Development of a new loudness model in consideration of audio-visual interaction

Kai AIZAWA[1]; Takashi KAMOGAWA[2]; Akihiko ARIMITSU[3]; Takeshi TOI[4]

[1] Graduate school of Chuo University, Japan

[2, 3, 4] Chuo University, Japan

ABSTRACT

Recently, several researchers have been focusing on loudness perception of sound sources when the sounds are presented with the appearance of what radiates them. They concluded the visual information may increase/decrease the perceived loudness intensity. However, because these works concentrated mostly on subjective evaluation of actual sounds, there is a demand for an objective loudness model in consideration of the audio-visual interaction. Due to this reason, this paper focuses on its development. First of all, a new equal loudness curve considering the color effect is proposed for a limited frequency band based on experimental results of audio-visual stimuli consisting of sinusoidal sounds and color patches. Secondly, this function is applied to the loudness calculation to correct the effect of colors, whereas, the other computation processes are the same as a conventional loudness model of ISO532B. This is because it is considered that the mechanism of sound perception is not affected by colors but by our ear structures. Finally, several audio-visual stimuli are evaluated and the result is compared to result analyzed by the new loudness model using the corresponding conditions and the model accuracy is discussed.

Keywords: Loudness, Audio-visual interaction I-INCE Classification of Subjects Number(s): 63.1

1. INTRODUCTION

For more than twenty years, several researches have been published regard to the loudness perception induced by audio-visual stimulus. In their works, it was revealed that the perceived loudness of sounds could be increased/ decreased if the visual stimulus was presented and the direction and magnitude of the visual effect could depend on the characteristics of the audio-visual stimulus. The majority of their researches was regard to actual noises such as trains, environments, and cars and those were evaluated with the corresponding visual stimulus presented (1, 2, 3). Because there still was a problem that the color samples used in the experiment was a few, in order to estimate the color influence on the loudness perception, some papers have also been published about the human perception of colors and concluded that red or colors close to red were perceived as the loudest (4, 5). However, the effect was confirmed to depend on the audio stimulus presented based on the experimental result of the large number of audio-visual stimuli (6). To formularize the visual effect on the loudness, by categorizing the color with some color parameters, the effect was investigated using random noises (7). In addition, a paper was published regard to the reciprocal effect between a noise coming from the structure and its color and established an equation that would explain the effect of the color on the loudness (8). In spite of the large number of the researches published, it has been still unclear how visual stimulus acts fundamentally on the loudness because those researches focused mostly on the subjective perception of actual audio-visual stimuli.

Due to this, this paper concentrates on an objective loudness model development in consideration of audio-visual interaction. Figure 1 shows the overall research scheme of this research.

In order to reveal the fundamental relationship between the sounds and colors, first of all, loudness induced by color itself is evaluated using a pairwise comparison method with a color patch

[1] kai_aizawa@camal.mech.chuo-u.ac.jp, kai.aizawa.7@gmail.com
[2] takashi_kamogawa@camal.mech.chuo-u.ac.jp
[3] arimitsu@mech.chuo-u.ac.jp
[4] toi@mech.chuo-u.ac.jp

presented on a large monitor. Then, based on the evaluation result, the loudness induced by color (defined as color-loudness in this paper) is modeled as a function of the HSV color parameter; *hue*, *saturation*, and *value* (those are written in italic form for easy understanding).

Secondly, in order to estimate the effect of color on the perceived loudness, five sinusoidal sounds with three different SPL's on the equal loudness curve are combined with some color patches and the impression is evaluated. In this evaluation, some sinusoidal sounds are presented without any visual stimuli to have the reference distance and the change in the evaluation is converted into the actual difference in SPL. Here, application of the color-loudness function is considered. Because some parameters in the function are considered to be influenced by the audio-visual interaction, the least square method is performed to optimize those parameters. Then, it is possible to estimate the change in SPL induced by a color patch that is not used in the evaluation. By applying this effect into 1/3 octave spectrum, loudness influenced both by sound and color (defined as interacted-loudness) is calculated based on ISO532B; a widely-known loudness model and the only one defined by the ISO.

Finally, the interacted-loudness model developed is validated based on the result of the evaluation and, in order to perform further validation, an additional evaluation is performed using combinations of narrow-band/broad-band random noises and color patches to discuss the possibility of the model.

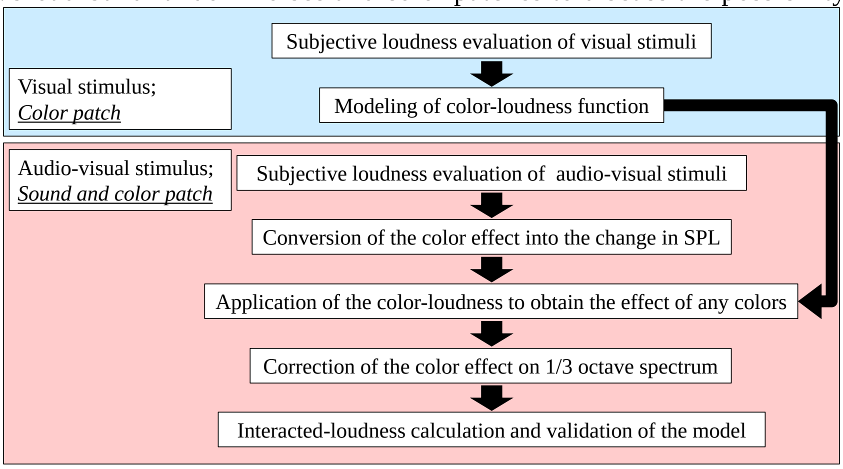


Figure 1 – Overall research scheme

2. COLOR-LOUDNESS MODEL BASED ON PSYCHOLOGICAL EXPERIMENT

2.1 Outline of the Color-loudness Evaluation

2.1.1 Conditions of the Evaluation

An evaluation is performed to identify tendency of the color-loudness. Several color patches are used and, based on the result of the evaluation, an equation that can explain the tendency of the color-loudness is established. Scheffe's pairwise comparison method revised by Nakaya is used for this psychological experiment because reiterations of the same pair do not need to be evaluated. For any possible combinations, change in the loudness of two patches is judged using seven-point-scale.

