thermodynamic entropy and the psychological quantity obtained by showing contents such as particle animations simulating the molecular motion to participants. We had found that the verbal expressions such as "degree of diffusion", "degree of randomness" and "elusiveness" are relatively suitable in representing the entropy concept than other verbal expressions. In this report, we investigated the correlation between the entropy and the psychological quantity obtained by showing several contents which show binary dot pattern images to participants. We examined the correlation between the entropy values calculated by several different algorithms and the psychological quantity by above three verbal expressions. In the comparisons between the binary dot patterns which have natural randomness that may exist in real photo picture, we obtained that the psychological quantity by above three verbal expressions have high correlation with entropy values. We found that the entropy values calculated by means of measuring the length between each black pixel has highest correlation to the psychological quantity.

Keywords: Entropy, Thermodynamics, Information Theory, Language Expression, Visual Expression.

1. INTRODUCTION

The concept of entropy, which is known as measure of fuzziness, disorder, or value now, was firstly established as the second law of thermodynamics by Clausius and Boltzmann in 1850 and has been extended to the economics theory, information theory, and others (Greven, Keller, and Warnecke, 2003). However, because of its wide range of application, the concept of entropy may be interpreted in a variety of ways. Therefore, it is difficult to understand its essence precisely.

The concept of entropy, originally introduced in the context of irreversibility of the second law of thermodynamics, is defined as quantity representing the states of matter such like heat, pressure, or content. Heat, pressure, or content are quantity which we can perceive, therefore those are assumed to be easy to understand. On the other hand, although entropy is defined as quantity representing the sates of matter as well, it is not well known whether entropy itself could be “perceived” or not. Therefore, we assume that this is one of the causes which make the concept of entropy to be difficult to understand. Being interested in whether entropy may be felt by our kansei or not, authors has been investigating the efficient way of expressing the concept of entropy, especially in verbal expression.

In our previous study (Sato and Suganuma, 2013b), we conducted an experiment in which we showed our participants several animations, simulating molecular motion. We made eight animations varying the range of molecular motion (content), and/or speed of molecular motion (heat), resulting in different entropy. Our participants were asked to view two animations presented simultaneously, along with several verbal expressions (in Japanese). They chose which animation suits better to the presented verbal expressions. As a result, we observed relatively high correlation between the actual entropy value of the animation and three verbal expressions. Those three verbal expressions were “degree of diffusion”, “degree of randomness”, and “degree of elusiveness” in Japanese. The term “degree of diffusion,” which the term used in the experiment was “kakusanno-doai” in Japanese, may include the meaning of diffusion, spreading, or scatter. The term “degree of randomness”, which the term used in the experiment was “ranzatsusa” in Japanese, may include the mening of, random, disorder, or mess. The term “degree of elusiveness,” which the term used in the experiment was “tsukamaenikusa” in Japanese, may include the meaning of elusive, or “to be hard to catch/touch”.

The concept of entropy is applied to information science as well as thermodynamics. As entropy in thermodynamics is based on molecular motion, we may consider that it represents the states of matter in dynamic situation. On the other hand, entropy in information science which is based on variation of symbols may be considered as entropy in static situation. In previous reports (Sato and Suganuma, 2013b; Sato, 2013), from thermodynamics point of view, we investigated for more appropriate verbal expression corresponding to the variation of entropy in dynamic situation based on kinetic theory of molecules. We considered that investigation of the relation between our kansei quantity and entropy in static situation is important as well as those in dynamic situation, to explore the nature of human understanding of entropy irrespective of dynamic or static entropy. Thus, in this study we examined the relation between kansei quantity raised by verbal expression and entropy of static binary image (example shown in figure 1).

Method of calculation based on information theory is common in calculation of entropy for binary images (Kamata, 2003). However, we could not find calculation method considering the kansei quantity. Therefore, we investigated the adequateness of calculation method in means of human kansei quantity on verbal expressions, as well.