Eight students with healthy eyes participate in the experiment and all of them are familiar to the psychological or psychoacoustical evaluation. Their ages are between 21 and 25.

Figure 2 illustrates the condition of the experiment. Two subjects sit down in front of a large TV monitor (SONY KD65X8500A) installed in a sound proof chamber. The visual stimulus, color patch, is presented on the monitor for 6 sec. The distance between the monitor and subjects is 500 mm and the distance between those two subjects is 500 mm. The angle of eye-sight occupied by the monitor is 93.5 degree on horizontal direction for each subject. The subjects determine the difference in the loudness for each pair and input the data using an interface developed by MATLAB-GUI®.

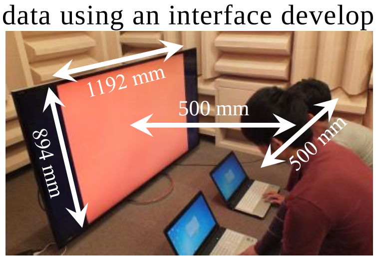


Figure 2 – Condition of the psychological experiment

2.1.2 Color Patches used in the Color-loudness Evaluation

Table 1 shows 27 color patches used in this evaluation. Those are selected based on the HSV parameters; the most common cylindrical-coordinate representation of color. *Hue* varies from 0 to 5π/3 rad with increment of π/3 rad. Each color patch has *saturation* and *value* of 0, 0.5, or 1.0. The possible combinations are 54 (= 6×3×3) but, once the *value* becomes to 0, the color is to be black; the darkest. For a special example, if the *saturation* is 0 and if the *value* is 0.5, the color is to be gray. White is made if both the *saturation* and *value* is 1. As a result, those are summarized in 27 samples.

In table 1, the HSV parameters regulated are compared to those measured in the chamber by using a colorimeter (KONICA MINOLTA CL-200A). The values of the "original" show the regulated parameters. The values of the "monitor" show the parameters measured at 100 mm away from the center of the monitor. The values of the "left/ right location" show the parameters at the eye location of the left/ right subject. By comparing these, there is a slight difference between the original and the stimulus exposed to the subjects but the correlations of those are high for each (R > 0.99). Due to this, the evaluation is considered to be worthy to the evaluation of the original stimulus itself.

Table 1 – Visual stimulus used in the evaluation

Color Number	*Hue(H), saturation(S)* and *value(V)* (*H*: 0-2π rad, *S*: 0-1.0, *V*: 0-1.0)											
	Original			Monitor			Left location			Right location		
	H	*S*	*V*	*H*	*S*	*V*	*H*	*S*	*V*	*H*	*S*	*V*
1	0.00	1.00	1.00	0.17	0.96	0.74	0.18	0.96	0.77	0.22	0.94	0.89
2	π/3	1.00	1.00	0.97	0.90	0.86	0.96	0.90	0.83	0.89	0.90	0.96
3	2π/3	1.00	1.00	1.63	0.89	0.77	1.71	0.90	0.73	1.74	0.90	0.73
4	π	1.00	1.00	3.16	0.73	0.79	3.17	0.77	0.83	3.05	0.59	0.84
5	4π/3	1.00	1.00	4.00	0.95	0.85	4.05	0.86	0.81	4.00	0.97	0.84
6	5π/3	1.00	1.00	4.98	0.70	0.82	5.00	0.70	0.79	4.97	0.70	0.81
7	0.00	0.50	0.50	0.16	0.59	0.30	0.22	0.64	0.34	0.22	0.64	0.34
8	π/3	0.50	0.50	0.91	0.59	0.33	0.88	0.61	0.35	0.82	0.63	0.37
9	2π/3	0.50	0.50	1.64	0.54	0.30	1.64	0.54	0.29	1.37	0.56	0.31
10	π	0.50	0.50	2.91	0.40	0.31	2.94	0.41	0.31	2.94	0.41	0.31
11	4π/3	0.50	0.50	4.08	0.49	0.27	4.08	0.48	0.27	4.10	0.47	0.27
12	5π/3	0.50	0.50	5.28	0.40	0.28	5.45	0.41	0.29	5.59	0.42	0.32
13	0.00	1.00	0.50	0.21	0.94	0.34	0.22	0.94	0.36	0.22	0.94	0.35
14	π/3	1.00	0.50	0.90	0.89	0.34	0.90	0.89	0.34	0.90	0.89	0.36
15	2π/3	1.00	0.50	1.59	0.87	0.28	1.57	0.87	0.29	1.54	0.87	0.29
16	π	1.00	0.50	3.03	0.56	0.30	3.03	0.55	0.30	3.04	0.57	0.31
17	4π/3	1.00	0.50	4.09	0.78	0.29	4.07	0.80	0.31	4.06	0.82	0.31
18	5π/3	1.00	0.50	5.19	0.62	0.29	5.38	0.61	0.31	5.19	0.62	0.30
19	0.00	0.50	1.00	0.21	0.62	0.76	0.21	0.62	0.77	0.21	0.63	0.80
20	π/3	0.50	1.00	1.02	0.62	0.87	1.03	0.62	0.91	1.02	0.63	0.93
21	2π/3	0.50	1.00	1.84	0.59	0.83	1.77	0.60	0.84	1.86	0.59	0.85
22	π	0.50	1.00	3.16	0.56	0.86	3.14	0.53	0.88	3.13	0.54	0.90
23	4π/3	0.50	1.00	3.93	0.69	0.83	3.93	0.70	0.85	3.93	0.70	0.84
24	5π/3	0.50	1.00	4.87	0.44	0.84	4.86	0.46	0.88	4.87	0.46	0.88
25	Any	Any	0.00	0.30	0.83	0.07	0.36	0.83	0.09	0.34	0.84	0.09
26	Any	0.00	0.50	1.47	0.10	0.32	1.54	0.10	0.34	1.35	0.10	0.34
27	Any	0.00	1.00	2.80	0.12	0.91	2.77	0.12	0.96	2.78	0.13	0.98

2.2 Result of the Color-loudness Evaluation

Figure 3 shows the evaluation result. The relative loudness is shown in terms of the *hue* angle in radian. Each color bar shows the score of the relative loudness of the corresponding colors. Here, because black, gray, and white could be made by any *hue*, the loudness scores of those are shown at every *hue* angle. The yardstick value of this experiment is calculated as 0.27, which means that two test pieces with having difference larger than the yardstick can be said to have significant difference. As a result, red (No.1) and yellow (No.2) patch are confirmed to be the loudest in those 27 samples. On the other hand, the quietest color is confirmed to be gray (No.26). In addition, the result is highly similar to the color-loudness perception of Japanese people examined in the reference paper (4).

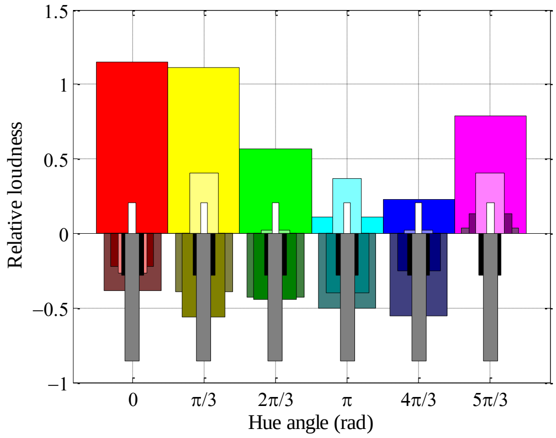


Figure 3 – Result of the evaluation

2.3 Modeling of the Color-loudness Function

2.3.1 Concept of the Function

The HSV unit is chosen because there is a possibility that the interacted-loudness can be explained by using the periodic function of *hue* as reported by Sawa (8).

Due to this, first of all, the color-loudness is modeled as a function of *hue*. Secondary, some coefficients in the function are estimated using a function of *saturation* and *value*. As a result, a color-loudness function is established in terms of *hue*, *saturation*, and *value*, which means that the function can calculate the perceived loudness based on the color parameter. Practically, the loudness needs to be able to calculate its absolute value but, in this section, just the relative value is enough to be utilized because the result of the function is converted to the absolute value afterward.

2.3.2 Estimation of the Coefficients in the Periodic Function

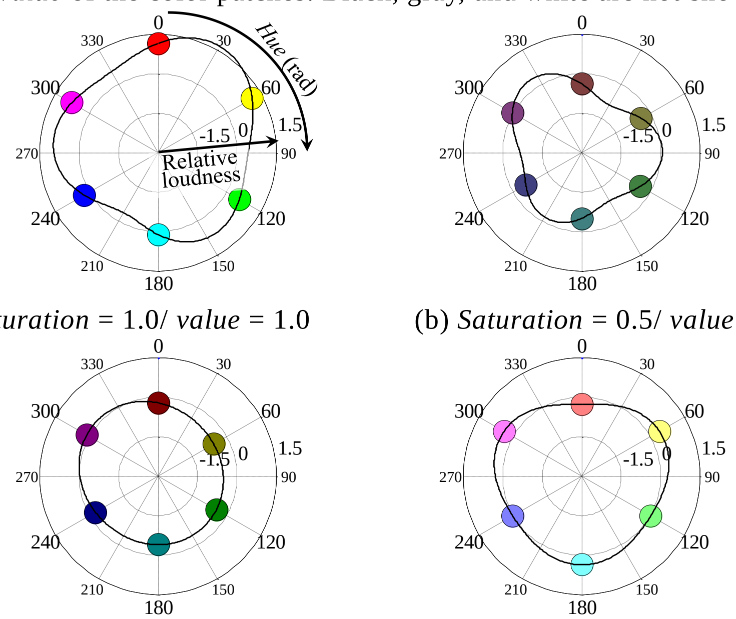
A fundamental form of the periodic function in terms of *hue* is written as shown in equation (1).

$$Loudness_{color} = A_0 + \sum_{k=1}^{i} A_k \cos(k\omega h - \theta_k) \qquad (1)$$

, where A_0 is the mesor, A_k is the amplitude of the periodic function, k is the order of the period, ω is equal to 1 rad/s because the fundamental period T is equal to 2π, h is the *hue* of the visual stimulus, and θ_k is the phase of each period.

By performing the least square method to minimize the difference between the evaluation score and the values calculated by using the equation (1), these coefficients are determined. Based on the result of the equation, the order of the function i is set to three, which means the color-loudness is represented as an equation consisting of three cosine functions with different frequencies and phases.

As a result, the coefficients explaining the evaluation score are determined. Figure 4 shows the result of the estimation. The functions are shown with black line on each figure. The evaluation scores are also plotted onto the circular coordinate systems using color circle markers in those figures. In this process, the coefficients are determined separately for each of those four conditions of the *saturation* and *value* of the color patches. Black, gray, and white are not shown here.