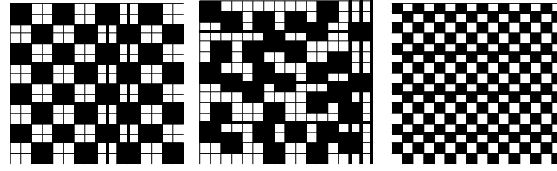


Figure 1: Binary image example of black and white (number of white and black pixels are the same)

2. CALCULATION METHOD OF ENTROPY VALUE OF BINARY FORMAT IMAGES

There are numbers of method of entropy calculation on binary images. Calculation method based on amount of information has been established in context of information science and image processing. However, when taking the correspondence relation between kansei quantity on verbal expressions and actual entropy value of binary images, there are little arguments on how to calculate the entropy value based on kansei quantity caused by verbal expression, or, to say, amount of entropy interpreted by our kansei. In the following, we discussed on the calculation methods of entropy value on binary images.

2.1. Calculation method of entropy value based on Mean Information Content

In the field of information science, entropy has been commonly used to evaluate the efficiency of information compression or information transfer rate of digital images. However, we assume that applying this calculation method in following experiment is questionable.

In general, amount of information entropy S may be calculated by formula (1) defined by Shannon.

$$S = \sum p_i \cdot \log_2(1/p_i) \quad (1),$$

where, p_i is occurrence probability of every event, and $\sum p_i$ is limited to 1. Assuming that every symbol which transmit information having equal occurrence probability, $p = 1/M$ stands, where M is total number of the symbols. Therefore,

$$S = M \cdot 1/M \cdot \log_2(M) = \log_2 M \quad (2),$$

stands.

However, as formula (2) depends on occurrence probability of the symbols, entropy value found by formula (2) varies according to the difference of symbol itself or its size. Moreover, in field of information science, calculation algorithm of entropy may vary depending on the information compression method.

In case of image entropy calculation, when the size of the image is $I \times J$ and each pixel has Z bit of luminance (color) depth and taking the luminance value as a symbol, entropy S may be found by following formula

$$S = - \sum_{i=0}^{2^Z-1} N_i / N_{pic} \log_2(N_i / N_{pic}) \quad (3).$$

Here, N_i is number of pixels having luminance value of Z_i , and N_{pic} is number of pixels in the image.

Based on formulas (1)~(3) and variation of how we set the symbols, following three types of entropy may be considered.

(a) information entropy of the binary image 1: taking one bit (whether a pixel is white or black) as a symbol.

(b) information entropy of the binary image 2: taking two bit (area of 2×2) as a symbol

(c) information entropy of the binary image 3: taking the length of the sequence of black pixel

2.2. Calculation method of entropy value of image based on the probability of a thermodynamic phenomena

In previous section, we have proposed calculation methods of entropy on binary images based on the concept of information theory. However, in thermodynamics, the concept of entropy is often explained in stochastic aspects with the number of possible microscopic states in text books or study guide of entropy. Here, we applied the concept of entropy in thermodynamics on entropy calculation of binary images.

One of the most important formulas in the field of statistical thermodynamics is the entropy formula defined by Boltzmann. The entropy S may be found by

$$S = k \cdot \ln W \quad (4),$$

where k is Boltzmann constant (gas constant of single molecule), and W is a microscopic probability of a molecule's thermodynamic phenomenon (the number of possible microscopic states). Alternating the probability of thermodynamic phenomenon into number of cases of black pixel to appear, we considered two calculation methods of entropy of binary images.

(d) entropy value based on the probability of black pixels to appear in succession.

In case that the probability of black pixel to appear is $1/2$, p_i the probability of black pixels to appear in succession would be as shown in Figure 2. When the probability that all of 16 pixels will become black is set to 1, entropy is calculated in quest of the ratio of the occurrence probability of each pattern as probability of a thermodynamic phenomenon.