(a) *Saturation* = 1.0/ *value* = 1.0 (b) *Saturation* = 0.5/ *value* = 0.5

(c) *Saturation* = 1.0/ *value* = 0.5 (d) *Saturation* = 0.5/ *value* = 1.0

Figure 4 – Approximation of the evaluation score using the periodic functions

In order to understand the tendency of the coefficients in the equation (1), the relationship between the coefficients and *saturation* is considered. In the color samples, the *saturation* varies from 0.0 to 1.0. However, most of the color patches have the *saturation* of more than 0.5. In case when the *saturation* is equal to 0.0, the evaluation score of the black patch is used. In this approximation, second order equation of *saturation* is used as shown in equation (2). This equation is used for the estimation of A_0 but, for all of the other coefficients, the same form is used.

$$A_0(s) = a_{A_0} s^2 + b_{A_0} s + c_{A_0} \qquad (2)$$

Next, estimation of three coefficients in the equation (2) is performed. For each of those three, approximations are performed in terms of the *value*. An example is shown in equation (3). This is the equation for the estimation of a_{A_0} which is used for the approximation function of A_0.

$$a_{A_0}(v) = a_{a_{A_0}} v^2 + b_{a_{A_0}} v + c_{a_{A_0}} \qquad (3)$$

By combining the equation (3) into the equation (2), equation (4) is obtained. As a result, the coefficient of the periodic function is written as a function consisting of the *saturation* and *value*.

$$A_0(s,v) = a_{A_0}(v) s^2 + b_{A_0}(v) s + c_{A_0}(v) \qquad (4)$$

As a result, equation (5) is obtained by performing the same procedure to every coefficient in the equation (1) and by combining the equation (4) into the equation (1). This is the color-loudness function. Because this equation is a function of the *hue*, *saturation*, and *value*, by identifying these HSV parameters in the equation, calculation of the relative loudness for any color is now possible.

$$Loudness_{color}(h,s,v) = A_0(s,v) + \sum_{k=1}^{i} A_k(s,v) \cos\{k\omega h - \theta_k(s,v)\} \qquad (5)$$

2.3.3 Validation of the Color-loudness Function

By applying the HSV parameters of the patches used in the color-loudness evaluation into the color-loudness function, the relative loudness is calculated to judge the modeling accuracy. Although the values calculated by using the color-loudness function are left out, the estimation is confirmed to reproduce the same value as the evaluation for these color patches used because the actual evaluation score and the values calculated are confirmed to be identical (correlation coefficient R = 1.00).

3. LOUDNESS MODEL INDUCED BY AUDIO-VISUAL STIMULUS

3.1 Outline of the Interacted-loudness Evaluation

An evaluation is performed to identify the similarity between the color-loudness and the interacted-loudness. If the application of the color-loudness to the interacted-loudness is possible, estimation of the interacted-loudness of any color that is not used in the evaluation is also possible.

First of all, several combinations of the visual stimuli and audio stimuli are used to identify the relationship. Based on the result presented in the previous section, several color patches are chosen as the visual stimuli. The color number of the patches selected is No.1, 2, 3, 5, 26, and 27. No.1 and 2 are chosen because they might be the loudest as shown in the color-loudness function. No.26 is chosen as the quietest color. In addition to those three, No.3 and 5 are chosen because they are the fundamental colors; green and blue. No.27 is chosen as an example of a special color.

The evaluation is performed under the same condition as the color-loudness evaluation but, for this case, the subjects are asked to wear headphones (STAX SR303) to expose the audio stimuli. Table 2 shows the properties of the sounds. Those are chosen because they have the same loudness levels. The increment of the loudness value is 20 phon for each. Because these loudness' are calculated by a MATLAB® code provided by GENESIS® based on ISO532B, the SPL of the sounds used are different from the values on the latest equal loudness curve of ISO226:2003. This is because ISO532B is developed to reproduce the previous equal loudness of ISO226:1987.

Table 2 – 15 sounds used in the evaluation

Frequency, Hz	SPL, dB		
500	42.24	61.87	81.10
1000	39.84	59.72	79.27
2000	41.70	61.23	80.85
4000	37.94	57.94	77.63
8000	41.59	61.80	82.35
Loudness, phon	40.00	60.00	80.00

The color patch is presented on the monitor for 6 sec with the evaluation sound played. The pairwise comparison method and the number of the subjects are the same as the previous section.

In order to convert the evaluation score into the absolute value, the reference distance in the relative loudness needs to be measured. Due to this, every sound is exposed without any visual stimuli presented in case of the evaluation of the reference. For this evaluation of the sound-itself, a notification is shown on the monitor just before the sound is played and then the subjects are asked not to watch at the monitor and to evaluate the perception of the sound itself.

Because the evaluation is performed for each frequency separately, five sets of results are obtained and each axis of the results is not consistent to the other frequencies. The total test pieces used in the evaluation for each frequency are 21 (3 levels×6 visual stimuli and without any visual stimulus presented), resulting in 210 pairs to be evaluated; 1050 pairs are evaluated for total.

3.2 Result of the Interacted-loudness Evaluation

Figure 5 shows the result of the pairwise comparison. In each figure, each color bar shows the evaluation score of the sounds and the corresponding color patch. The x-marker shows the result of the evaluation for the sound-itself. The first left seven bars and marker show the evaluation scores for the combination of the audio stimuli of 40 phon. The middle seven shows the result of 60 phon and the last seven shows the result of 80 phon. Each figure contains the error bar which represents the standard deviation of the result. On the right end of each figure, the yardstick value is shown.

Based on the result, the basic tendency of the interacted-loudness is seen but the magnitude and direction of the effect are confirmed to depend both on the sound specification of the audio stimulus such as sound pressure and frequency and on the color specification of the visual stimulus.