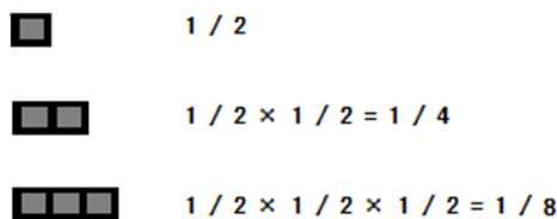


Figure 2: Probability of black pixel(s) to appear consequently

(e) Entropy on the basis of the distance between the black pixels in every pair

When people look at a binary image, it is thought that it feels as a two-dimensional spread. Therefore, it is thought that the method of calculating entropy for each line or each sequence one-dimensionally does not fully reflect the feeling of the two-dimensional spread. Therefore, we considered a method of computing the entropy value radially by means of measuring the distances of all pixels. Here, it assumes that the probability that a black pixel may appear is equally given to the whole picture, and we pay its attention to the distance between the two black pixels A and B as

shown in Figure 3. In this case, the occurrence probability of B to A is proportional to length l of the circumference which is calculated by measuring the radius r .

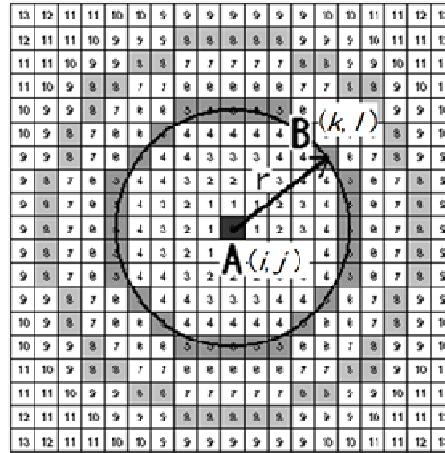


Figure 3: Existence probability which increases with the increase in the distance r from the pixel A to B

Fig. 3 shows how many pixels B can exist in the same distance from the pixel A located at the center, if a length of 1 pixel is defined as 1. In this figure, numbers written on the each pixel show the distance from the central pixel A. For example, pixels at a distance of 5 from pixel A are 28 pixels shown shaded dark in Fig. 3. Similarly, there are 48 pixels at a distance of 8 from the pixel A (lightly shaded in Figure 3.). Thus, the longer the distance between one pixel to pixel A is, the higher the probability of existence of the pixel is. If this probability is made into the probability of a thermodynamic phenomenon, the probability of a thermodynamic phenomenon is proportional to the circumference length which is calculated from the length between two pixels. Therefore, the entropy S_{AB} between two pixels can be calculated by the following formula as $W=r$.

$$S_{AB} = \ln r \quad (5).$$

Furthermore, in an actual program the following formula (6) will be applied.

$$S = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n Z_{i,j} \cdot Z_{k,l} \cdot \ln \left(\sqrt{(k-i)^2 + (l-j)^2} \right) \quad (6).$$

Here, i, j, k , and l are the coordinate values of two pixels A (i, j) and B (k, l), respectively, as shown in Figure 3. Moreover, at the time the pixel is white, Z becomes 0 and at the time the pixel is black, Z becomes 1. The value computed with this algorithm shows a spatial spread of a black pixel directly.

3. EXPERIMENTS WHICH EXAMINE THE CORRELATION BETWEEN CALCULATING VALUE OF ENTROPY AND PSYCHOLOGICAL QUANTITY FROM VERBAL EXPRESSION

In chapter 2, we discussed about the calculation methods of entropy values on binary images. We have proposed calculation methods based on the average amount of information and number of possible microscopic states. Our interest is in the relation between the difference of entropy (caused

by the difference of calculation method) and kansei quantity caused by verbal expressions. Hence, we conducted an experiment testing the relationship between the entropy value and kansei quantity caused by the verbal expressions. To minimize the task load of participants, verbal expressions used in this experiment was limited to three. Those were Japanese terms which mean "degree of diffusion", "degree of disorder", and "degree of elusiveness". These verbal expressions were chosen from our previous studies (Sato and Suganuma, 2013a, b; Sato, 2013), which we found the kansei quantity caused by verbal expressions to have relatively higher correlation to the actual entropy value in dynamic situation.