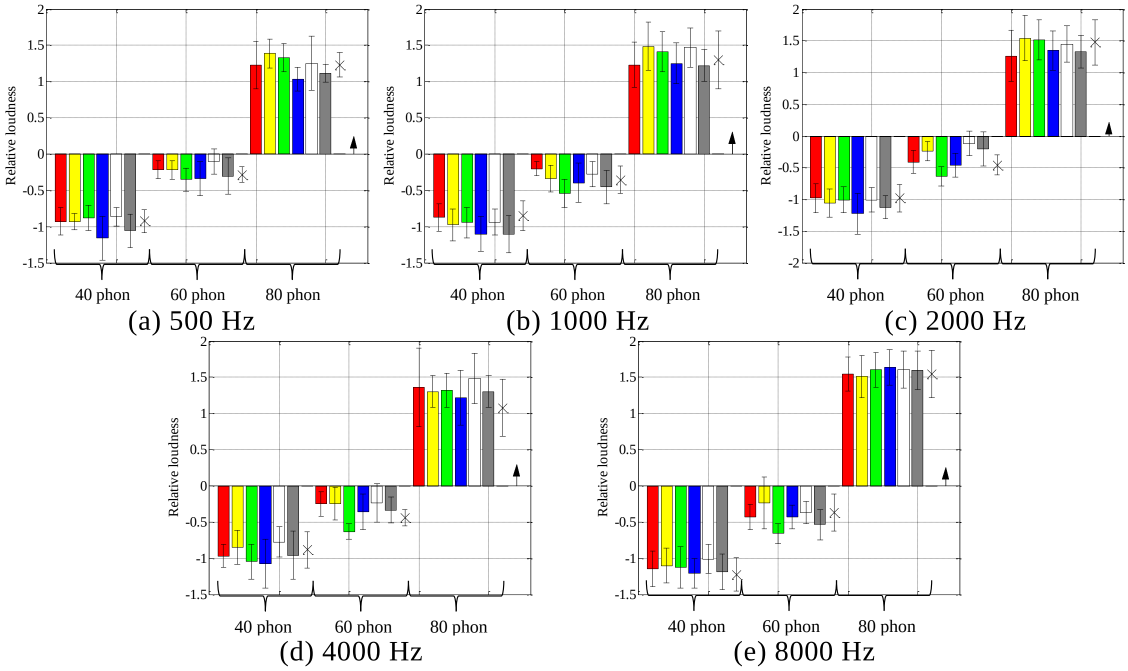


Figure 5 – Evaluation result of the audio-visual stimuli (sinusoidal sounds and color patches)

3.3 Conversion of the Evaluation Result into the Actual SPL

The result of the evaluation is converted into the actual SPL. The scaling factor is identified using the evaluation scores of the references. Because three references are used in each evaluation, the scaling factor can be calculated in three ways; scaling factor calculated using 40 and 60 phon, 40 and 80 phon, and 60 and 80 phon. For the smoother scaling, the average of these three scale factors are calculated, which means these three scaling factors are summed up and divided by three, and then, the answer is to be used as $scale(freq)$. Each scaling factor is calculated using equation (6).

$$\mathrm{scale}_{\mathrm{high\text{-}low}}(\mathrm{freq}) = \frac{\mathrm{diff}_{\mathrm{SPL}_{ref}(high-low)}(freq)}{\mathrm{diff}_{\mathrm{SCORE}_{ref}(high-low)}(freq)} = \frac{SPL_{ref(high)}(freq) - SPL_{ref(low)}(freq)}{SCORE_{ref(high)}(freq) - SCORE_{ref(low)}(freq)} \quad (6)$$

, where $SPL_{ref(high)}$ is the SPL for the higher reference that is used and $SPL_{ref(low)}$ is for the lower, and $SCORE_{ref(high)}$ is the evaluation score of the higher reference and $SCORE_{ref(low)}$ is for the lower.

For example, in case of the scaling calculation using 40 and 60 phon, the SPL and evaluation score of the reference of 60 phon are treated as the “higher” and those of 40 phon are treated as “lower”.

By using the scaling factor calculated and the evaluation scores of the reference, the difference from the reference, the color effect on SPL difference, can be calculated. The conversion is performed using equation (7). Here, the conversion of the result for 40 phon is shown.

$$diff_{color-40}(freq) = \frac{SCORE_{color-40}(freq) - SCORE_{ref-40}(freq)}{SCORE_{ref(highest)}(freq) - SCORE_{ref(lowest)}(freq)} \times scale(freq) \quad (7)$$

, where $SCORE_{color-40}$ is the evaluation score of a color patch of 40 phon, $SCORE_{ref-40}$ is the reference score of 40 phon, $SCORE_{ref(highest)}$ is the reference score of the maximum loudness and $SCORE_{ref(lowest)}$ is the reference score of the minimum. In this case, the reference of 80 phon is treated as the maximum and the reference of 40 phon is treated as the minimum. The reason why the computation of the denominator in the equation (7) is added is to regularize the calculation result.

By applying this equation to every set of the evaluation results, the relative change is converted into the change in SPL. Then, it is now possible to plot the evaluation result on the equal loudness curve obtained by ISO532B. Figure 6 shows the result. In this figure, linear approximation is performed between the every two evaluation results obtained. For example, the effect at frequencies between 500 and 1000 Hz of 40 phon is estimated using the result of 500 and 1000 Hz of 40 phon.