3.1. Experiment method

Participants. Seventy-seven students of Kanagawa Institute of Technology participated the experiment as a course credit. They were all sophomores at department of mechanical engineering, faculty of engineering. At the time the experiment was conducted, they were not taught about entropy yet. Therefore, we assumed that most of participants were not familiar with the concept of entropy at that point.

Stimuli. We made nine binary images as shown in Figure 4 (images A ~ I). Three verbal expressions on which participants made evaluation of binary images were Japanese phrases which mean "degree of diffusion", "degree of disorder", and "degree of elusiveness."

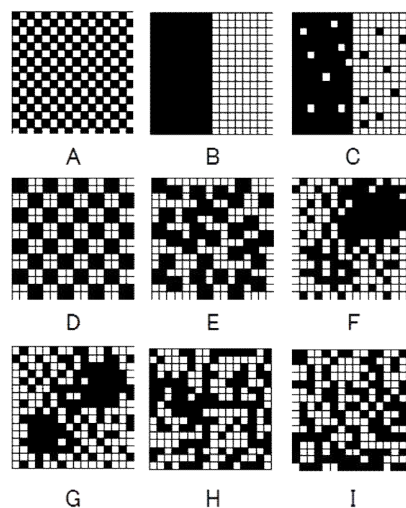


Figure 4: Binary image patterns used in the experiment

Procedure. The experiment was conducted in pair comparison method manner. In each trial, we simultaneously presented two binary images chosen from nine images shown in Figure 4. Three verbal expressions were shown as well. Participants were asked to choose which binary image seems to be more suitable for each verbal expression. The combination of binary images resulted in 36 pairs. Thus we collected 108 responses (36 image presentations x 3 verbal expressions) from one participant.

3.2. Experimental result

Entropy values of binary images for calculation methods (a) ~ (e) are shown in Figure 5. According to Figure 5, we can see that entropy values and the rank order of binary images largely differ due to the difference of calculation methods. Especially, images A and D showed large difference as it was ranked higher in calculation method (a), (e), and (d), but ranked lower in the other calculation methods.

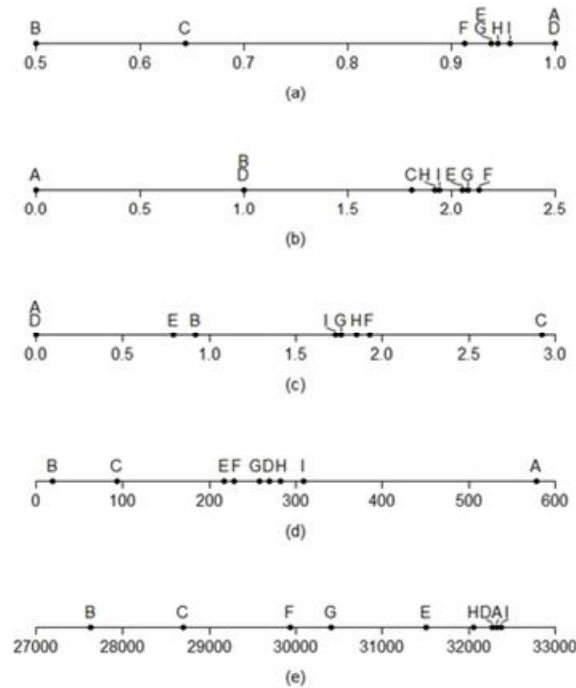


Figure 5: Comparison of the calculated value of the entropy S by five calculation methods. Charts (a), (b), (c), (d), and (e) show the result of three entropy values by means of different calculation methods of information content, entropy value based on the probability of black pixels to appear in succession, and entropy value based on the probability of black pixels to appear in succession, respectively

According to Thurstone's paired comparison method, we ranked and scored nine images for three verbal expressions separately. For each verbal expression, participants' choice was tabulated as win-loss matrix (chosen to be suitable as winning). Then, the win-loss numbers were converted into win-loss ratio, and then converted into z-scores. The average of the z-scores for each image was calculated as the score of the image. These scores are assumed to reflect the kansei quantity caused by verbal expressions (we would refer this score as f). Figure 6 shows the difference of kansei quantity f of binary images for three verbal expressions. Here, image A, which showed a swing in entropy value rank, showed a similar tendency in f value as well. Participants considered image A to be suitable for "degree of diffusion" and "degree of elusiveness," however, not for "degree of disorder."

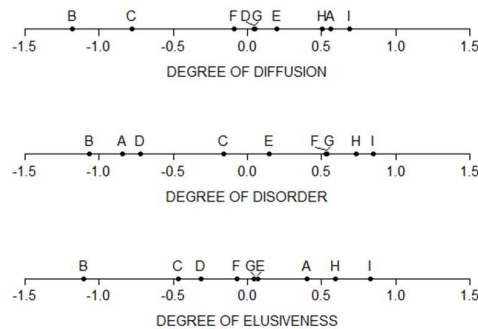


Figure 6: Comparison of the amount f of psychological quantity by three verbal expressions.

In table 1, we assessed the correlation (Pearson's correlation) between the entropy values for binary images by five calculation methods and kansei quantity f caused by three verbal expressions. The relation between the entropy value and the kansei quantity were verified for all the combinations (5 x 3) beforehand. We assumed that the relation between the entropy value and the kansei quantity is linear, and thus we used Pearson's correlation for evaluating their relation in this paper. Note that it was not monotonic increase for combinations of the calculation method and the verbal expression which has relatively low correlation in Table 1.

Table 1: The correlation coefficients (Pearson's r) for verbal expression and calculation methods (left half), along with cross correlations between verbal expressions (right half)

Verbal Expressions	Calculation Method of Entropy Value					Verbal Expressions		
	(a)	(b)	(c)	(d)	(e)	Degree of Diffusion	Degree of Disorder	Degree of Elusiveness
Degree of Diffusion	0.92	0.02	-0.24	0.82	0.95		0.51	0.95
Degree of Disorder	0.41	0.83	0.61	0.05	0.30	0.51		0.67
Degree of Elusiveness	0.78	0.17	0.04	0.72	0.83	0.95	0.67	

3.3. Discussion of the experimental result

As we pointed out earlier, result for binary images A and D showed a characteristic tendency for both the entropy value S and kansei quantity f . Their rank/score differed largely in both S and f , according to the calculation method or verbal expression.

3.3.1. Consideration about the effect of geometric regularity of the image on participants' response

In comparison of kansei quantity f caused by verbal expressions (Figure 6), image A had relatively high f for "degree of diffusion", but low f for "degree of disorder." In our previous study (Sato and Suganuma, 2013b; Sato, 2013) these two verbal expressions had relatively high correlation with actual entropy value of molecular animation. Therefore, we assume that these two verbal expressions are efficient in describing the concept of entropy. From that point of view, tendency which image A showed is apparently contradicted and thus the characteristic of kansei quantity f for image A seemed interesting.

However, it seems plausible for image A to have seemingly contradictory results. In case of kansei quantity for "degree of diffusion", because image A had black pixels spread entirely for the image, it may be seen as "diffused" to the participants. On the other hand, as image A had uniform and geometrical pattern, it may as well seen as "ordered" image, and resulted in low kansei quantity for "degree of disorder." In both the cases above, we assume that participant made their decision by their thought rather than through their instinct. This may be caused by their interpretation of the verbal expressions and the geometric regularity of the images we presented. The "kansei" quantity in this case may not reflect participants' intuition, but it may be a result of their logical thinking. In that sense, some of our images in the experiment may not be suitable for investigating the relation between the entropy and kansei quantity caused by verbal expressions.

Thus, to eliminate the effect of geometric regulation on kansei quantity we would exclude some of the images from the following discussion. Eliminated images were image A, B, C, D, and E. Image D had similar pattern to image A, and may be considered as geometrically regulated pattern as well. Image B had distinct regions in its both sides. Therefore we considered that it had geometrically regulated pattern. Image C was basically the same to image B, besides few "particles". Image E was relatively "natural," however it was consisted of four pixels in a block, which was relatively

unnatural, compared to the remaining images.