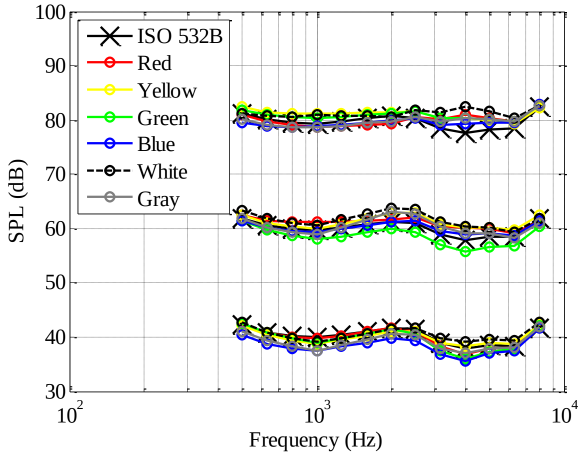


Figure 6 – Result of the evaluation in the absolute SPL

In addition, the color effect between 40 and 80 phon is estimated using the least square method. As a result, at every frequency, second order equations are established for each color based on the effect at the three different levels. By using this approximation, the change in SPL induced by the six kinds of visual stimuli at any SPL existing between 500 and 8000 Hz can be estimated. The equal loudness curve shown in the figure 6 is based on ISO532B which is based upon ISO226:1987 but, in figure 7, the new equal loudness curve defined in ISO226:2003 is utilized and the effect of the color on the curve is estimated based on the approximation mentioned above.

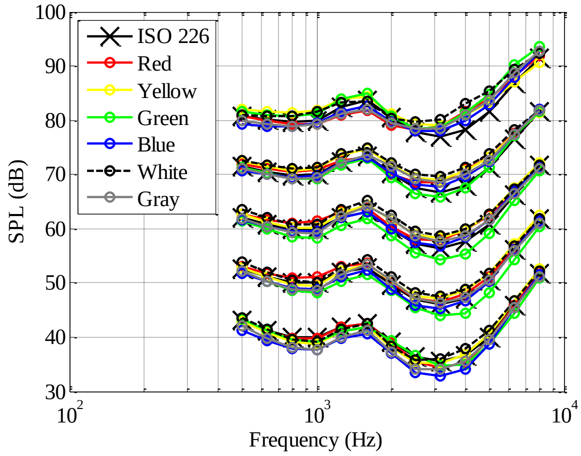


Figure 7 – Result of the color effect on the equal loudness curve of ISO226:2003

3.4 Application of the Color-loudness Function to the Interacted-loudness

In order to overcome the disadvantage of the small number of the colors used in the interacted-loudness evaluation, the color-loudness function shown in the equation (5) is expanded to explain the fluctuation of the interacted loudness. In the equation (5), the coefficients depend on the HSV parameters but these are considered to be affected also by the effect of audio-visual interaction.

Due to this, the least square method is performed to minimize the difference between the already-known color effects and values calculated using equation (8). In the optimization process, the mesor α_0, the amplitude of each periodic function α_k, and the common phase shift in all of the periodic functions τ are determined. The reason why the phase shift is determined constantly to every periodic function is that, if the accuracy of the function is high by using this constant phase shift, the loudness induced by audio-visual stimulus can be explained in terms of the base-phase shift of the color-loudness function. Because those parameters to be determined are considered to depend on the SPL and frequency, this optimization is performed to every set of SPL and frequency that are used in the evaluation of the interacted-loudness. Then, the estimation of the interacted-loudness for any color including six colors used in the interacted-loudness evaluation can be performed.

$$diff\,(SPL, freq, h, s, v) = \{A_0(s,v) + \alpha_0(SPL, freq)\} + \sum_{k=1}^{i}\{A_k(s,v) + \alpha_k(SPL, freq)\}\cos\{k\omega h - \theta_k(s,v) + \tau(SPL, freq)\} \quad (8)$$

Table 3 shows the matrix of the correlation coefficients between the evaluation and estimation. A series of the changes induced by each color obtained in the evaluation for each frequency is compared to those obtained by the equation analyzed with the same color conditions. If the equation (8) can exactly explain the color effect as measured in the evaluation, the diagonal components of the correlation matrix, which is written with red letters, should have the value of 1. Based on the results, the correlation gets worse at some points, which means the application of the equation (8) may be impossible by changing the mesor, amplitude, and phase. However, because this is seen only for a particular color and loudness, the application of the color-loudness is considered to be possible.

After the post-process of the evaluation result described in the section 3.3, the estimation of the color effect on any arbitrary spectrum is available for the six colors. To this spectrum obtained, the application of the color-loudness shown here is performed separately. Then, the color effect of any possible combination in HSV unit at any frequency with any SPL can be estimated.

Table 3 – Correlation between the evaluation and estimation

(a) 40 phon

	Estimation: Red	Yellow	Green	Blue	White	Gray
Evaluation: Red	**0.35**	0.45	0.90	0.77	0.09	0.45
Yellow	0.95	**0.96**	0.08	0.38	0.96	0.80
Green	0.27	0.32	**0.99**	0.70	0.06	0.56
Blue	0.59	0.48	0.56	**0.93**	0.69	0.94
White	0.96	0.98	0.05	0.23	**0.93**	0.69
Gray	0.89	0.83	0.29	0.57	0.98	**0.90**

(b) 60 phon

	Estimation: Red	Yellow	Green	Blue	White	Gray
Evaluation: Red	**0.98**	0.29	0.58	0.62	0.35	0.13
Yellow	0.43	**0.98**	0.63	0.63	0.77	0.71
Green	0.62	0.52	**1.00**	0.82	0.22	0.12
Blue	0.66	0.86	0.78	**0.85**	0.59	0.46
White	0.41	0.84	0.26	0.16	**0.99**	0.93
Gray	0.24	0.95	0.36	0.21	0.97	**0.96**

(c) 80 phon

	Estimation: Red	Yellow	Green	Blue	White	Gray
Evaluation: Red	**1.00**	0.63	0.92	0.76	0.95	0.95
Yellow	0.56	**0.95**	0.84	0.25	0.65	0.59
Green	0.91	0.82	**0.99**	0.68	0.96	0.93
Blue	0.86	0.33	0.72	**0.98**	0.89	0.94
White	0.91	0.76	0.96	0.80	**0.98**	0.97
Gray	0.96	0.46	0.84	0.91	0.96	**0.98**

3.5 Description of the Interacted-loudness Model

The computation of the loudness ISO532B starts with the third octave analysis of a time domain signal. The filter banks of each octave are applied to the time domain signal. By calculating the O/A values of the signals filtered by each octave filter, the 1/3 octave band spectrum can be obtained. Then, the spectrum is converted into the excitation level for each critical band. Next, correction of low frequency recognition is performed as the same as sound field and outer-ear correction. Finally, the main loudness is calculated and the masking curve is applied. By calculating the area of the spectrum, the O/A loudness value is calculated and this value is treated as the perceived loudness.