3.3.2. Binary image to be ranking by common sensibility amount in words of three

The remaining images were F, G, H and I (Figure 7). They all had random (unregulated) pattern of black and white pixels. These images, in contrast to the eliminated images, looked relatively natural. They had patterns which may be seen in part of actual photographs. Interestingly, the kansei quantity f for these images ranked in the same manner for three verbal expressions (Table 3).

Table 2 shows the calculated entropy values, kansei quantity caused by verbal expressions, and, cross correlation of kansei quantity for images F, G, H, and I. We found high correlation between every combination of verbal expressions which was close to 1. It is interesting that even for the verbal expression “degree of elusiveness” had high correlation to other verbal expressions. The terms “disordered” or “diffusion” are adverb/adjective which indicates the state of the object directly. On the other hand, “elusiveness” is to some extent a reflection of our capability. Therefore, the process of decision may differ between “disordered”, “diffusion” and “elusiveness”.

Although there may be a difference of decision, the tendency of three verbal expressions for four images was alike. Therefore, we assume that feeling, intuition, or kansei in this situation is strongly related to the nature of entropy.

Table 2: The correlation coefficients from the comparisons only using binary format image F, G, H, and I

Verbal Expressions	Calculation Method of Entropy Value					Verbal Expressions		
	(a)	(b)	(c)	(d)	(e)	Degree of Diffusion	Degree of Disorder	Degree of Elusiveness
Degree of Diffusion	0.92	-0.96	-0.59	0.97	1.00		0.99	1.00
Degree of Disorder	0.87	-0.91	-0.54	0.94	0.97	0.99		0.99
Degree of Elusiveness	0.91	-0.95	-0.57	0.96	0.99	1.00	0.99	

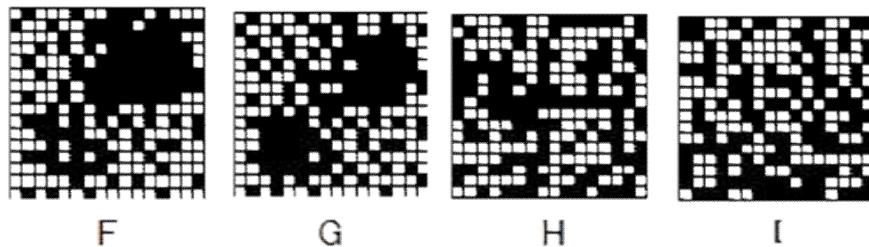


Figure 7: The binary images used on comparisons by natural expression

3.3.3. Consideration about the calculation method of entropy value of the binary format image based on the amount of sensitivities

Here, based on kansei quantity f for images F, G, H, and I, we will discuss about the relation of kansei quantity and calculation methods.

The correlation coefficient between the calculation methods and kansei quantity is shown in table 2. We found negative correlation for calculation methods (b) and (c). Possible reason for calculation method (b) was its lack of spatial resolution. That is, in images F~I, brightness of the pixels (black or white) altered in relatively higher frequency. However, the symbol in calculation method (b) was defined as 2x2 block of pixels, which turned out as lower frequency of brightness alteration. Therefore, calculation method (b) might have been inadequate for calculating entropy for image F~I.

Calculation methods (b) and (c) are algorithms which measure the diverseness of the symbols. For instance, calculation method (c) may be assumed as measuring the diverseness of length of the

symbols. If we take a look on images F and G, image F has single large chunk and image G has two chunks. When measuring the diverseness along horizontal axis, because image F has larger chunks, it would result in containing variations of series, compared to image G. Therefore, image F would result in larger entropy with calculation method (c). However, taking a look on both images, image F seems to be more coherent than image G, and does not match the calculation result from methods (b) and (c). We assume that such contradiction was caused by the characteristic of calculation methods (b) and (c), which the diverseness in these methods refer to those in one dimensional array of pixel data. On the other hand, when we see an image, we see it as two-dimensional space. Thus we assume that image information entropy may have less correlation with entropy which our kansei receive.

For calculation methods (a), (d), and (e), we observed higher correlation for all of them in table 2 in comparison to those shown in table 1. The difference between table 1 and 2 was that whether images A~E were included (table 1) or excluded (table 2) when assessing the correlation between entropy value and kansei quantity caused by verbal expression. In table 1, when all the binary images were included, correlation coefficient for calculation methods (a), (d), and (e) were 0.92, 0.82, and 0.95, respectively for verbal expression "degree of diffusion". Correlation between verbal expression "degree of disorder" were 0.41, 0.05, and 0.30 for calculation methods (a), (d), and (e) respectively. However, when images A~E were excluded (table 2), correlation between verbal expression "degree of disorder" were above 0.9 for calculation methods (a), (d), and (e).

As we discussed earlier, images A~E seemed not suitable for our objective. Therefore, when assessing the correlation between entropy values and kansei quantity caused by verbal expressions, results obtained from images A~E might have behaved as a noise. The fact that we observed higher correlation between calculated entropy values [calculated by methods (a), (d), and (e)] and kansei quantity caused by verbal expression when images A~E were excluded, suggests that those calculation methods reflects the attribute of entropy of binary image which our kansei receives. Especially, correlations for calculation method (e) were relatively high and close to 1 for all the verbal expressions. Those were 1, 0.97, and 0.99 respectively for "degree of diffusion", "degree of disorder", and "degree of elusiveness". Calculation method (e) is taking the distance between pixels on its basis irrespectively to horizontal or vertical direction. Differently from calculation methods (b) or (c), which take the one-dimensional distance between the pixels, calculation method (e) quantifies two-dimensional spread. We assume that this characteristics of calculation method (e) matched to the "entropy" we feel, and thus we assume that this calculation method was relatively suitable for calculating the entropy which participants felt.

4. CONCLUSION

In this study, we conducted an experiment investigating the relation between the entropy values for binary images and kansei quantity caused by verbal expressions describing the entropy. We asked participants to select a image which seemed to suit to verbal expressions in a pair comparison manner. Verbal expressions we presented to our participants were "degree of diffusion", "degree of disorder", and "degree of elusiveness". Furthermore, we applied variety of calculation methods for entropy calculation, and investigated for efficient calculation reflecting the human intuition or kansei on recognition of entropy. The results we obtained were as follows.

(1) From the comparison of kansei quantity caused by three verbal expressions ["degree of diffusion", "degree of disorder", and "degree of elusiveness"] for images which may exist in nature, we observed relatively high correlations between them. This result suggests that kansei which

functioned when evaluating the impression on binary images with relatively natural pattern was based on innate aspect of entropy.

(2) From the result of images selected in (1), it was suggested that entropy value based on information theory, which was used for index of amount of information, may differ from the "entropy" which human perceive from two-dimensional diversity. Furthermore, it was suggested that calculation method based on distance between two pixels to be relatively adequate to calculate or to measure the "entropy" which human perceive.

REFERENCES

- Greven, A., Keller, G., Warnecke, G.(2003). Entropy: Princeton University Press, USA
- Sato, T., Suganuma, M.,(2013a). Verbal expression of the entropy concept: Procedia - Social and Behavioral Sciences, 97, 443-447, from <http://www.sciencedirect.com/science/article/pii/S1877042813037026>.
- Sato, T., Suganuma, M.,(2013b). Consideration of Expression Method of the Entropy Concept: Correlation between the thermodynamic entropy obtained from the molecule movement animation and the psychological quantity from language expression, International Journal of Affective Engineering, 12(2), 303-309, from https://www.jstage.jst.go.jp/article/jjske/12/2/12_303/_article
- Sato, T., (2013). Verbal Expressions of Entropy: Relation between Thermodynamic Entropy and Psychological Amount of Entropy by Verbal Expressions, 7 th Conference of Japanese Society for Cognitive Psychology.
- Kamata, S.,(2003). Information Science & Engineering, Saience Corporation, Tokyo:Japan.