Because the loudness computation utilizes the 1/3 octave spectrum, the visual effect for each octave band needs to be considered. However, because the color effect within each octave band is considered to be almost constant, the effect measured by the sinusoidal sounds is treated as the effect on each octave. Then, the corrected octave spectrum is used in the loudness computation.

4. VALIDATION OF THE INTERACTED-LOUDNESS MODEL

4.1 Sinusoidal Example

An analysis is performed using the interacted-loudness model under the same condition of the evaluation performed with the sinusoidal sounds and color patches shown in the section 3.

Table 4 shows the comparison of the analysis and the evaluation. The values shown on the left side of each slash indicate the result of the analysis and the values on the right indicate the evaluation score. The values of the "maximum difference" indicate the absolute difference from each analysis/ evaluation value of the "sound-itself". The correlation coefficients between the evaluation and analysis of each color for each combination of the intensity and frequency are also shown.

As a result, high correlations are confirmed as shown with red letters in the table 4, which means that this interacted-loudness model can explain the tendency of the human perception of the stimuli used in the evaluation. Because both of the correlation shown in the table 3 (difference in SPL) and the table 4 (difference in loudness) is entirely high, it can be said that the difference on the equal loudness at each octave frequency can be treated as the difference in that octave band; the color effect on the octave spectrum can be calculated using the equation (8).

However, the correlation in the table 4 gets worse at some frequencies and at some SPL's and this indicates the incomplete accuracy of the loudness model. In this model, the color-loudness function is used to estimate the change in SPL and the phase shift in the function is common to every order of the period. Because the loudness model developed cannot explain the human perception completely, for the better modeling, the phase may need to be modified separately for each order of the period.

Table 4 – Comparison of the analysis/ evaluation of the sinusoidal sounds

Loudness, phon	Frequency, Hz	SPL, dB	Loudness, phon / Relative loudness							Maximum difference	Correlation
			sound itself	red	yellow	green	blue	white	gray		
40	500	42.25	40.03 / -0.92	39.93 / -0.93	39.90 / -0.93	40.73 / -0.88	37.60 / -1.16	39.93 / -0.86	38.54 / -1.06	2.43 / 0.30	0.96
	1000	39.90	40.04 / -0.85	38.81 / -0.88	39.52 / -0.98	38.98 / -0.95	37.11 / -1.10	38.64 / -0.94	37.58 / -1.11	2.93 / 0.23	0.80
	2000	41.70	40.00 / -0.98	39.55 / -0.98	38.57 / -1.06	40.24 / -1.01	37.58 / -1.23	39.40 / -1.01	37.47 / -1.13	2.53 / 0.24	0.88
	4000	37.95	40.00 / -0.88	38.13 / -0.97	40.01 / -0.85	38.55 / -1.05	36.60 / -1.08	40.90 / -0.77	39.43 / -0.96	3.40 / 0.30	0.91
	8000	41.65	40.04 / -1.23	40.69 / -1.15	40.68 / -1.10	41.10 / -1.13	40.43 / -1.21	41.45 / -1.01	40.48 / -1.18	1.41 / 0.20	0.90
60	500	61.90	59.93 / -0.29	60.78 / -0.21	60.43 / -0.22	58.96 / -0.35	59.12 / -0.34	61.10 / -0.11	60.12 / -0.30	1.17 / 0.24	0.93
	1000	59.75	59.95 / -0.36	61.38 / -0.20	59.90 / -0.34	58.15 / -0.54	59.41 / -0.40	60.38 / -0.28	59.09 / -0.45	1.80 / 0.34	0.99
	2000	61.25	59.89 / -0.46	59.93 / -0.41	61.67 / -0.24	58.88 / -0.64	59.08 / -0.46	62.62 / -0.12	61.45 / -0.20	2.73 / 0.52	0.96
	4000	57.95	60.00 / -0.44	62.36 / -0.25	62.16 / -0.25	57.63 / -0.63	60.57 / -0.36	62.07 / -0.24	60.29 / -0.33	2.37 / 0.39	0.98
	8000	61.80	59.97 / -0.37	59.63 / -0.43	60.52 / -0.23	58.68 / -0.66	59.77 / -0.43	59.95 / -0.37	59.46 / -0.54	1.29 / 0.43	0.99
80	500	81.10	79.95 / 1.23	80.12 / 1.23	80.78 / 1.39	80.15 / 1.33	78.49 / 1.04	80.08 / 1.25	78.81 / 1.11	1.46 / 0.35	0.96
	1000	79.30	79.69 / 1.30	80.29 / 1.23	81.23 / 1.49	80.09 / 1.41	79.44 / 1.25	80.38 / 1.47	79.16 / 1.22	1.54 / 0.27	0.77
	2000	80.85	79.61 / 1.48	79.08 / 1.26	80.40 / 1.54	79.44 / 1.51	78.88 / 1.35	79.80 / 1.45	79.23 / 1.33	0.79 / 0.28	0.80
	4000	77.65	79.75 / 1.08	83.20 / 1.36	82.07 / 1.30	81.58 / 1.32	80.87 / 1.22	83.76 / 1.48	82.40 / 1.30	4.01 / 0.26	0.91
	8000	82.40	79.68 / 1.54	79.64 / 1.54	79.70 / 1.51	79.71 / 1.60	79.72 / 1.64	79.90 / 1.61	79.53 / 1.60	0.22 / 0.13	0.23

4.2 Narrow-band/ Broad-band Random Noise Example

Here, some combinations of the random noises and color patches are used as input stimuli to the interacted-loudness model developed in the previous section. In order to identify the human perception of the loudness for these stimuli, another psychoacoustical experiment is conducted under the same condition as the previous experiment. Two sinusoidal sounds are also used in the experiment to have a reference distance in the resulting evaluation score. In this case, because 1000 Hz/ 40 and 80 phon are used, the evaluation score for these two is worthy to the loudness value of 40 and 80 phon; the distance between these two evaluation scores is equal to the difference of 40 phon.

Table 5 shows the result of the evaluation and the analysis. The evaluation score is converted into the actual loudness value based on the distance between the two reference evaluation score.

As a result, high correlation (R = 0.84) is confirmed between the evaluation and analysis for the entire result. In addition, the direction of the color effect estimated using the analysis is the same as the evaluation result except for the case of 2000-4000 Hz random noise/ 60 phon. This might be coming from the inaccuracy of the interacted-loudness model but, for overall, the model can be said to be significant because most of the evaluation result is equal to the fluctuation of the human perception of the interacted-loudness. Furthermore, the yardstick value of this evaluation is 0.31 in the evaluation axis, which leads 4.11 phon in loudness. Due to this, some combinations affected by the colors with having the significant difference are shown with the red letters. For those two pairs, the loudness model provides the same tendency as the evaluation result. Based on those facts found in this comparison, the accuracy of the model is confirmed at a constant level.

Table 5 – Comparison of the evaluation/ analysis result of the random noise

Number	Condition				Evaluation		Analysis
	Type	Frequency, Hz	Loudness curve, phon	Visual stimulus	Relative loudness	Loudness, phon	Loudness, phon
1	Sinusoidal	1000	40	-	-1.85	40.00	39.90
2		1000	80	-	1.17	80.00	79.30
3	Random	500-1000	60	Green	-0.98	51.52	72.85
4				Blue	-0.94	52.05	73.44
5		1000-2000	60	Yellow	-0.78	54.17	76.45
6				Green	-0.84	53.38	74.36
7		2000-4000	60	Blue	-1.08	50.20	75.96
8				Gray	-1.18	48.87	76.59
9			80	Red	1.23	**80.79**	**95.31**
10				-	0.88	**76.16**	**93.95**
11		4000-8000	80	Blue	1.10	**79.07**	**93.11**
12				White	1.46	**83.84**	**94.37**
13		500-8000	60	Red	0.67	73.38	89.18
14				Yellow	0.65	73.11	89.48
15				Green	0.51	71.26	87.03

5. Conclusions

A loudness model in consideration of audio-visual interaction was proposed based on the result of the evaluation of both the loudness induced by visual stimulus itself and the loudness induced by the audio-visual effect. The color-loudness model developed based on the result of the visual evaluation was applied to the loudness model affected both by the audio and visual stimulus. Although the visual stimuli used in the audio-visual evaluation was limited, as a result of this application, difference in SPL at any SPL/ frequency caused by the visual stimulus of any color could be calculated based on the interacted-loudness model developed. The description of the concept for this model was mentioned in the body of the paper, and in order to verify the model, the interacted-loudness values obtained by the model for some test cases were compared to the human perception induced by the color patches and sinusoidal sounds or random sounds. As a result, the interacted-loudness model was confirmed to provide the well-similar results to the actual human sensory but, in order to exclude some disagreements and to verify the model in higher level, more trials should be taken by using the other combination of color and sound or by using the actual noise in the real world. In addition to this, the color-loudness model needs to be re-considered because, in this research, the coefficients in the periodic function of hue is estimated using the function of saturation at first and then the coefficients in the function of saturation is estimated using the function of value. Because both the saturation and value is considered to affect the coefficients in the periodic function at the same time, the estimation of the coefficients should be performed based on the function consisting both of saturation and value. Those considerations are undergoing and will be presented in the near future.

REFERENCES

1. Fastl H. Audio-visual interaction in loudness evaluation. Proc ICA 2004; 4-9 April 2004; Kyoto, Japan 2004. Tu3.X1 (II-1161)
2. Rader T, et al. Crosscultural effects in audio-visual interactions. Proc Technical Committee of Noise and Vibration of the Acoustical Society of Japan; 18 June 2004; Tokyo, Japan 2004. N-2004-31
3. Menzel D, et al. Influence of vehicle color on loudness judgement. J Acoust Soc Am. 2008; 123(5):2477-2479
4. Fastl H. Cross-cultural comparison of colour evaluation using semantic differential. Proc DAGA 2008; 10-13 March 2008; Dresden, Germany 2008. p.385-386
5. Menzel D, et al. Crying colour and their influence on loudness judgement. Proc NAG/DAGA 2009; 23-26 March 2009; Rotterdam, The Netherlands 2009. p.1528-1531
6. Menzel D, et al. Colour-influence on loudness judgement. Proc ICA 2010; 23-27 August 2010; Sydney, Australia 2010. p.175-180
7. Ki-Hong K, et al. The effect of Chroma of color on perceived loudness caused by noise. Proc INTER-NOISE 11; 4-7 September 2011; Osaka, Japan 2011.
8. Sawa T, et al. Sound design of door closing sound of box type structure in consideration of the color. Proc Acoustical society of Japan 2013 Spring meeting; 13-15 March 2013; Tokyo, Japan 2013. p.1555-1558 (written in Japanese)